Universidade Estadual de Campinas – Unicamp

Instituto de Física "Gleb Wataghin"

Liset de la Fuente Rosales

A Monte Carlo Study of the Direct and Indirect DNA Damage Induced by Ionizing Radiation.

Estudo do Dano Direto e Indireto Induzido ao DNA pela Radiação Ionizante Usando o Método de Monte Carlo

Campinas

2018

Universidade Estadual de Campinas

Instituto de Física "Gleb Wataghin"

Liset de la Fuente Rosales

A Monte Carlo Study of the Direct and Indirect DNA Damage Induced by Ionizing Radiation.

Estudo do Dano Direto e Indireto Induzido ao DNA pela Radiação Ionizante

Usando o Método de Monte Carlo

Tese apresentada ao Instituto de Física "Gleb Wataghin" como parte dos requisitos exigidos para obtenção do título de Doutora em Ciências

Dissertation presented to the Institute of Physics "Gleb Wataghin" of the University of Campinas in partial fulfillment of the requirements for the degree of Doctor in Science

Supervisor/Orientador: Mario Antonio Bernal Rodriguez

Este exemplar corresponde à versão final da Tese defendida pela aluna Liset de la Fuente Rosales, e orientada pelo Prof. Dr. Mario Antonio Bernal Rodriguez

Campinas

2018

Agência(s) de fomento e nº(s) de processo(s): CNPq, 190154/2013-6

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca do Instituto de Física Gleb Wataghin
Maria Graciele Trevisan - CRB 8/7450

F952m

Fuente Rosales, Liset de la, 1986-
A Monte Carlo study of the direct and indirect DNA damage induced by ionizing radiation / Liset de la Fuente Rosales. – Campinas, SP : [s.n.], 2018.

Orientador: Mario Antonio Bernal Rodriguez.
Tese (doutorado) – Universidade Estadual de Campinas, Instituto de Física Gleb Wataghin.

1. Dano ao DNA. 2. DNA. 3. Radiações ionizantes. 4. Íons pesados. 5. Monte Carlo, Método de. I. Bernal Rodriguez, Mario Antonio, 1972-. II. Universidade Estadual de Campinas. Instituto de Física Gleb Wataghin. III. Título.

Informações para Biblioteca Digital

Título em outro idioma: Estudo do dano direto e indireto induzido ao DNA pela radiação ionizante usando o método de Monte Carlo
Palavras-chave em inglês:
DNA damage
DNA
Ionizing radiation
Heavy ions
Monte Carlo method
Área de concentração: Física
Titulação: Doutora em Ciências
Banca examinadora:
Mario Antonio Bernal Rodriguez
Sandro Guedes de Oliveira
Jean Rinkel
Elisabeth Mateus Yoshimura
Paulo Roberto Costa
Data de defesa: 28-06-2018
Programa de Pós-Graduação: Física

MEMBROS DA COMISSÃO JULGADORA DA TESE DE DOUTORADO DE **LISET DE LA FUENTE ROSALES - RA: 161025** APRESENTADA E APROVADA AO INSTITUTO DE FÍSICA “GLEB WATAGHIN”, DA UNIVERSIDADE ESTADUAL DE CAMPINAS, EM 28/06/2018.

COMISSÃO JULGADORA:

- Prof. Dr. Mario Antonio Bernal Rodriguez - (Orientador) - IFGW/UNICAMP
- Prof. Dr. Jean Rinkel - IFGW/UNICAMP
- Prof. Dr. Sandro Guedes de Oliveira - IFGW/UNICAMP
- Prof. Dr. Elisabeth Mateus Yoshimura – INSTITUTO DE FÍSICA - USP
- Prof. Dr. Paulo Roberto Costa – UNIVERSIDADE DE SÃO PAULO - USP

A Ata de Defesa, assinada pelos membros da Comissão Examinadora, consta no processo de vida acadêmica do aluno.

CAMPINAS
2018

To my family.

Acknowledgements

First, I would like to extend my sincerest thanks to my supervisor Dr. Mario Bernal Rodriguez and his family for their support, kindness, and friendship during these four years. I appreciate that he agreed to be my mentor. It has been an honor to work with him. I have learned a lot and gained more experience as a researcher.

Thanks to my family, I couldn't have reached this far without them. I am a result of my parents love, education, and efforts and I am trying harder to be a good example to my dear brother who I am sure is going to be a successful scientist. Also, I would like to mention my grandparents, my godmother, my sister in law and all my relatives. Special thanks to Ibis and Ivan and to their respective families.

I also want to thank the Brazilian National Council for Scientific and Technological Development (CNPq) and the Academy of Sciences for the Third World (TWAS) for funding this research and giving me the opportunity to study in this great university, specially in the "Gleb Wataghin" Institute of Physics. I am very grateful for this chance and to all the professors and colleagues that I have met, not only in our research group but also in the Applied Physics Department.

Among all colleagues and friends at work, I can not forget to mention Prof. Alessandra Tomal and Prof. Sandro Guedes who have been aware of the development of this work and gave important advice to improve it. I also want to mention the Surface Physics Group and Prof. Richard Landers, who shared not only coffee but also interesting talks.

