

# SCIENTIFIC REPORTS



OPEN

## Simulation of early DNA damage after the irradiation of a fibroblast cell nucleus using Geant4-DNA

Sylvain Meylan<sup>1</sup>, Sébastien Incerti<sup>2,3</sup>, Mathieu Karamitros<sup>3,4</sup>, Nicolas Tang<sup>1</sup>, Marta Bueno<sup>1</sup>, Isabelle Clairand<sup>1</sup> & Carmen Villagrasa<sup>1</sup>

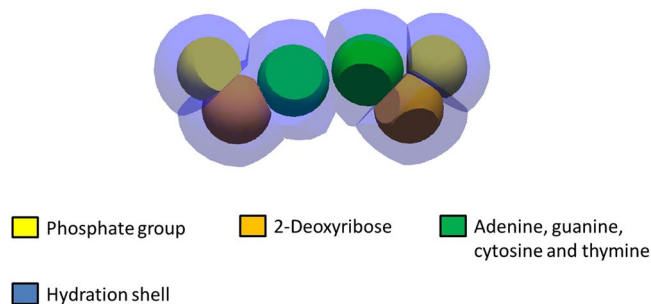
In order to improve the understanding of the mechanisms involved in the generation of early DNA damage, a new calculation chain based on the Geant4-DNA toolkit was developed. This work presents for the first time the simulation of the physical, physicochemical and chemical stages of early radiation damage at the scale of an entire human genome (fibroblast, male) and using Geant4-DNA models. The DnaFabric software was extended to generate and export this nucleus model to a text file with a specific format that can be read by Geant4 user applications. This calculation chain was used to simulate the irradiation of the nucleus by primary protons of different energies (0,5; 0,7; 0,8; 1; 1,5; 2; 3; 4; 5; 10; 20 MeV) and the results, in terms of DNA double strand breaks, agree with experimental data found in the literature (pulsed field electrophoresis technique). These results show that the simulation is consistent and that its parameters are well balanced. Among the different parameters that can be adjusted, our results demonstrate that the criterion used to select direct strand break appears to have a very significant role on the final number of simulated double strand breaks.

The biological effects of ionising radiation are an active field of interdisciplinary research that aims to improve our understanding of their deleterious nature and our ability to predict them. Improvements might have applications in many fields including medicine, radiation protection and space exploration. Better predictive capabilities would improve the accuracy of radiotherapy and hadron therapy as well as of estimates of their risks. One way of addressing this prediction uses a mechanistic approach to study the chain of physical and chemical events triggered by irradiation within a cell and leading to very early radiation induced effects. Many such studies focus on damage to the DNA molecule, considered highly sensitive to radiation<sup>1–6</sup>.

In this work, we use a mechanistic approach with Monte Carlo simulations and we focus on the damage to DNA induced by radiation. Specifically designed Monte Carlo codes, known as track structure codes<sup>7,8</sup>, must be used to adapt the study of the initial energy deposition of ionising radiation to the DNA scale (only a few nanometers). Geant4-DNA<sup>9–12</sup> processes are an extension of the Geant4<sup>13</sup> Monte Carlo code that makes possible the track structure simulations used in this work. Moreover, the simulation must be performed within a geometrical model of the DNA target to be able to compute relevant values, such as DNA double strand breaks (DSBs). This model should be accurate enough to discriminate between the physical and chemical interactions that occur within the sensitive volumes of the DNA.

The DNA geometrical models currently used in this research field range from very simple representations based on cylinders<sup>14,15</sup> to highly complex, advanced and promising depictions describing the DNA components atomistically<sup>4,16</sup>. Their complexity generally makes it hard to adapt them to the different biological conditions that may influence DNA topology, although the lack of complete knowledge of the organisation of the DNA within a cell nucleus may require this adaptation. That is, although the double helix structure of the DNA has been well described, not yet the case for the higher levels of DNA organisation such as chromatin distribution within the chromosome territories. Furthermore, the organisation of the DNA within a nucleus is also dynamic and changes with the cell cycle and the cell type. DnaFabric software<sup>17</sup> was therefore developed to facilitate the generation of complex DNA models that can go from a few pairs of nucleotides to whole-genome representations. This software

<sup>1</sup>IRSN, Institut de Radioprotection et de Sûreté Nucléaire, BP17, 92962, Fontenay-aux-Roses, France. <sup>2</sup>Univ, Bordeaux, CENBG, UMR 5797, F-33170, Gradignan, France. <sup>3</sup>CNRS, IN2P3, CENBG, UMR 5797, F-33170, Gradignan, France. <sup>4</sup>Notre Dame Radiation Laboratory, 102 Radiation Research Building Notre Dame, Indiana, 46556, USA. Correspondence and requests for materials should be addressed to S.M. (email: [sylvain.meylan@symalgo-tech.com](mailto:sylvain.meylan@symalgo-tech.com)) or C.V. (email: [carmen.villagrasa@irsn.fr](mailto:carmen.villagrasa@irsn.fr))



**Figure 1.** The nucleotide pair model formed by 6 DNA constituents and the surrounding hydration shell (blue)<sup>17</sup>.

makes it possible to generate, modify, and visualise complex DNA geometries which can also be exported for use in Geant4-DNA calculations.

This work presents for the first time the simulation of the physical, physicochemical and chemical stages of early radiation damage at the scale of an entire human genome (fibroblast, male) and using Geant4-DNA models. This simulation takes the form of a calculation chain that is based on several Geant4-DNA user applications and several analysis programs. In the end, the simulation determines the DNA damage produced by the irradiation. This paper presents the first results obtained with this calculation chain for proton irradiation at different energies and compares them with available experimental data. This comparison makes it possible to set some relevant parameters for the calculation and analysis hypothesis.

## Modelling the DNA within a cell nucleus

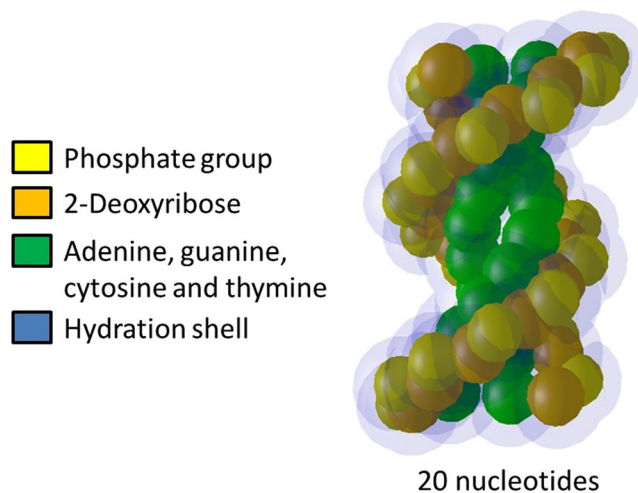
**DnaFabric software.** DnaFabric is a C++ program to generate, edit, display and export complex DNA geometrical models from the nucleotide scale to the entire DNA content of a cell-nucleus. A previous paper<sup>17</sup> described an early version of the software and presented a first set of DNA geometrical models. That first version, however, was unable to deal with geometries composed of more than  $10^5$  elements; in practice, it could only generate and manipulate a DNA fibre of roughly 18 kbp. Recent improvements enable it to work with a cell nucleus filled with an entire human genome of 6 Gbp ( $36 \cdot 10^9$  distinct volumes). The geometry generated can then be exported to a text file (extension “.fab2g4dna”) with a specific format that can be read by Geant4 user applications.

Among the various improvements to DnaFabric is a new module (“Engines”), which allows users to implement a simulation to modify a predefined DNA geometrical model. This module includes tools that can work with the hierarchical organisation of the DnaFabric geometrical models and perform multi-threaded simulations while updating the geometry rendered on the visualisation screen. Furthermore, the hierarchical organisation of the DNA models is now based on a graph structure that can define several memory-light placeholder objects used as references to a single memory-heavy object. This refinement of the hierarchical organisation allows DnaFabric to deal with billions of heavy object instances. The visualisation module was also modified to enable it to render such a huge number of objects. In practice, a level of detail (LOD)<sup>18</sup> management system was implemented to define several 3D representations for each geometrical object. Thus, an object far from the viewpoint can be displayed as a low-detail representation, while an object close to the view point is, on the contrary, fully detailed.

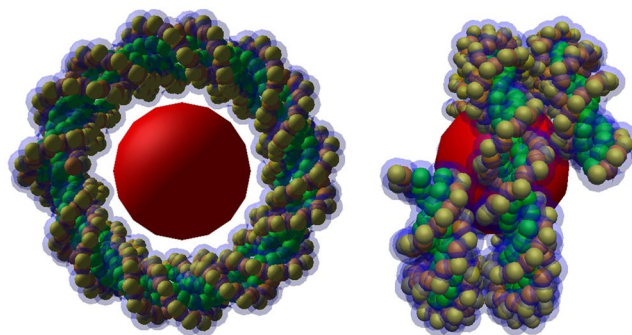
**DNA model.** The DNA model used in this work was built with DnaFabric and its elementary pre-implemented geometrical models: a nucleotide pair, histone protein, nucleosome, linker and 5 voxels filled with hetero-chromatin fibres. This section describes these built-in DNA models only briefly, since most have previously been described<sup>17</sup>.