I want to recognize the support of the GEANT4-DNA collaboration team, specially Dr. Sebástien Incerti, Dr. Mathieu Karamitros, and Dr. Ziad Francis. Their advice were important to carry out this research.

Also, the collaboration with Dr. Nadia Falzone, from Oxford University, and Dr. Nicole Ackerman, from Agnes Scott College, for the amazing experience of working together and all the knowledge gained from them.

During these years, I have met amazing friends with whom I have shared good and hard times together. Among them, I want to mention Rodrigo Seraide, Rosanne,

Erina, Andreza, Denis, Azam, Rita, Edson, and Idalicio. My special thanks to Teresinha Fazan and her sisters for being my Brazilian family. Also, to the friends I have made at home: Carol, Clarinha, Ingrid, Mariana, and Michelle. It has been a real pleasure meeting you guys.

Of great importance are my friends and previous colleagues from my home country, Dr. Rodolfo Alfonso, Dr. Eduardo Larrinaga, MSc. Ileana Silvestre, MSc. Yudy Ascencion, Misleidy Napoles, MSc. Michel Rodriguez, MSc. Rogelio Diaz, Grechel Freixas, and many others who have influenced my professional career and personal life.

Finally, I want to thank my partner in life, Fernando, for being my complement, for encouraging me, and for all his love. I am a very lucky woman because I have you.

I am sure there are many others that I did not mention but to all I am grateful.

Resumo

A simulação por Monte Carlo (MC) é uma poderosa ferramenta para estudar os efeitos biológicos induzidos pela radiação ionizante em seres vivos. Vários códigos MC, com diferentes níveis de complexidade, são comumente usados em campos de pesquisa como nanodosimetria, radioterapia, proteção radiológica e areoespacial. Este trabalho apresenta uma ampliação de um modelo existente para fins radiobiológicos, a fim de levar em conta o dano indireto e mixto induzido no DNA por radiações ionizantes. O kit de ferramentas de simulação GEANT4-DNA foi usado para simular a etapa física, pré-química e química do dano inicial no DNA induzidos por prótons e partículas alfa. O meio usado nas simulações foi a água líquida. Foram gerados dois arquivos de saída, um contendo eventos de deposição de energia dentro da região de interesse (ROI), e outro com a posição das espécies químicas produzidas pela radiólise d'água, de 0,1 ps até 1 ns. As informações contidas nos dois arquivos foram sobrepostas em um modelo de material genético com resolução atômica, consistindo de várias cópias de fibras de cromatina de 30 nm. A configuração do B-DNA foi usada. O foco deste trabalho foi o dano indireto produzido pelo ataque do radical hidroxilo ($\bullet OH$) ao grupo açucar-fosfato, normalmente através da abstração do hidrogênio. A abordagem seguida para explicar o dano indireto no DNA foi o mesmo usado por outros códigos radiobiológicos. O parâmetro crítico aqui considerado foi o raio de reação, calculado a partir da equação de difusão de Smoluchowski. Os rendimentos de quebra de cadeia simples, dupla e total produzidos por mecanismos diretos, indiretos e mistos são relatados. Resultados consistentes com outros trabalhos de simulações e experimentais foram encontrados, mesmo sem seguir qualquer processo de ajuste. Até aonde nós sabemos, esta é a primeira vez que o código GEANT4-DNA é combinado com um modelo atômico do DNA para estudar o dano químico induzido por radiações ionizantes.

Palavras-chaves:dano indireto, DNA, nanodosimetria, GEANT4-DNA, Monte Carlo, prótons, partículas alfas

Abstract

Monte Carlo (MC) simulation is a powerful tool to study biological effects induced by ionizing radiation on living beings. Several MC codes, with different level of complexity, are commonly used in research fields such as nanodosimetry, radiotherapy, radiation protection, and space radiation. This work presents an enhancement of an existing model for radiobiological purposes, in order to account for the indirect and mixed DNA damage induced by ionizing particles. The GEANT4-DNA simulation toolkit was used to simulate physical, pre-chemical and chemical stages of the early DNA damage induced by protons and alpha particles. Liquid water was used as the medium for simulations. Two phase-space files were generated, one containing energy deposition events inside the region of interest (ROI), and another one with the position of chemical species produced by water radiolysis from 0.1 ps up to 1 ns. The information contained in both files was superposed on a genetic material model with atomic-resolution, consisting of several copies of 30-nm chromatin fibers. The B-DNA configuration was used. This work focused on the indirect damage produced by the hydroxyl radical ($\bullet OH$) attack on the sugar-phosphate, normally through hydrogen abstraction. The approach followed to account for the indirect damage in DNA was the same used by other radiobiological codes. The critical parameter considered here was the reaction radius, which was calculated from the Smoluchowski's diffusion equation. Single, double, and total strand break yields produced by direct, indirect and mixed mechanisms are reported. Results consistent with simulated and experimental works were found, even without following any fitting process. To the best of our knowledge, this is the first time the GEANT4-DNA code is used in conjunction with a DNA atomic resolution model for studying the chemical damage induced by ionizing radiations.

Key-words: indirect damage, DNA, nanodosimetry, GEANT4-DNA, Monte Carlo, protons, alpha particles.

List of Figures

List of Tables

List of abbreviations and acronyms

LNT	Linear non-threshold
SPE	Solar particle events
GCR	Galactic cosmic rays
RBE	Relative biological effectiveness
LET	Linear energy transfer
DNA	Deoxyribonucleic acid
A-	adenine
T-	Thymine
C-	Cytosine
G-	Guanine
U-	Uracil
RNA	Ribonucleic acid
IR	Ionizing radiation
TUNEL	Deoxynucleotidyl transferase dUTP nick end labeling assay
ELISA	Enzyme-linked immunosorbent assay
bp	base pair
SSB	Single strand breaks
DSB	Double strand break
G1	Gap phase 1 of cell cycle
S	DNA synthesis of cell cycle
G2	Gap phase 2 of cell cycle

M	Mitosis phase of cell cycle
NHEJ	Non-homologous end-joining
HR	Homologous recombination
DNA-Pkcs	DNA-dependent protein kinase catalytic subunits
MRN	Protein complex Mre11/Rad50/Nbs1
RPA	Replication protein A
ssDNA	Single stranded DNA
PFGE	Gel electrophoresis
gH2AX	Quantifying gamma histone H2A
MC	Monte Carlo
TSC	Track structure codes
SB	Strand break
PDB	Protein Data Bank
HZE	Heavy ions of high charge and energy
BEB	Binary encounter Bethe
IAM	Independent atom model
OER	Oxygen Enhancement Ratio
ASCII	Human-readable format
MPI	Message Passing Interface
MT	Multithread
PBS	Portable Batch System
GFRMd	Física Radiológica Médica research group
ROI	Region of interest

SSB+	Complex SSB yield
SHP	Site-Hit Probability
G(t)	Time dependent radiolytic yield
SSBY	Single strand break yield
DSBY	Double strand break yield
TSBY	Total strand break yield
RSD	Relative standard deviation
RBE	Relative Biological Effectiveness

Contents

1 Introduction

1.1 Rationale

In the late 1800s, several important findings occurred: the X-rays by Roentgen in 1895, the radioactivity by Becquerel in 1896, and the radioactive elements polonium and radium by Pierre and Marie Curie in 1898. Those discoveries were followed by a prompt use of ionizing radiation for diagnosing and treating malignant and non-malignant diseases, based on the advantages they offered and completely ignoring their biological effects. The first radiographic studies "accidentally" led to the use of this type of radiation as an apparent cure of skin cancer. The first steps in radiation therapy, which is currently a well established clinical practice, did not consider the extent of the radiation biological effects and harmful side effects did not take so long to appear.

Nowadays, it is known that living beings in general are being unavoidably exposed to ionizing radiations, regardless of the source they come from: natural (cosmic rays radiation and radon gas emanation from rocks on earth) or artificial (human activities, medical practices to diagnose and treat diseases, radioactive material resulting from nuclear weapons testing, energy generation, nuclear power plants accidents like Chernobyl-1986 or Fukushima-2011) [5]. Hence the need to understand and prevent ionizing radiation effects.

The origin of the branch of science concerned with the effects induced by ionizing radiation in living beings, known as Radiobiology, can be related to Leopold Freund's experiments in 1896 using X-rays, although only in 1927 Hermann Joseph Muller published experimental data showing radiation-induced genetic effects. As a matter of fact, fractionated schemes during conventional cancer treatments, typically from a 1.8 to 2 Gy daily dose, were suggested by several early radiobiological experiments.

The accumulated experience about ionizing radiation effects comes mainly from epidemiological studies, experimental assays, occupational exposure, and cancer treatments. In radiation protection, to estimate the risk of induced cancer and hereditary effects, a linear non-threshold (LNT) relationship between the dose and the effect is assumed. However, epidemiological data has large uncertainties at low doses and the LNT

hypothesis may not be adequate to estimate the health risk [6, 7, 8].

During manned space flights, the crew is exposed to a high level of ionizing radiations and most of them comes from radiation types that are not experienced on earth, like protons or electrons coming from the solar wind, solar particle events (SPE), and galactic cosmic rays (GCR) [9]. Furthermore, there is a huge amount and variety of secondary particles resultant from the interaction of primary particles with spacecraft or human tissues. For instance, relativistic iron ions coming from GCR produce a huge number of ionizations in the medium. Since there is no experience dealing with this type of radiation field on earth, estimating radiation risks in this case is difficult and the uncertainties involved are very high, which is a major limitation to the length of space missions and the amount of shielding that should be used.

Nowadays, the use of high energy protons and heavy ions for radiotherapy treatments is expanding abroad. For such radiation beams, the same biological effect occurs for a different value of absorbed dose when compared to a high energy photon beam. The current methodology is to quantify the delivered dose in ion beam therapy by means of the relative biological effectiveness (RBE) [11]. There is an international concern about the different methods used to derive the RBE, so the focus is to join all efforts to establish a generic consensus [12]. As a consequence, there is a lot of on-going researches to introduce new dosimetric quantities, which allow discriminate physical processes from the biological ones. Moreover, the most effective screening method for early stage breast cancer is mammography, which uses low energy photon beams. In this case, it should be considered that the RBE increases when the photon energy decreases [13].