*From the nucleotide pair to the chromatin fibre.* Six different spherical volumes were implemented in DnaFabric to represent the DNA constituents used as base units in our model: phosphate, deoxyribose, adenine, guanine, thymine and cytosine. They were used to build nucleotide pairs (the base unit of DNA) such as that presented in Fig. 1. The spherical base units were then cut to ensure that they do not overlap and thus to facilitate the use of the geometry in Geant4-DNA. Additionally, each nucleotide pair was wrapped in a volume representing 24 water molecules<sup>16,17,19</sup> to model the inner hydration shell of the DNA. Indeed, it is believed that the inner hydration shell that can transfer an ionisation from itself to the DNA<sup>19</sup>. The position and volume of each constituent within the nucleotide pair was calculated from PDB file data provided by the Glactone project<sup>20</sup>. The use of 6 spheres in the nucleotide pair model was chosen instead of an atomistic representation because it speeds-up the computations while not impacting the final outcome since an atomistic level of details is not required in our work.

The B-DNA double helix, which is the most common type of DNA double helix found in living cells<sup>21</sup> was built with pairs of nucleotides, by stacking several of them along the z axis as described in a previous publication<sup>17</sup>. This produced a B-DNA double helix similar to that depicted in Fig. 2. It was then wisted around a complex of histone proteins, represented by a single red sphere with a radius of 2.4 nm to form a nucleosome, such as that depicted in Fig. 3.



**Figure 2.** B-DNA double helix built from a stack of 10 pairs of nucleotides.



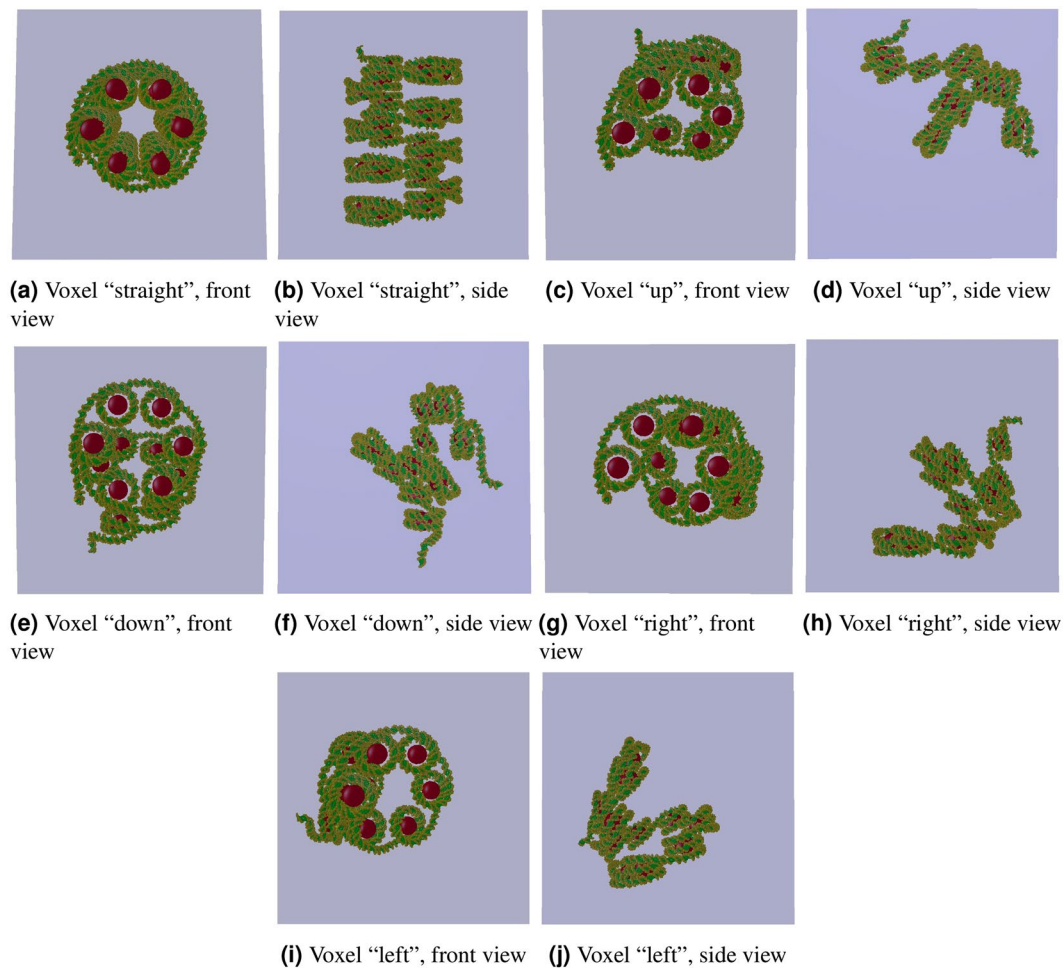
**Figure 3.** A nucleosome composed of a complex of histone proteins (red sphere) and of a twisted B-DNA double helix.

*Chromatin fibre and voxels.* Several nucleosomes were helically placed and linked together to create a continuous chromatin fibre<sup>17</sup>. In this work, pieces of the fibre (23 nucleosomes) were shaped and oriented to form a set of five different voxel configurations: “straight”, “right”, “left”, “up” and “down” voxels. They are represented in Fig. 4 and their quantitative characteristics summarised in Table 1.

**Modelling of a fibroblast cell nucleus.** Finally, a model of a fibroblast cell nucleus was built and filled with the DNA content of the human male genome. The external shape of the nucleus is ellipsoidal (half-axes dimensions: 9.85, 7.1 and 2.5  $\mu\text{m}$ ), and the genome modelled in a hierarchical form: the 5 voxels described above are used to fill chromatin domains. Each domain is represented by a sphere with a radius of 500 nm that contains several hundreds voxels ( $\sim 10^6$  pairs of nucleotides). Each domain belongs to a human chromosome, which is attributed to a spatial region of the cell nucleus: the chromosome territory. The number of domains to be placed in each chromosome territory is specified in Table 2 and is proportional to the number of base pairs (bp) within each chromosome territory.

It should be noted that the process of filling such a cell nucleus model in the G0/G1 phase is itself a three-stage simulation. It requires first the generation of the 46 human chromosomes and their empty domains. DnaFabric does this by randomly positioning one cylinder per chromosome in the cell nucleus. Each cylinder contains all the spherical domains of the chromosome in a “condensed” form. The next stage of the simulation involves “relaxing” the genome in order to obtain a distribution of the domains consistent with the G0/G1 phase. DnaFabric simulates this process according to the model previously described in the literature<sup>22</sup>. Once the relaxed genome is built, the domains are filled with DNA by adding voxels within each domain with a filling algorithm implemented in the “Engines” module. This algorithm generates DNA loops within each domain and ensures that the DNA chromatin fibre is continuous in each chromosome territory.

Once the filling process is complete, the nucleus model can be exported to a “.fab2g4dna” file for use in the Geant4-DNA simulations. Figure 5 illustrates the fibroblast cell nucleus used in this work and the DNA structure at different scales.



**Figure 4.** The 5 voxels implemented in DnaFabric: “straight”, “up”, “down”, “right” and “left”.

Characteristics	Voxels				
	Straight	Up	Down	Right	Left
Number of nucleosomes	23	11	11	11	11
Number of nucleotide pairs	4860	2466	2470	2459	2461
Pitch (nm)	10	17	17	17	17
Helix diameter (nm)	10.46	10.46	10.46	10.46	10.46
Number of nucleosomes per turn	6	6	6	6	6
Delta angle (deg)	60	60	60	60	60
Fibre radius (nm)	17	17	17	17	17
Fiber length (nm)	50	39	39	39	39
Voxel length (nm)	50	50	50	50	50
DNA volume (nm <sup>3</sup> )	$5.7 \cdot 10^3$	$2.9 \cdot 10^3$	$2.9 \cdot 10^3$	$2.9 \cdot 10^3$	$2.9 \cdot 10^3$

**Table 1.** Characteristics of the 5 voxels: “straight”, “up”, “down”, “right” and “left”.

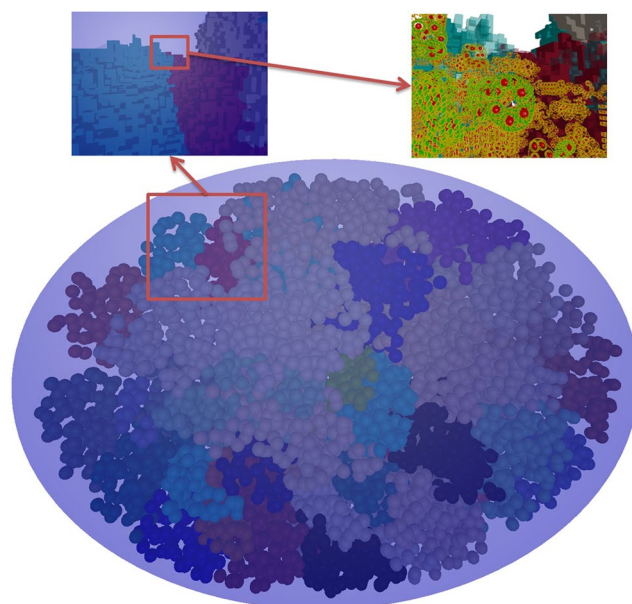
**Modelling the irradiation of a cell nucleus.** A calculation chain was developed to simulate the physical, physicochemical and chemical stages triggered by irradiation of a cell nucleus. The modular structure of the chain makes it possible to separate the main stages of the simulation, thus improving the readability of the code and allowing users to run the modules independently.

The calculation chain comprises 7 programs and 7 scripts to be executed in a specific order. The sequencing of these programs is illustrated schematically in Supplementary Figure 1.

The simulation of the physical, physicochemical and chemical stages in this chain uses a slightly modified version of the Geant4.10.1 source code. The user applications contained in the chain can handle the files exported by DnaFabric (“fab2g4dna”) that enable it to consider the full content of the human genome ( $6.4 \cdot 10^9$  nucleotide pairs) during the simulation. The aim of the calculation chain is to compute the yields of DSBs generated per

Pairs of chromosomes	Chromosomes	Number of domains
1	1–2	2 × 250
2	3–4	2 × 242
3	5–6	2 × 198
4	7–8	2 × 190
5	9–10	2 × 182
6	11–12	2 × 171
7	13–14	2 × 159
8	15–16	2 × 145
9	17–18	2 × 138
10	19–20	2 × 134
11	21–22	2 × 135
12	23–24	2 × 133
13	25–26	2 × 114
14	27–28	2 × 107
15	29–30	2 × 102
16	31–32	2 × 90
17	33–34	2 × 83
18	35–36	2 × 80
19	37–38	2 × 59
20	39–40	2 × 64
21	41–42	2 × 47
22	43–44	2 × 51
23	X	156
23	Y	57

**Table 2.** The 23 pairs of chromosomes in the human genome (source: [www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)).



**Figure 5.** A relaxed and filled fibroblast cell nucleus. The three views presented use three different levels of zoom. The main view (bottom) shows the whole nucleus from a distant point from which only the domains are visible. On the top left, the view is that of a zoom in on the red boxed area of the main view. At this level, it is possible to see the voxels as cubes. On the top right, the view comes from another zoom in on the red boxed area of the top left view. In this view, the detail of the DNA within the voxels is at a molecular level.

primary particle from both the direct (physical stage) and indirect (chemical stage) effects. Nevertheless, the user can also access intermediate results to, for example, separate the direct from the indirect DNA strand breaks (SBs) or determine the base damage and its location.



The calculation chain can be divided into four parts:

- A Geant4-DNA user application to simulate the physical stage (“phys\_geo”).
- A Geant4-DNA user application to simulate the physico-chemical and chemical stages (“chem\_geo”).
- A clustering algorithm to reveal DNA cluster damage with a user-specified distance parameter (“DBScan”).
- Several analysis routines and scripts to synchronise the elements of the calculation chain and to process the results it generates.

The main hypotheses, assumptions and parameters associated with each of these parts are detailed in the sections below.

**Simulation of the physical interactions within the DNA geometry.** The physical interactions between the incident protons (including secondary electrons) and the DNA target were simulated with the physical models present by default in Geant4-DNA (version 10.1) and already detailed in the literature<sup>9,10,12</sup>. The DNA volumes described in section title:dna\_models were filled with liquid water, which constitutes an approximation for biological medium, to simulate the physical interactions because the Geant4-DNA models available in version 10.1 use interaction cross sections in liquid water only.

The geometrical description of the DNA used in this user application (“phys\_geo”) comes from six “.fab-2g4dna” files. The first file describes the cell nucleus and contains the position and type of each of the voxels within it. The other files detail the DNA content of the five voxels introduced in section title:dna\_models. The voxels and nucleus in the user application are imported by a parser that generates the corresponding Geant4 geometry. A cutting algorithm is also used to deal with the geometrical overlaps that can appear after conversion into Geant4 geometry. This algorithm is executed once for each of the 5 voxels included in the simulation.

The introduction of a cell nucleus model and its DNA content into the “phys\_geo” user application requires the use of specific features. Specifically, the inclusion of several million voxels into the Geant4 simulation necessitates a parametrisation process to keep the amount of memory required by the simulation at a reasonable level (<10 GB). Parameterisation in Geant4.10.1 allows the user to define a volume in memory only once and then to use this definition to represent a large number of identical volumes in the simulation, which reduces the amount of required memory. This method cannot be used, however, if different types of volumes are parameterised in the same area and if the multithreading mode is enabled, as it is in our case to speed up the simulation. We therefore modify Geant4 to enable us to parameterise the 5 different types of voxels and resolve this issue; principally, some variables were made thread-local to avoid data race issues when the multithreading mode is activated. This modification finally allowed the simulation of the physical stage in a cell nucleus filled with around  $6.4 \cdot 10^9$  pairs of nucleotides.

Two datasets are stored in an ntuple generated by the ROOT-CERN library<sup>23</sup> during this simulation. The first set is used to calculate the DNA damage induced by the physical interactions (direct effects) and the second to generate the input data for the chemical stage (indirect effects). The simulations can thus be separated so that those for the different stages can be run independently when necessary. The first dataset corresponds to the physical interactions located in the DNA volumes: 2-deoxyribose, phosphate, adenine, guanine, thymine, cytosine and their hydration shells. Specifically, the information recorded there concerns the type of interaction (ionisation, elastic etc.), the particles (type and energy) and the DNA molecules involved (name and spatial localisation). The second set of data includes the water molecules that have been ionised or excited and the solvated electrons (physical characteristics and position). Water molecules and solvated electrons are saved only if they are located within a voxel to limit the size of the output file.

**Physicochemical and chemical stages.** Like the physical stage, the physicochemical and chemical stages are simulated with a Geant4 user-application (“chem\_geo”) built especially to take the geometrical description of the DNA exported from DnaFabric into account. This time, however, the general idea is to consider the DNA model not as a group of Geant4 physical volumes but as a set of spatially ordered molecules that should not diffuse over time. The modifications introduced in the Geant4-DNA chemistry module (version 10.1)<sup>24</sup> to make this possible are summarized below:

- A number was associated with each DNA molecule to indicate its DNA strand (1 or 2).
- The capability to deal with coefficients of diffusion equal to zero was introduced to specify that each DNA molecule should be treated as a “static” object.
- A new type of reaction was added to prevent one of the reactants from disappearing. It implements the absorption capacity of the histone proteins, which will remain in the simulation regardless of the molecules they may have absorbed.

The DNA molecules presented in Table 3 are thus included in the simulation and a specific parser is included in the “chem\_geo” user application. The parser reads the output of the physical stage simulation to introduce unstable water molecules and solvated electrons into this chemical part of the simulation. The parser then processes the “.fab2g4dna” files to generate and place in the appropriate space the DNA molecules to be included in the simulation. The set of reactions shown in Table 4 was added during this work to the default set of chemical reactions of the Geant4-DNA chemistry model (see also Table 5) to allow the DNA molecules to react with the chemical species induced by irradiation. In particular, only the OH• radical was considered able to react with the DNA molecules<sup>25,26</sup>. The latter is an acceptable approximation in this work because the reactions between

DNA molecule	Radius (nm)
2-deoxyribose	0.29
Phosphate	0.27
Adenine	0.3
Thymine	0.3
Guanine	0.3
Cytosine	0.3

**Table 3.** Characteristics of the DNA constituents taken into account during the simulation of the chemical stages<sup>17</sup>.

Reaction	Reaction rate ( $10^9 M^{-1} \cdot s^{-1}$ )
2-deoxyribose + $OH^\bullet$	2.5
Adenine + $OH^\bullet$	6.10
Guanine + $OH^\bullet$	9.20
Thymine + $OH^\bullet$	6.40
Cytosine + $OH^\bullet$	6.10
Histone + molecule $\rightarrow$ histone <sub>modified</sub>	—

**Table 4.** Reactions added to the default Geant4-DNA chemical module<sup>27–29</sup>. The last reaction simulates the histone protein as an “absorber”: histone absorbs any molecule that directly touches it (no reaction rate is attributed to it).

Reaction	reaction rate ( $10^{10} M^{-1} \cdot s^{-1}$ )
$H^\bullet + e_{aq}^- + H_2O \rightarrow OH^- + H_2$	2.65
$H^\bullet + OH^\bullet \rightarrow H_2O$	1.44
$H^\bullet + H^\bullet \rightarrow H_2$	1.20
$H_2 + OH^\bullet \rightarrow H^\bullet + H_2O$	$4.17 \cdot 10^{-3}$
$H_2O_2 + e_{aq}^- \rightarrow OH^- + OH^\bullet$	1.41
$H_3O^+ + e_{aq}^- \rightarrow H^\bullet + H_2O$	2.11
$H_3O^+ + OH^- \rightarrow 2H_2O$	14.3
$OH^\bullet + e_{aq}^- \rightarrow OH^\bullet$	2.95
$OH^\bullet + OH^\bullet \rightarrow H_2O_2$	0.44
$e_{aq}^- + e_{aq}^- + 2H_2O \rightarrow 2OH^- + H_2$	0.50

**Table 5.** Default reactions of the Geant4-DNA chemical module (version 10.1)<sup>11</sup>.

2-deoxyribose and  $e_{aq}^-$  or  $H^\bullet$  are associated with reaction rates that are much lower<sup>27</sup> than the one associated with the reaction involving 2-deoxyribose and  $OH^\bullet$ .