The passage of the ionizing radiation through biological targets is expressed in the form of tracks of charged particles which leave stochastic patterns of ionizations and excitations on the irradiated material. If these energy deposition events are described with conventional average macroscopic quantities, their contribution seems to be small. However, at the microscopic scale (cellular and subcellular dimensions) this contribution can be significantly large. In several cases, the use of the well-known absorbed dose quantity is not adequate [14]. For instance, during the irradiation of tissues and organs with short range charged particles or low energy photons from incorporated radionuclides, the use of high linear energy transfer (LET) radiations, low energy microbeams, and also the exposure at low absorbed dose levels. The heterogeneity observed in the energy deposition

for a given radiation field critically depends on the size of the site we are studying. It has been observed that not only those sites with energy deposits inside them may produce biological changes but also other sites around or near without energy deposition events may also exhibit biological responses.

Because of their small size, biomolecules are potential targets for this random and discrete clusters of energy deposition events along charged particle tracks, which can induce a measurable biological effect or not, that will depend on several factors: probability of the specific molecule being hit by the ionizing radiation, abundance of this molecule in the cell, its importance for the cell healthy functioning, how many copies of this molecule are present inside the cell and cell capacity to react to the loss of working copies, importance of the cell to the structure or function of the corresponding tissue or organ, etc. Early experiments with mammalian systems showed that the principal sensitive sites for radiation induced biologic effects, including cell killing, carcinogenesis and mutation, are inside the cell nucleus [15]. Specifically, the deoxyribonucleic acid (DNA) is considered the most critical target inside the cell nucleus [16, 17, 18].

Three different approaches to the radiobiological problem can be pointed out: experimental [13, 19, 20, 21], theoretical, and computational [22, 23, 24, 25, 26, 27, 28]. The theoretical approach is based on mathematical models [29, 30, 31] and the computational one is known as radiation biophysics [32] or computational radiobiology. In general, the computational approach has as main goal to characterize the primary properties of radiation tracks or microdosimetric quantities. This is extremely useful specially in cases where experiments are not feasible. If the aim is to study the energy deposition pattern for different radiation qualities, it is necessary to describe the charged particle track structure. With the considerable increase of the speed and general capabilities of modern computers, it is possible to use more physically realistic stochastic models. General-purpose Monte Carlo codes [33, 34, 35, 36, 37] are not suitable for the detailed description of the interaction of ionizing radiation with matter at the nanometric scale, mainly because they normally use a condensed history method grouping several collisions in one single step. The need of such detailed description of particle tracks gave rise to specific-purpose Monte Carlo codes, referred to as track structure codes (TSC) [27, 38, 39, 40, 41]. These codes allow to calculate the energy deposited at the nanometric scale modeling particle tracks "step-by-step" in gaseous media or liquid water. The strength of these TSC, when they

are capable of simulating the DNA damage caused by radiation tracks, lies in their genomic resolved models when compared to experimental techniques for the same purpose. Indeed, experiments are not able to resolve the DNA damage at a single base-pair (bp) level. The most common experimental method for DNA damage measurements, pulsed field gel electrophoresis (PFGE), has a resolution of about 1 Mbp. Nowadays, TSC is the only method capable to account for the DNA damage at a single base-pair.

In 2001, a new project was initiated by Prof. P. Nieminen at the European Space Agency with the purpose of extending the GEANT4 toolkit capabilities in order to estimate the biological effects of ionizing radiation in the perspective of future space manned missions. As a result of what is known as the GEANT4-DNA collaboration, a set of physics processes for microdosimetry in liquid water capable to track electrons down to the eV scale was firstly delivered in 2007 [42]. So far, several improvements have been introduced and an open source toolkit named GEANT4-DNA is now available. The last version of the GEANT4-DNA package provides a complete set of models for describing event-by-event electromagnetic interactions of charged particles with liquid water and also takes into account the water radiolysis and the corresponding chemical species transport [43].

As part of the GEANT4-DNA collaboration group, the Medical Radiological Physics Group (GFRMd) of the Institute of Physics "Gleb Wathaghin" (IFGW) has been working for several years in developing biophysical models to validate the GEANT4-DNA package for Radiobiology research, with special focus on the estimation of early biological effects induced by ionizing radiations [1, 24, 25]. These works included the development of geometrical models of the human genetic material for three DNA configurations (B, A and Z) and the estimation of the direct DNA damage induced by photons and heavy charged particles. After the impact of ionizing radiation on biological systems, several processes take place which are classified according to the corresponding time scale: physical, chemical, and biological. This enforced the idea that only studying the direct DNA damage during the physical stage does not completely describes the early damage to this molecule.