The physicochemical and chemical stages are not simulated in the cell nucleus as a whole; instead, reactions are limited to particular voxels. Moreover, reactants can react with one another only if they are produced by the same track (independent track approximation). The separation of the chemical stage simulation within the different voxels is due to the need to minimise memory use and simulation time. Considering the entire nucleus with its human genome simultaneously during the chemical stage would have required loading about  $36 \cdot 10^9$  individual molecules, which exceeds the limitations of not only the chemistry module (Geant4-DNA version 10.1) but also current hardware. Fragmentation of the simulation to isolated voxels allowed us to reduce the amount of memory required drastically. In such a configuration, each voxel contains no more than 10000 individual molecules, which is easily manageable. The drawback of this separation is the need to run numerous different simulations (one simulation per event/voxel pair). On the other hand, it facilitates the distribution of the chemical stage simulations on multiple threads through the use of pseudo-parallelism (one simulation per thread).

The physicochemical stage is simulated with the default “dissociation channels” given in the chemistry module<sup>11</sup>. The dissociation channels describe how an unstable water molecule that has been ionised or excited during the physical stage will decay into chemical species. These chemical species are then randomly placed in a sphere of 1 nm centered on the position of the former unstable water molecule. The resulting chemical species then diffuse and react with each other and with solvated electrons or DNA molecules during the chemical stage. The simulation of the chemical stage takes place in several time steps during which all the molecules move according to their diffusion coefficients<sup>11</sup>. Two potentially reactant molecules can trigger a reaction alongside this movement, initiated either through spatial proximity determined at the end of each time step or during a time step, through the so-called “Brownian Bridge” technique. Scavenging reactions that decrease the number of  $OH^\bullet$  radicals available

to damage the DNA were not specifically modelled in this calculation. A simplification of these scavenging reactions was taken into account by different methods: histone reactions, voxel spatial limitation and, most importantly, by limiting the chemical stage simulation time to 2.5 ns<sup>30</sup>. Other similar simulation codes<sup>31</sup> use a 10 ns duration but they take into account the scavenging of the chemical species through random absorption of the radicals at each time step.

**Calculation of strand breaks and double strand breaks.** *Determination of direct strand breaks.* The data generated during the simulation of the physical stage does not allow direct DNA damage to be identified immediately. More specifically, the cartography of all the interactions that take place within the DNA is available but does not necessarily correspond to direct damage ( $SB_{\text{direct}}$ ) to the DNA molecule. Determinations of which interactions leads to an SB or to base damage and of whether the damage occurs in the atom in which the energy was deposited or if a charge transfer occurs, are still the subject of active research. In general, ionisation and excitation occurring within the DNA are considered able to induce DNA damage under some conditions<sup>1</sup>. It is also commonly accepted that ionisation taking place within the DNA hydration shell can lead to direct DNA damage through a charge transfer process<sup>32</sup> and that a dissociative attachment<sup>33,34</sup> can create a resonance effect able to alter the DNA structure<sup>35,36</sup>. The latter finding implies that electrons with energies inferior to those required to ionise or excite DNA molecules can still lead to DNA damage. Precise and complete data about the process by which physical interactions causes direct DNA damage remain sparse. Modelling thus requires making assumptions, and in mechanistic simulations, it usually assumes selection based on the amount of energy deposited in sensitive parts of the DNA. The amount of energy and the sensitive volumes change with each simulation code<sup>31,37</sup>.

In this work, the criterion chosen to calculate the number of  $SB_{\text{direct}}$  from the energy depositions registered during the simulation of the physical stage is a cumulative deposited energy of at least 17.5 eV<sup>38–40</sup> in the combined phosphate and 2-deoxyribose (hydration shell included) constituents of a nucleotide pair, the region generally known as the “backbone” of the DNA double helix. Nevertheless, a linear probability was also tested to estimate the influence of this selection process on the amount of DNA damage. This probability increases linearly from 0 for a deposited energy less than 5 eV, to 1 when the deposited energy exceeds 37.5 eV<sup>31</sup>.

It should be noted that energy depositions are computed within DNA but with data about liquid water which constitutes an approximation.

*Determination of indirect strand breaks.* During the simulation of the chemical stage, every chemical reaction defined in the code is saved in an output file for later analysis. In the current implementation of the analysis, used for this work, only reactions between  $OH^{\bullet}$  and 2-deoxyribose can generate an indirect SB ( $SB_{\text{indirect}}$ ). Those reactions, however, are not all necessarily considered indirect SBs. Instead, when such a reaction is detected, a uniform probability of ~40% ( $=\frac{2}{5}$ ) is applied to decide whether it converts into an indirect SB. This probability is applied because the structure of the DNA chain allows only 2 of every 5 reactive sites of the 2-deoxyribose molecule to be reached by the  $OH^{\bullet}$ <sup>41,42</sup>. It also implies that an average of 11% of all the chemical reactions between  $OH^{\bullet}$  and DNA will lead to an indirect SB.

*Calculation of double-strand breaks and their complexity.* After the  $SB_{\text{direct}}$  and  $SB_{\text{indirect}}$  are identified and localized within the DNA geometrical model, a clustering algorithm is used to calculate the number of DSBs. In this work, a DSB is defined as a cluster containing at least two SBs separated by less than 10 bp and with at least one SB per strand. This clustering takes place in a merging process that starts by forming initial clusters of SBs separated by less than 10 bp. The clusters are then merged if they share one of their points (that is, one SB). At the end of the merging procedure, the clusters obtained are composed of at least two SBs. This work describes the final number of SBs contained in each cluster as the cluster (or DSB) complexity. It should be noted that this definition of DSB complexity does not include base damage, as the clustering algorithm considers only the SBs. Similarly, clusters formed of two or more SBs that are all located on the same strand are identified as complex single-strand breaks (SSBs) here. Like for the DSBs, their complexity indicates the number of SBs in the cluster. Any isolated SB not belonging to any cluster is considered a simple SSB.

Finally the calculation chain presented here computes the number of DSBs by applying a set of default parameters that can be easily changed by the user. These are:

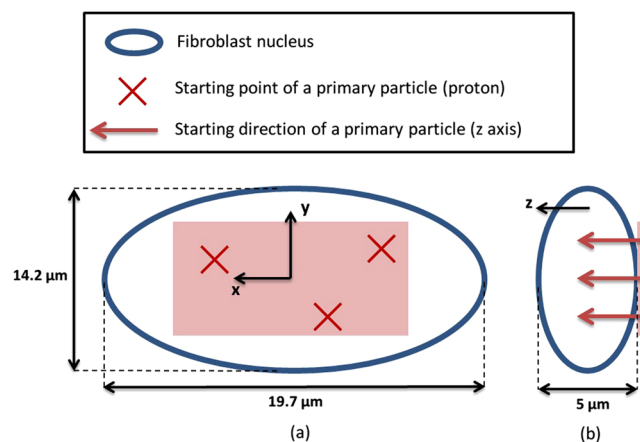
- A 17.5 eV threshold for the energy deposited in the backbone of a nucleotide to determine an  $SB_{\text{direct}}$ .
- 40% of the chemical reactions between  $OH^{\bullet}$  and a sugar (2-deoxyribose and phosphate) give rise to an  $SB_{\text{indirect}}$ .
- A chemical stage duration of 2.5 ns.
- A cluster defines a DSB if it comprises at least two SBs located on opposite strands and separated by less than 10 bp.

*Analysis required to compare simulated results with experimental data obtained by pulsed field gel electrophoresis.* The simulation performed with the calculation chain provides a set of DSBs and SSBs with their associated complexity per simulated primary particle (pp). However, further steps are required to compare the yield of DSB/pp to experimental data obtained with a technique known as pulsed field gel electrophoresis<sup>43–46</sup>. In this case, our simulated results require further processing to take experimental constraints into account:



Energy (MeV)	$LET_{d,\infty}$ (keV/ $\mu\text{m}$ )
0.5	47.9
0.7	36.5
0.8	32.6
1	27.2
1.5	19.9
2	16.1
3	11.8
4	9.6
5	8.0
10	4.6
20	2.6

**Table 6.**  $LET_{d,\infty}$  of the protons used in the simulations performed.  $d$  represents the mean distance travelled within the cell nucleus by each proton and  $\infty$  means that the energy depositions of all the secondary particles are taken into account.

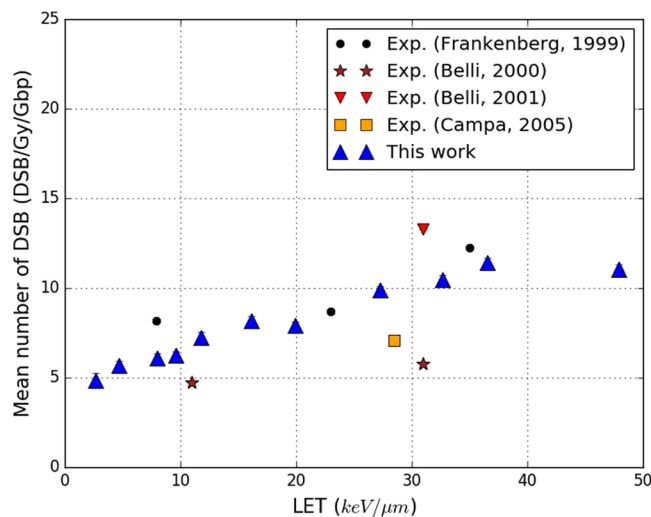


**Figure 6.** Source of the protons used as primary particles in the simulations. (a) Represents a view from the top of the fibroblast nucleus and (b) a profile view.

1. DSBs are detected indirectly since experimental data are basically numbers of DNA fragments resulting from DSBs generated during the irradiation. Knowing that two DSBs are required to create a DNA fragment, it is possible to deduce a number of DSBs from the number of detected fragments. However, the technique has low resolution and some of the smaller fragments are not detected; this artificially decreases the number of experimental DSBs.
2. Experimental results are not given as DSBs per primary particle (proton) but as DSB per Gy and per Gbp (DSB/Gy/Gbp). Furthermore, it should be noted that several Gy are delivered to the cell nucleus in each experimental irradiation and that extrapolation is used to obtain the number of DSB/Gy/Gbp.

To take these two points into account, an additional analysis routine was added to the calculation chain. In this analysis, the position of each simulated DSB in the human genome is used to calculate the fragment size. If the fragment size is lower than the detection threshold of the experimental data (10000 bp), it is removed. A final number of fragments per primary particle (pp) is therefore obtained, which takes into account the resolution constraints of pulsed field gel electrophoresis experiments. This number of fragments corresponds to a number of distant DSBs per primary particle ( $DSB_{distant}/pp$ ). It is, then, multiplied by the yield of primary particles required to deposit 1 Gy in the cell nucleus and divided by the number of Gbp included in the human genome (diploid cell:  $\sim 6.4$  Gbp).

**Configuration of the simulations.** This work simulated the irradiation of a fibroblast cell nucleus. The geometrical model used to represent the fibroblast nucleus was generated as explained in section title:models\_cell\_nucleus and filled with the DNA content of a human male genome. The primary particles used in these simulations were mono-energetic protons of 0.5 MeV, 0.7 MeV, 0.8 MeV, 1 MeV, 1.5 MeV, 2 MeV, 3 MeV, 4 MeV, 5 MeV, 10 MeV and 20 MeV.  $LET_{d,\infty}$  associated with these energies in the cell nucleus were calculated (with Geant4) and are shown in Table 6. All secondary electrons were taken into account in the calculations of the  $LET_{d,\infty}$  to simulate the electronic equilibrium caused by the broad beam irradiations reported in the literature<sup>43–46</sup>. The primary proton source is represented by a square surface ( $16 \times 12 \mu\text{m}$ ) placed above the cell nucleus. The direction of the particles is parallel to the Z axis. Figure 6 illustrates this configuration. The primary proton source covers only



**Figure 7.** Yields of DSB/Gy/Gbp simulated in this work, when applying the default parameters of the calculation chain (“This work”). The results depend on the LET of the protons used as primary particles. Experimental data from the literature are also presented: Frankenberg *et al.*<sup>43</sup> (black points), Belli *et al.*<sup>45,46</sup> (brown stars and red triangles) and Campa *et al.*<sup>44</sup> (orange squares). Statistical uncertainties are shown for this work unless they are too small to be seen.

92% of the volume of the nucleus to avoid the simulation of tracks in the area near the border of the nucleus. Indeed, there is only 3% of the nucleus DNA in these area because the filling algorithm is less effective in such restricted spaces.

The output of the simulation is a mean number of DSBs per track that is converted to a DSB yield per Gy and per Gbp using the following normalization:

$$N_{DSB/Gy/Gbp} = N_{DSB/event} \cdot \frac{E_{1Gy}}{\bar{l} \cdot LET_{d,\infty}(E_p) \cdot n} \times F \quad (1)$$

with  $N_{DSB/Gy/Gbp}(s_{bp})$  the number of DSB computed per Gy and Gbp,  $N_{DSB/event}$  the number of DSB per track simulated,  $E_{1Gy}$  the amount of deposited energy (eV) required to get one Gy in the nucleus volume,  $\bar{l}$  the mean path of each track in the nucleus,  $LET_{d,\infty}(E_p)$  the LET of the primary with an energy of  $E_p$ ,  $n$  the number of Gbp in the nucleus and  $F$  a factor to take into account that only 92% of the nucleus volume and 97% of the DNA is irradiated. Specific computations were done to determine the parameters of this equation and the results are:

$$\begin{cases} E_{1Gy} = 4570 \text{ keV} \\ \bar{l} = 3,5 \text{ } \mu\text{m} \\ n = 6.4 \text{ Gbp} \\ F = 0.95 \end{cases}$$

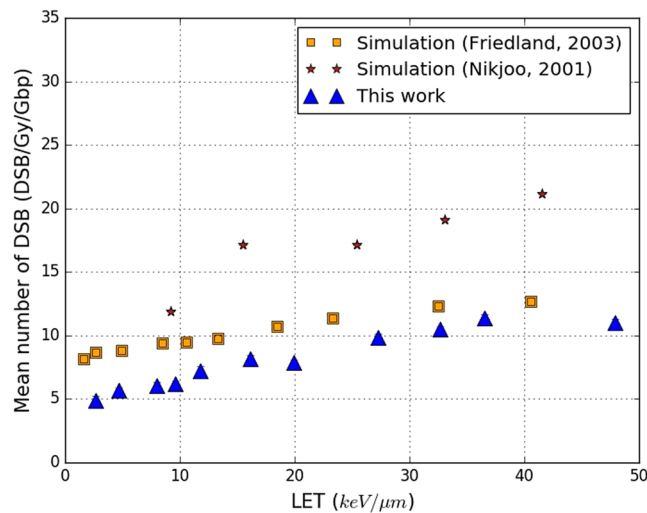
The statistical relevancy of the simulation results is controlled by a dedicated module that starts new batches of 1000 primaries until the statistical uncertainty on the DSB yields is lower than a user specified value. In this work, this value was set to 2% which means that around 5000 primaries were simulated for each energy. Simulations were performed in parallel on a computer cluster which each node was in charge of computing results for one LET value. There were 24 threads per node and the simulations lasted around 3 weeks (depending on the particle energy).

Initially, the default parameters of the calculation chain were used to compute the number of DSBs in the DNA and compare them with both experimental<sup>43–46</sup> and simulated data<sup>31,40</sup> from the literature. A minimum fragment size of 10000 bp was used to compute the number of DSB/pp for all the results of section title:dna\_damages.

In a second simulation, the criterion used to determine the  $SB_{direct}$  (17.5 eV threshold) was modified in a sensitivity analysis to estimate its influence on the final number of DSBs. The threshold was changed to 12.5 and 30 eV and replaced by the linear probability presented above.

## DNA strand breaks simulated with Geant4-DNA

**Comparison of the results with data from the literature.** Figure 7 shows the yield of DSB/Gy/Gbp simulated in this work compared to experimental data measured by pulsed field gel electrophoresis by: Frankenberg *et al.*<sup>43</sup>, Campa *et al.*<sup>44</sup> and Belli *et al.*<sup>45,46</sup>. Our simulation reproduced the experimental conditions of Frankenberg *et al.*<sup>43</sup>. The results of Campa *et al.*<sup>44</sup> and Belli *et al.*<sup>45,46</sup> came from V-79 Chinese hamster cells, and the lowest fragment size was higher than reported by Frankenberg *et al.*<sup>43</sup>: 23000 bp in the work of Belli *et al.*<sup>45,46</sup> Despite these differences, the results are included in Fig. 7 to illustrate the scatter of the data. In general, experimental data show that the yield of DSB/Gy/Gbp increases with the LET of the primary protons: the data from



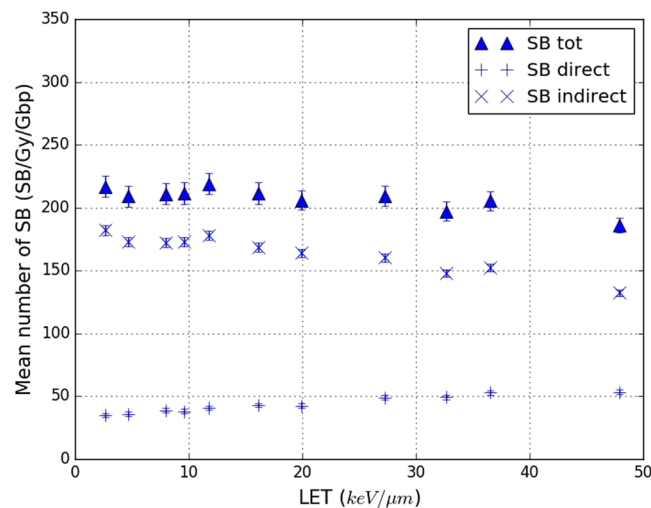
**Figure 8.** Yields of DSB/Gy/Gbp simulated in this work (blue triangles) when applying the default parameters of the calculation chain (“This work”). Simulated results obtained with the PARTRAC code<sup>31</sup> (orange squares) and the KURBUC code<sup>40</sup> (brown stars) are also included. Results are shown as a function of the LET of the primary protons. Statistical uncertainties are shown for this work unless they are too small to be seen.