The first aspect to be considered when modelling early DNA damage is the geometrical model that is going to be used. The models that have been developed so far differ on their organizational level. The most simple model usually use geometrical shapes like

cylinders to represent a DNA segment, not considering the atomic composition. A more detailed description is the volume model, where the DNA is divided into slices and each slice is placed into volumes representing the nucleobases and the sugar-phosphate backbone. Some sophisticated DNA models with atomic resolution are also available, but their use in track structure codes is limited to a few examples [1, 27]. Recent works are implementing more realistic human cells using those DNA models as base unit [26, 27, 28, 44, 45, 46]. Particularly, the PARTRAC suite uses a DNA model with a full atomic description to build complex target geometries such as human fibroblasts [27] or lymphocytes [47]. Those complex geometries combined with the description of the three initial stages (e.g. physical, physico-chemical, and chemical) after ionizing radiation exposure allow to account for the radiation induced DNA damage. Other recent works performed by Meylan et al. [28] and Lampe et al [46] also implemented realistic cell geometries using a nucleotide pair as the base unit for the DNA model without describing the atomic composition of the bases and sugar-phosphate backbone. Meylan et al. modelled a fibroblast nucleus using the DnaFabric software [48], which was extended in order to provide a new simulation chain based on GEANT4-DNA to estimate early DNA damage. The work presented by Lampe et al. also took advantages of the GEANT4-DNA capabilities to study the impact of background radiation on a bacterial genome.

The geometrical model can be implemented either in the track structure code or in an independent radiobiology oriented software, where particle tracks are superimposed on the DNA model. With this model and the track structure code able to simulate the processes or stages mentioned above, it is possible at least to extract all the information about ionizations, excitations, and the chemical species resultant from the water radiolysis inside the ROI.

The main objective of this thesis is to estimate the early biological damage yields caused by ionizing radiations, including the indirect damage, taking advantage of the GEANT4-DNA package capabilities [43, 49, 50, 51] and the existing geometrical model with atomic resolution developed by Bernal et al.[1]. The GEANT4-DNA track structure toolkit was used. In addition, a home made code developed for radiobiology purposes was enhanced for also predicting the indirect and mixed damage induced in the DNA by heavy charged particles. It should be remarked that, unless the other TCS, GEANT4-DNA is an open source code so our work can be accessed by anyone. To the best of our knowledge,

it is the first time the GEANT4-DNA code is used in conjunction with a DNA atomic resolution model for studying the chemical damage induced by ionizing radiations.

1.2 Hyphotesis

"It is possible to develop a new biophysical model able to account for the indirect damage to DNA with atomic resolution which combined with the GEANT4-DNA toolkit allows to estimate early biological damage caused by ionizing radiation. The feasibility of this model can be benchmarked against simulations and experimental measurements performed by other authors."

1.3 Objectives

Main: "To develop a new biophysical model to consider the indirect damage to the DNA molecule induced by chemical species resulting from water molecule dissociation."

Secondary objectives:

- To study the production and transport of chemical species resulting from water molecule dissociation with the GEANT4-DNA toolkit.
- Development of a new biophysical model to account for the indirect damage induced by chemical species.
- Validation of the new biophysical model by determining indirect DNA damage yields after irradiation with energetic ions (protons and α-particles).
- To support the GEANT4-DNA project during the development of the Monte Carlo code for radiation transport simulation and the generation and transport of chemical species.

1.4 Thesis structure

This thesis was organized in four chapters and the conclusions. Chapter 1 provides an introduction to the research topic, emphasizing the importance, relevance, and

application of it. This chapter also identifies the main and secondary objectives pursued in this research. Chapter 2 gives a theoretical background of the subject, mainly in Radiobiology and the use of the Monte Carlo method applied to radiobiological studies. Also, a state of the art of the topic specially focused on track structure codes and details about biophysical models available is presented. Chapter 3 presents a detailed description of the codes developed in this work, all the validation tests carried out, a description about the geometry employed, the irradiation setups, and the simulations performed. In this chapter, reference is also made to the implementation of parallel programming and the computational resources available to perform the simulations. Chapter 4 shows the results obtained in correspondence with each task explained in the Methods. Finally, the conclusions are drawn with some suggestions on how to improve the current version of the code.

2 Theoretical background

2.1 Radiobiology. Interaction of ionizing radiation with biological systems

In the introduction, it was mentioned that the most important feature of the interaction of ionizing radiation with matter is the random and discrete (to a good approximation) nature of the energy deposition. The major contribution to the energy deposited inside a microscopic biological target is due to secondary electrons. Three main groups, taken into account their initial energy have been identified: spurs (100 eV or less), blobs (100-500 eV), and short tracks (500-5000 eV). A representation of those groups is shown in Fig. 1.

The fact that two different types of radiation deliver to the medium the same amount of energy imparted per unit mass does not necessarily mean that they produce equivalent biological effects. The microdosimetric pattern of energy deposition is what determines the biological effectiveness. To be able to compare these energy deposition events for different types of radiation, the concept of linear energy transfer (LET) was introduced by the International Commission of Radiation Units and Measurements (ICRU) [53]. The LET ($keV/\mu m$) is defined as the average energy transferred to electrons and lo-

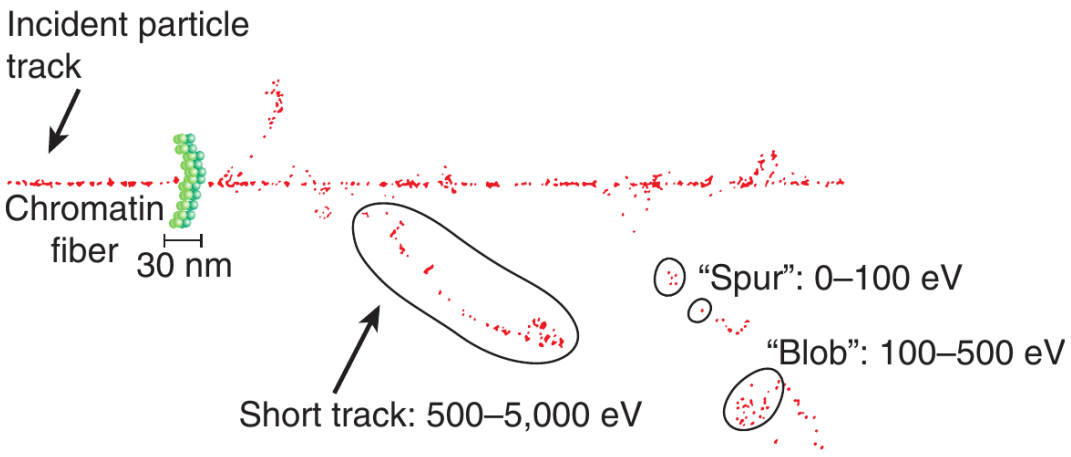

Figure 1 – Alpha particle track though an absorbing medium, illustrating the random and discrete energy depositions events along the track. A segment of chromatin is also shown approximately to scale. Source: [52]

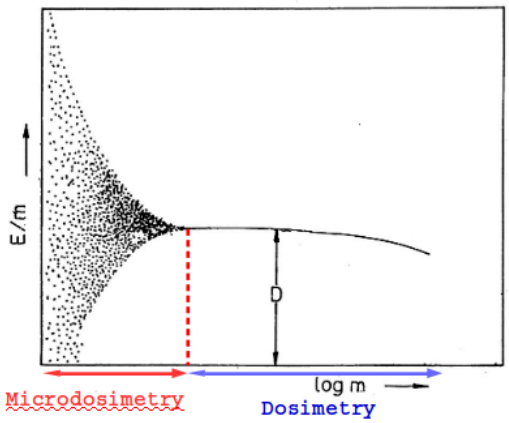

Figure 2 – Stochastic character of the energy deposition process. Source: [54]

cally imparted to the medium by a charged particle of specific energy per unit path-length:

$$LET = \frac{\Sigma_i dE_i}{\Sigma_i dx_i} \tag{2.1}$$

where dE_i is the energy lost by the particle along step dx_i.

The importance of the stochastic character of the energy deposition along a radiation track is shown in Fig. 2. The energy deposited per unit mass (m) is plotted against the mass of regions with decreasing volume centered at the same point in an uniform medium and subjected to several irradiations of equal duration. For large volumes, absorbed dose converge to a given value, while as the site volume decreases absorbed dose fluctuation increases. Thus, for large enough volumes, the concept of absorbed dose is a good representation of the deposited energy, a situation that belongs to the macrodosimetry domain. For smaller volumes, the randomness of energy deposition becomes apparent, which belongs to the microdosimetry domain.

All biomolecules are susceptible to being hit by energy deposition "clusters". However, the most critical target is the DNA because of the role it plays in the correct functioning of the cell and in the transmission of the genetic information to the progeny.

The damage induced by ionizing radiation (IR) in DNA is classified as direct and indirect (see Fig. 3). The direct damage refers to the direct action of primary or secondary particles on biological molecules (DNA, etc.), done through ionizations and excitations. As a result, one or more chemical bonds may be broken, leaving atoms or molecules with unpaired electrons, which are very reactive and have short lifetime. The resulting

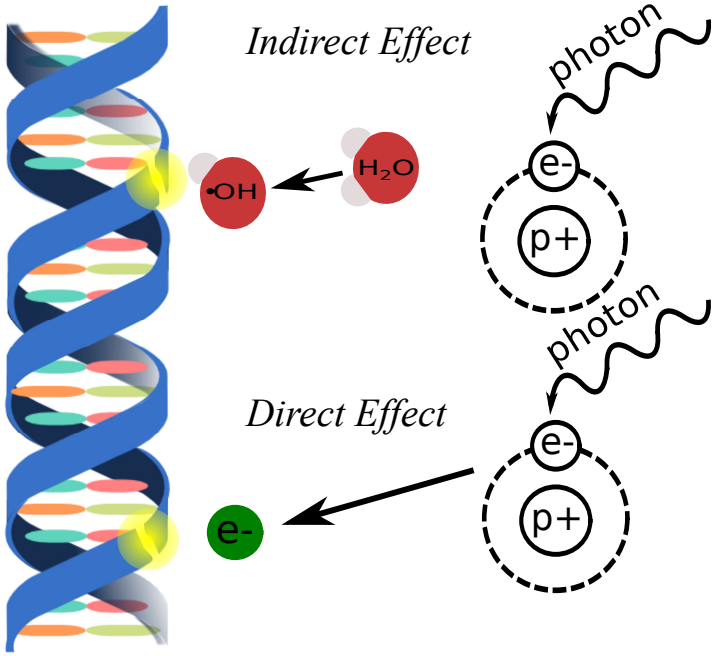

Figure 3 – Representation of direct and indirect action pathways.