Frankenberg *et al.*<sup>43</sup> start at 8.16 DSB/Gy/Gbp for 7.9 keV/μm and end at 12.22 DSB/Gy/Gbp for 35 keV/μm. The scattering of the data from Belli *et al.*<sup>45,46</sup> and Campa *et al.*<sup>44</sup> around those of Frankenberg *et al.*<sup>43</sup> illustrates the weight of the uncertainties associated with this kind of experimental measurement and the influence of biological factors such as cell type. The results obtained in this work also increase with the LET starting at 5 DSB/Gy/Gbp for 2.6 keV/μm and up to 11.3 DSB/Gy/Gbp for 47.9 keV/μm. Overall, the agreement between our results here and the experimental data is good. This is especially true for the data of Frankenberg *et al.*<sup>43</sup> around a LET of 20 keV/μm. Interestingly, our results are slightly lower than the experiment for LET lower than 15 keV/μm and higher than 30 keV/μm.

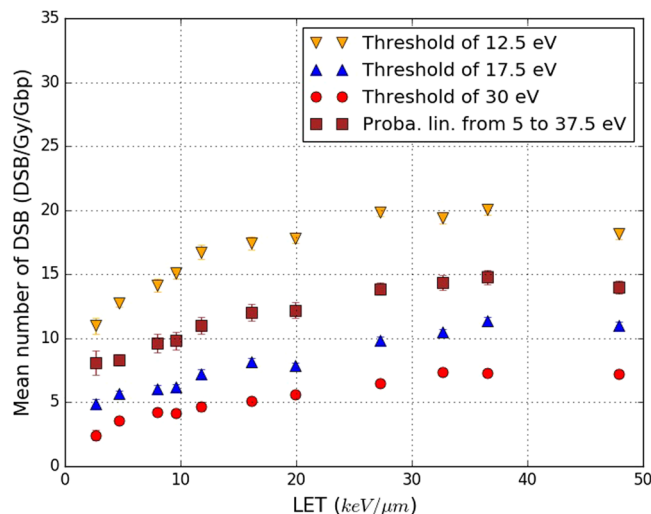
Figure 8 compares the yield of DSB/Gy/Gbp calculated in this work with results presented by Friedland *et al.*<sup>31</sup> and Nikjoo *et al.*<sup>40</sup> simulated respectively with the PARTRAC and KURBUC codes. As previously, our results are based on the default parameters of the calculation chain. It should be noted that some hypotheses differ from those used in PARTRAC or KURBUC. One example is the  $SB_{direct}$  selection criterion in PARTRAC<sup>31</sup>, which uses the linear acceptance probability previously described (between 5 and 37.5 eV). The duration of the chemical stage also differs: it is of  $10^{-9}$  s in KURBUC<sup>40</sup> and  $2.5 \cdot 10^{-9}$  s in our work. Despite these differences, the results with PARTRAC and KURBUC are comparable to the ones obtained in this work because all of them use similar methodology and the same experimental measurements as references (see Fig. 7). Figure 8 shows that the yield of DSB/Gy/Gbp increases with the LET for all the simulation codes. However, this increase appears to be linear with KURBUC but not with either PARTRAC or our results. In both of the latter cases, the increase of the DSB/Gy/Gbp is less accentuated above a LET value of 35 keV/μm. Furthermore, for LET higher than 35 keV/μm, our results are close to those with PARTRAC (relative difference less than 10%). On the other hand, our results are lower than those with PARTRAC for all LET lower than 35 keV/μm. For example, our results are 35% lower than those with PARTRAC for a LET of 4.6 keV/μm. The results with KURBUC are higher than both those with PARTRAC and our findings. Indeed, the difference between the yields of DSBs computed with KURBUC and in this work varies between 5 and 10 DSB/Gy/Gbp for all LET considered.

The results shown in Figs 7 and 8 above are the yields of DSB/Gy/Gbp calculated by simulating the physical, physicochemical and chemical stages and then processing the SBs produced during the physical and chemical stages to determine these rates. Nonetheless, we can extract from these results the total number of strand breaks ( $SB_{tot}$ ) at the origin of the DSBs and SSBs (cf. section title:  $sb\_calc$ ). We can also discriminate between the SBs from the physical stage (that is,  $SB_{direct}$ ) and from the chemical stage (the  $SB_{indirect}$ ). Figure 9 shows the yield of  $SB_{tot}$ ,  $SB_{direct}$  and  $SB_{indirect}$  simulated in this work as a function of the LET of the protons used as primary particle. As previously, the results are presented per Gy and per Gbp. The total number of SBs obtained in the simulation ( $SB_{tot}$ ) is almost constant at around 220 for LET values below 20 keV/μm. Nonetheless, for LET values higher than 20 keV/μm, the number of  $SB_{tot}$  decreases to 185 for a LET of 47.9 keV/μm. The number of  $SB_{direct}$  is quite stable until LET values of 20 keV/μm but substantially lower than those for either  $SB_{tot}$  or  $SB_{indirect}$ . Specifically, there are around 40  $SB_{direct}$  for LET values below 2.6 to 20 keV/μm, equivalent to only 20% of the  $SB_{tot}$  and 24% of the  $SB_{indirect}$ . For LET values higher than 20 keV/μm, the number of  $SB_{direct}$  progressively increases to 52 at 47.9 keV/μm.

**Influence of the criterion used to identify direct single-strand breaks.** Figure 10 presents four sets of results for the yield of DSB/Gy/Gbp calculated as a function of the LET of the primary protons. They differ in the selection criterion used to identify the  $SB_{direct}$  generated during the physical stage of our simulation, as explained in section title:  $configs$ . Figure 10 shows that the variations of the yield of DSB/Gy/Gbp with

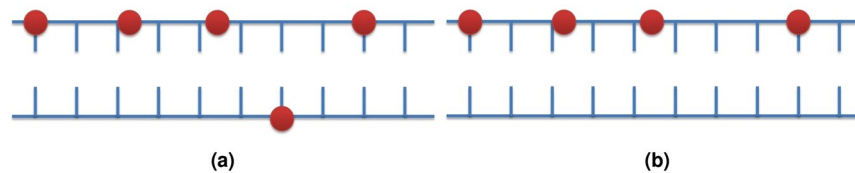


**Figure 9.** Yields of total, direct and indirect strand breaks (SBs) obtained in this work (default parameters) as a function of the LET of the primary protons. Triangles represents the total number of SBs (SB tot), “+” the SBs from the physical stage (SB direct) and “x” the SBs coming from the chemical stage (SB indirect). Statistical uncertainties are shown unless they are too small to be seen.



**Figure 10.** Yields of DSB/Gy/Gbp simulated in this work, as a function of the LET of the primary protons. The calculation chain (“This work”) was used with its default parameters except for the selection criterion used to identify strand breaks produced during the physical stage (SB<sub>direct</sub>). Four sets of results were calculated according to four different selection criteria: threshold of 12.5 eV (yellow triangles), threshold of 17.5 eV (blue triangles, default configuration), threshold of 30 eV (red points) and the linear acceptance probability described at the end of section Configuration of the simulations (red squares). Statistical uncertainties are shown unless they are too small to be seen.

the LET are similar for all four selection criteria. In all cases, the number of DSB/Gy/Gbp increases from 2.6 to 20 keV/μm with an increase factor of roughly 1.7, and then remains almost constant from 20 to 47.9 keV/μm. The use of a threshold of 12.5 eV for the energy deposited in the backbone of a nucleotide yields to the highest number of DSB/Gy/Gbp. The yield of DSB/Gy/Gbp starts at 12 for a LET of 2.6 keV/μm and increases to 18 for a LET of 47.9 keV/μm. The threshold of 17.5 eV (default) reduces the number of DSB/Gy/Gbp by half, it starts at 6 for a LET of 2.6 keV/μm and increases to 11 at a LET of 47.9 keV/μm. A further increase of the deposited energy threshold to 30 eV decreases the simulated numbers of DSB/Gy/Gbp to 3.5 at a LET of 2.6 keV/μm and 7 at 47.9 keV/μm. These three sets of results clearly demonstrate the increasing of the deposited energy threshold (physical stage) decreases the total number of simulated DSB/Gy/Gbp (physical and chemical stages) for all the LET considered in this work. Finally, the use of the linear acceptance probability (see the end of section title-configs) produces a higher number of DSB/Gy/Gbp than the results obtained with the default criterion of our calculation chain (threshold of 17.5 eV) but a lower number than computed with an energy threshold of 12.5 eV.



**Figure 11.** Two clusters [(a) and (b)] of strand breaks (red points): clusters (a) is a double strand-break (DSB) and (b) a single-strand break (SSB). Cluster (a) contains only one SB on the bottom DNA strand; when this SB is removed cluster (a) becomes cluster (b).

## Discussion

The comparison between our simulated results, computed in our calculation chain in its default configuration, and the experimental data from the literature (see Fig. 7) shows good agreement between them for the yield of DSB/Gy/Gbp. This agreement is especially good with the data of Frankenberg *et al.*<sup>43</sup> whose experimental conditions were reproduced in our simulations. Nonetheless, we note the shortage of experimental data available in the literature for proton projectiles with energies between 0 and 20 MeV; moreover, the values for those that do exist are quite scattered. This scattering may be due to the different experimental conditions, in particular the use of different type of cells or different minimum fragment size thresholds, but also from the numerous uncertainties associated with the pulsed field gel electrophoresis technique<sup>43</sup>. Nevertheless, one important consideration is that the general agreement of our simulated results with the experimental data does not mean that the simulation reproduced accurately all the processes involved in the creation of early DNA damage. The default configuration in the calculation chain must use some parameters that are adjusted to keep the simulation balanced in terms of the number of DSBs calculated. In the end, the agreement observed in Fig. 7 shows that the set of default parameters chosen in this work is sufficiently relevant to ensure the consistency of the simulation with the experimental data. Moreover, these parameters are chosen within realistic ranges that can be explained or justified. At the same time, the final values of these parameters and the sensitivity of the final results offer evidence of the importance of the particular mechanism involving them.

The numbers of DSB/Gy/Gbp calculated in this work are similar to those calculated with other simulation codes, as illustrated in Fig. 8. Our results are of the same order of magnitude and vary similarly with the LET of the primary protons. This is especially true for the PARTRAC simulation code which uses an approach very close to that used in this work. It should be noted that KURBUC and PARTRAC also have their own sets of parameters and hypotheses that have been adjusted to ensure their consistency. The differences between these parameters is likely to explain the discrepancies between the yields of simulated DSB/Gy/Gbp. For example, the PARTRAC code uses the linear probability as default selection criterion<sup>31</sup> while KURBUC and our calculation chain use a threshold of 17.5 eV<sup>40</sup>. Other elements specific to our simulation may also influence the discrepancies: the use of the physical models of the version 10.1 of Geant4, our DNA geometry (more specifically the hydration shell dimensions) and the chemical reactions considered in the simulation.

Figure 9 demonstrates that 80% of the SB, produced by the simulation of the irradiation of a fibroblast cell nucleus were created during the chemical stage which is slightly higher than the ~70% previously reported<sup>31</sup>. The slight increase of the yield of direct SB with the LET was also not reported previously. As for the DSB yields, the parameters and hypothesis specific to our simulation are likely to explain these discrepancies, especially the use of a 17.5 eV threshold. The fact that 80% of the simulated SB were created during the chemical stage raises questions about the real influence of the  $SB_{\text{direct}}$  selection criterion on the final calculation of the number DSB/Gy/Gbp in our work. That is, the  $SB_{\text{direct}}$  selection criterion influences only the output of the physical stage, whereas the final number of DSB/Gy/Gbp considers the outputs of both the physical and chemical stages. Figure 10 illustrates the significant impact of this selection criterion on our results. The DSB/Gy/Gbp results depend strongly on the selection criterion chosen, even if their variation with the LET remains similar: an increase phase followed by stabilisation of the number of DSB/Gy/Gbp. This means that the choice of this criterion, even though it can influence only 20% of the total number of SBs simulated (the  $SB_{\text{direct}}$ ) can result in very substantial differences in the final number of DSB/Gy/Gbp computed through the simulation of the physical, physicochemical and chemical stages.

The significant influence of the number of  $SB_{\text{direct}}$  on the number of DSB/Gy/Gbp is explained by the DSB determination process (DBScan algorithm), which is conditioned in turn by the creation of a specific type of SB cluster (see section title:ssb\_dsb), one that contains at least one SB on each of the two strands of the DNA. Therefore, it is possible for a cluster identified as a DSB to become an SSB if the SB of an opposite strand disappears (see Fig. 11). Modification of the  $SB_{\text{direct}}$  selection criterion may thus be able to change clusters identified as DSBs into SSBs and to alter substantially the number of simulated DSB/Gy/Gbp.

Note that the simulations performed in this work do not consider the data related to reactions between  $OH^{\bullet}$  and DNA bases, although some of these reactions may lead to the appearance of DNA SBs. The processes involved in this conversion are rather complex<sup>47,48</sup>. Overall, their contribution to the number of SBs is considered low enough to ignore in this work.

## Conclusion

The calculation chain presented in this work is designed to simulate early DNA damage and is the first simulation tool based on Geant4-DNA that is able to fully simulate the physical, physicochemical and chemical stages of irradiation damage at the scale of a human cell nucleus. Extensions of the DNA models included in DnaFabric software were presented and used to create a fibroblast cell nucleus model filled with the content of the human



male genome (diploid cell,  $\sim 6.4 \cdot 10^9$  pairs of nucleotides). The model was then exported to a file for use in Geant4-DNA simulations. A set of those simulations was created and integrated in the calculation chain, making it possible to simulate the physical, physicochemical and chemical stages that follow the irradiation of a cell nucleus. The simulation of these three stages thus took fully into account the fibroblast cell nucleus model previously generated with its  $\sim 6.4 \cdot 10^9$  nucleotide pairs.

Simulations were performed to reproduce the irradiation of the fibroblast cell nucleus by primary protons of different energies (0–20 MeV) in order to compute the resulting yields of DSB/Gy/Gbp. The results were then compared with data from experiments that used pulsed field gel electrophoresis<sup>43–46</sup>. Comparison of the simulated and experimental results required inclusion of constraints related to the low resolution of the experiments. This resulted in setting a minimum DNA fragment size of 10000 nucleotide pairs<sup>43</sup>, so that any DNA fragment smaller than that was considered too small to be experimentally detected. Thus, the simulation identified those fragments but ignored them for the calculation of the final number of DSB/Gy/Gbp. In the end, agreement between our results and the experimental data was good and confirmed the coherence of the calculation chain introduced in this work. Our results were also compared with simulated data obtained with other simulation codes<sup>31,40</sup>. The discrepancies observed between our results and those of the other simulation codes illustrate variations that can result from different parameter adjustments and, specifically, different  $SB_{\text{direct}}$  selection criteria. One selection criterion was shown to influence the number of DSB/Gy/Gbp calculated in our simulation very substantially, although it directly impacts only 20% of all the SBs.

We are currently working at including the reactions of the  $e_{\text{aq}}^-$  hydrogen chemical species and DNA in the simulation. The addition of these reactions together with the use of the data related to the reaction between  $OH^\bullet$  and DNA bases will enable us to introduce base damage in the simulation. Consideration of base damage is required to extend the simulation from the computation of DSBs to the calculation of chromosome aberrations. Furthermore, the addition of recently published DNA cross sections<sup>49,50</sup> in the simulation is ongoing work. Their introduction in the simulation will allow to fill the DNA geometry with a composite material which physical properties are closer to DNA than liquid water in terms of interaction probability and amount of energy deposited. The use of these DNA cross sections together with the latest Geant4 physical models for liquid water<sup>51</sup> will improve the simulation of the physical stage.

The calculation chain created in this work was developed as part of the Geant4 and Geant4-DNA collaborations; the code will be made publicly available in a suitable form for the user community.