radicals can react between them causing crossed neutralization, and with oxygen and lipids where chain reactions may be triggered. The direct damage is dominant for high LET radiations. Additionally, since water is the most abundant molecule in the cells, representing 70% to 80% of total cell mass, most of the energy deposited by the radiation is going to be absorbed in cellular water. The interaction of ionizing radiation with water generates short-lived H_2O^+ radical-cations, fast electrons, and electronically excited water molecules H_2O^*, which are responsible for the indirect damage.

The H_2O^+ and H_2O^* radicals are unstable and decay in a time frame of 10^{-13} s to form the radicals $\bullet OH$ and $H\bullet$:

$$\begin{aligned} IR + H_2O &\longrightarrow H_2O^+ + e^- \\ H_2O + H_2O^+ &\longrightarrow H_3O^+ + \bullet OH \\ &\textit{and} \\ IR + H_2O &\longrightarrow H_2O^* \\ H_2O^* &\longrightarrow H_2O + photon \quad emitted \\ &\textit{or} \\ H_2O^* &\longrightarrow \bullet OH + H\bullet \end{aligned} \tag{2.2}$$

The most highly reactive radical species produced during water radiolysis is $\bullet OH$. It can also diffuse a short distance and react with critical target molecules producing another radical. The ejected secondary electrons may interact with water and produce hydroxyl ions and a hydrogen atom (a hydrogen radical $H\bullet$), or they may suffer several interactions with the medium until they reach thermal energies after 10^{-11} s. Thermalized electrons are then solvated by dielectric interactions with water molecules and form e^-_{aq}, a free electron in a solvent cavity surrounded by a sheath of orientated water dipoles. This e^-_{aq} could react with protons producing a $H\bullet$.

$$\begin{aligned} e^- + H_2O &\longrightarrow H_2O^- \\ H_2O^- &\longrightarrow OH^- + H\bullet \\ e^-_{aq} + H^+ &\longrightarrow H\bullet \end{aligned} \tag{2.3}$$

The relative yields of water radiolysis products depend on the medium pH and radiation LET. The concentration of these radicals are expressed in terms of the G value, which is defined as the number of radicals or molecules produced per unit energy deposit (100 eV).

From the mechanistic point of view, after a cell is irradiated it goes through several processes which are described by stages taking place on different time scales: physical stage, physico-chemical, chemical, bio-chemical, and biological stages [55]. The description of these stages is of crucial importance when simulating radiation damage to DNA or biological systems, specially the first three listed. The physical stage include the interactions between primary and secondary particles and the atoms of which the tissue is compounded, normally it elapses during the first 10^{-18} s (DNA molecule) or 10^{-14} s (mammalian cell) after the system was irradiated. The physical stage is closely related with the next two stages since at the end of it the water molecule is left excited or ionized and sub-excitation electrons are going to be generated. During the physico-chemical stage (10^{-14}-10^{-12} s), several processes occur like ion-molecule reaction, dissociation, autoionization of excited states, thermalization of sub-excitation electrons (solvation electrons), and hole diffusion, to mention some of them. This stage leads to the formation of the radicals resultant from water radiolysis as was explained above. The chemical stage (10^{-12}-10^{-6} s) describes the reaction and diffusion of the chemical species. Finally, the

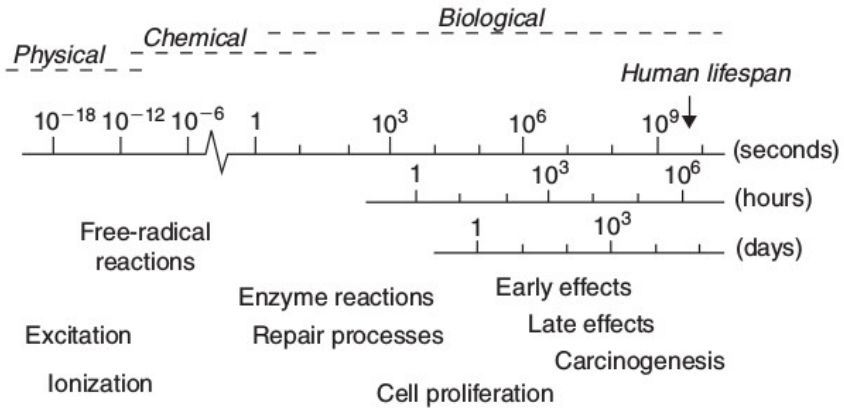

Figure 4 – Time scale of effects of radiation exposure in biological systems. Source:[55]

biological stage include all subsequent processes, beginning with enzymatic reactions that act on the residual chemical damage. During this stage, mechanisms of repair are activated. In some cases, these mechanisms fail, leading to cell death. The killing of stem cells and the subsequent loss of cells give rise to early manifestations of normal tissue damage during the first weeks and months after the irradiation. A secondary effect of cell killing is compensatory cell proliferation which may occur in normal tissues and in tumours. At later times after the irradiation the "late reactions" appear, these include fibrosis, telangiectasia of the skin, spinal cord damage and blood vessel damage. The time scale of the observable effects could last several years after the exposure.