## References

- Goodhead, D. Initial events in the cellular effects of ionizing radiations: clustered damage in DNA. *Int. J. Radiat. Biol.* **65**, 7–17 (1994).
- Friedland, W., Jacob, P. & Kundrat, P. Mechanistic simulation of radiation damage to DNA and its repair: on the track towards systems radiation biology modelling. *Radiation Protection Dosimetry* **143**, 542–548 (2010).
- Friedland, W. & Kundrat, P. Track structure based modelling of chromosome aberrations after photon and alpha-particle irradiation. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis* **756**, 213–223 (2013).
- Bernal, M. A. *et al.* Performance of a new atomistic geometrical model of the B-DNA configuration for DNA-radiation interaction simulations. *Journal of Physics: Conference Series* **490**, 012150 (2014).
- Santos, M. D., Villagrasa, C., Clairand, I. & Incerti, S. Influence of the chromatin density on the number of direct clustered damages calculated for proton and alpha irradiations using a Monte Carlo code. *Progress in Nuclear Science and Technology* **4**, 449–453 (2014).
- Nikjoo, H., Taleei, R., Liamsuwan, T., Liljequist, D. & Emfietzoglou, D. Perspectives in radiation biophysics: From radiation track structure simulation to mechanistic models of DNA damage and repair. *Radiation Physics and Chemistry* **128**, 3–10 (2016).
- Nikjoo, H., Uehara, S., Emfietzoglou, D. & Cucinotta, F. Track-structure codes in radiation research. *Radiation Measurements* **41**, 1052–1074 (2006).
- Lazarakis, P. *et al.* Comparison of nanodosimetric parameters of track structure calculated by the Monte Carlo codes Geant4-DNA and PTr. *Physics in Medicine and Biology* **57**, 1231 (2012).
- Incerti, S. *et al.* The Geant4-DNA project. *International Journal of Modeling, Simulation, and Scientific Computing* **1**, 157–178 (2010).
- Bernal, M. *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. *Physica Medica* **31**, 861–874 (2015).
- Karamitros, M. *et al.* Diffusion-controlled reactions modeling in Geant4-DNA. *Journal of Computational Physics* **274**, 841–882 (2014).
- Incerti, S. *et al.* Comparison of GEANT4 very low energy cross section models with experimental data in water. *Medical Physics* **37**, 4692–4708 (2010).
- Allison, J. *et al.* Geant4 developments and applications. *Nuclear Science, IEEE Transactions on* **53**, 270–278 (2006).
- Garty, G. *et al.* A nanodosimetric model of radiation-induced clustered DNA damage yields. *Physics in medicine and biology* **55**, 761 (2010).
- Bug, M. U., Surdutovich, E., Rabus, H., Rosenfeld, A. B. & Solov'Yov, A. V. Nanoscale characterization of ion tracks: MC simulations versus analytical approach (2012).
- Friedland, W., Dingfelder, M., Kundrat, P. & Jacob, P. Track structures, DNA targets and radiation effects in the biophysical Monte Carlo simulation code PARTRAC. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* **711**, 28–40 (2011).
- Meylan, S., Vimont, U., Incerti, S., Clairand, I. & Villagrasa, C. Geant4-DNA simulations using complex DNA geometries generated by the DnaFabric tool. *Computer Physics Communications* **204**, 159–169 (2016).
- Biljecki, F., Ledoux, H., Stoter, J. & Zhao, J. Formalisation of the level of detail in 3d city modelling. *Computers, Environment and Urban Systems* **48**, 1–15 (2014).
- Swarts, S. G., Sevilla, M. D., Becker, D., Tokar, C. J. & Wheeler, K. T. Radiation-Induced DNA Damage as a Function of Hydration: I. Release of Unaltered Bases. *Radiation Research* **129**, 333 (1992).
- Bernal, M. *et al.* An atomistic geometrical model of the B-DNA configuration for DNA–radiation interaction simulations. *Computer Physics Communications* **184**, 2840–2847 (2013).
- Leslie, A., Arnott, S., Chandrasekaran, R. & Ratliff, R. Polymorphism of DNA double helices. *J. Mol. Biol.* **143**, 49–72 (1980).
- Kreth, G., Finsterle, J., Von Hase, J., Cremer, M. & Cremer, C. Radial arrangement of chromosome territories in human cell nuclei: a computer model approach based on gene density indicates a probabilistic global positioning code. *Biophysical Journal* **86**, 2803–2812 (2004).
- Brun, R. & Rademakers, F. ROOT—an object oriented data analysis framework. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **389**, 81–86 (1997).
- Karamitros, M. *et al.* Modeling radiation chemistry in the Geant4 toolkit. *Progress in Nuclear Science and Technology* **2**, 503–508 (2011).
- Milligan, J. R., Wu, C. C. L., Aguilera, J. A. & Ward, J. F. Characterization of the reaction rate coefficient of DNA with the hydroxyl radical. *Radiation research* **146**, 510–513 (1996).

26. Roots, R. & Okada, S. Estimation of Life Times and Diffusion Distances of Radicals Involved in X-Ray-Induced DNA Strand Breaks or Killing of Mammalian Cells. *Radiation Research* **64**, 306–320 (1975).
27. Aydogan, B., Bolch, W. E., Swarts, S. G., Turner, J. E. & Marshall, D. T. Monte Carlo Simulations of Site-Specific Radical Attack to DNA Bases. *Radiation Research* **169**, 223–231 (2008).
28. Kreipl, M. S., Friedland, W. & Paretzke, H. G. Interaction of ion tracks in spatial and temporal proximity. *Radiation and Environmental Biophysics* **48**, 349–359 (2009).
29. Buxton, G. V., Greenstock, C. L., Helman, W. P. & Ross, A. B. Critical Review of rate constants for reactions of hydrated electrons, hydrogen atoms and hydroxyl radical in Aqueous Solution. *J. Phys. Chem* **17** (1988).
30. Abolfath, R. M., Carlson, D. J., Chen, Z. J. & Nath, R. A molecular dynamics simulation of DNA damage induction by ionizing radiation. *Physics in Medicine and Biology* **58**, 7143–7157 (2013).
31. Friedland, W., Jacob, P., Bernhardt, P., Paretzke, H. G. & Dingfelder, M. Simulation of DNA damage after proton irradiation. *Radiation Research* **159**, 401–410 (2003).
32. Swarts, S. G. *et al.* Mechanisms of Direct Radiation Damage in DNA, Based on a Study of the Yields of Base Damage, Deoxyribose Damage, and Trapped Radicals in d(GCAGCGTGC). *Radiation Research* **168**, 367–381 (2007).
33. Celiberto, R., Janev, R., Wadehra, J. & Tennyson, J. Dissociative electron attachment to vibrationally excited H<sub>2</sub> molecules involving the resonant Rydberg electronic state. *Chemical Physics* **398**, 206–213 (2012).
34. Laporta, V., Celiberto, R. & Tennyson, J. Dissociative electron attachment and electron-impact resonant dissociation of vibrationally excited O<sub>2</sub> molecules. *Physical Review A* **91** (2015).
35. Bald, I., Illenberger, E. & Kopyra, J. Damage of DNA by Low Energy Electrons. *Journal of Physics: Conference Series* **373**, 012008 (2012).
36. Boudaiffa, B., Cloutier, P., Hunting, D., Huels, M. A. & Sanche, L. Resonant formation of DNA strand breaks by low-energy (3 to 20 eV) electrons. *Science* **287**, 1658–1660 (2000).
37. Nikjoo, H., O'Neill, P., Goodhead, D. & Terrisol, M. Computational modelling of low-energy electron-induced DNA damage by early physical and chemical events. *International Journal of Radiation Biology* **71**, 467–483 (1997).
38. Lobachevsky, P. N. & Martin, R. F. DNA Strand Breakage by <sup>125</sup>I-Decay in a Synthetic Oligodeoxynucleotide: Quantitative analysis of fragment distribution. *Acta Oncologica* **35**, 809–815 (1996).
39. Prise, K. M. *et al.* Critical energies for SSB and DSB induction in plasmid DNA by low-energy photons: action spectra for strand-break induction in plasmid DNA irradiated in vacuum. *International journal of radiation biology* **76**, 881–890 (2000).
40. Nikjoo, H., O'Neill, P., Wilson, W. E. & Goodhead, D. T. Computational approach for determining the spectrum of DNA damage induced by ionizing radiation. *Radiation Research* **156**, 577–583 (2001).
41. Balasubramanian, B., Pogozelski, W. K. & Tullius, T. D. DNA strand breaking by the hydroxyl radical is governed by the accessible surface areas of the hydrogen atoms of the DNA backbone. *Proceedings of the National Academy of Sciences* **95**, 9738–9743 (1998).
42. Dizdaroğlu, M. & Jaruga, P. Mechanisms of free radical-induced damage to DNA. *Free Radical Research* **46**, 382–419 (2012).
43. Frankenberg, D. *et al.* Induction of DNA Double-Strand Breaks by 1 h and 4 h e Ions in Primary Human Skin Fibroblasts in the LET Range of 8 to 124 keV/um. *Radiation Research* **151**, 540–549 (1999).
44. Campa, A. *et al.* DNA DSB induced in human cells by charged particles and gamma rays: Experimental results and theoretical approaches. *International Journal of Radiation Biology* **81**, 841–854 (2005).
45. Belli, M. *et al.* DNA DSB induction and rejoining in V79 cells irradiated with light ions: a constant field gel electrophoresis study. *International Journal of Radiation Biology* **76**, 1095–1104 (2000).
46. Belli, M. *et al.* DNA Fragmentation in mammalian cells exposed to various light ions. *Advances in Space Research* **27**, 393–399 (2001).
47. Cadet, J. & Wagner, J. R. DNA Base Damage by Reactive Oxygen Species, Oxidizing Agents, and UV Radiation. *Cold Spring Harbor Perspectives in Biology* **5** (2013).
48. Bauer, N. C., Corbett, A. H. & Doetsch, P. W. The current state of eukaryotic DNA base damage and repair. *Nucleic Acids Research* **43**, 10083–10101 (2015).
49. Bug, M. U. *et al.* An electron-impact cross section data set (10eV–1keV) of DNA constituents based on consistent experimental data: A requisite for Monte Carlo simulations. *Radiation Physics and Chemistry* **130**, 459–479 (2017).
50. Bordage, M. *et al.* Implementation of new physics models for low energy electrons in liquid water in Geant4-DNA. *Physica Medica* **32**, 1833–1840 (2016).
51. Kyriakou, I., Šeňl, M., Nourry, V. & Incerti, S. The impact of new Geant4-DNA cross section models on electron track structure simulations in liquid water. *Journal of Applied Physics* **119**, 194902 (2016).

## Author Contributions

S.M. developed the simulation, analysed the results and drafted the manuscript based on discussions with C.V., S.I. and I.C.; C.V. contributed to the simulation design and analysis of the results; S.I. contributed to the developments of Geant4-DNA physics models and coordinated the Geant4-DNA project; M.K. contributed to the implementation of the chemical stage in the simulation; M.B. and N.T. contributed to the analysis of the results. All authors reviewed and approved the final manuscript. S.M. was employed by the IRSN from 2013 to 2017. C.V., N.T. and I.C. are still employed by the IRSN. S.M., C.V., S.I., I.C. and M.K. are members of the Geant4 and Geant4-DNA collaborations.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-017-11851-4>.

**Competing Interests:** The authors declare that they have no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017