So far we have mentioned that both direct and indirect actions may lead to the formation of free radicals, but we have not said how these radicals produce damage to DNA molecule [52]. Free radicals contain unpaired electrons and they are highly reactive, as it was mentioned before. They will undergo multiple reactions in order to acquire new electrons or get rid of the unpaired ones. Those reactions are considered slow compared to the time scale of the initial ionization events but are faster than normal enzymatic processes in a typical mammalian cell. It is considered that free radical reactions are completed in milliseconds after the irradiation. The $\bullet OH$ for instance can suffer hydrogen abstraction from other molecules and addition across carbon-carbon or other double bonds. Other complex macromolecules that have been converted to free radicals can undergo transmutations in the process of getting rid of the unpaired electron, and that could lead to the breakage of chemical bonds. For the specific case of the DNA molecule, these broken bonds may result in the loss of a base or an entire nucleotide, or an excision of the sugar-phosphate backbone, involving one or both strands. Sometimes these chemical

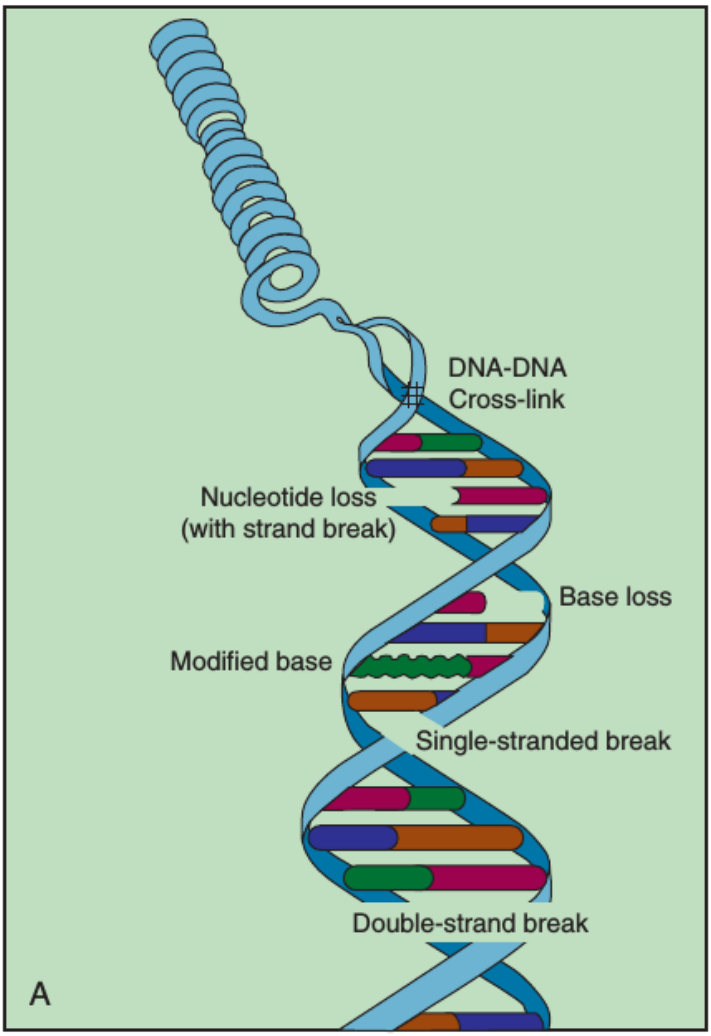

Figure 5 – Types of damage found in irradiated DNA. Source: [52]

broken bonds are rearranged, exchanged, or rejoined in inappropriate ways. Bases in the DNA can be modified by the addition of one or more hydroxyl groups (ex. thymine converted to thymine glycol), pyrimidines may become dimerized, or the DNA may become cross-linked to itself or to associated protein components. Since the energy deposition events are discrete, then the free radicals produced are also clustered, and they may undergo several chemical reactions and induce multiple damages in a highly localized area. This is know as "multiple or clustered damage sites".

2.2 DNA molecule description

One of the most important components inside the cell nucleus is the deoxyribonucleic acid (DNA). In this molecule the genetic information is stored to be used during the development, functioning, and reproduction of the cell. The role of this molecule is fun-

damental, that is the reason why it is considered one of the critical targets of ionizing radiation action. The DNA molecule is a polymer conformed by two twisted strands forming a double helix of approximately 2 nm of diameter and around 2 m length in humans [57]. The basic element of each strand is the nucleotide which is composed by a nitrogen base (A-adenine, T-thymine, C-cytosine and G-guanine) and a sugar-phosphate group. Both strands have complementary bases (A-T and C-G) linked by hydrogen bonds.

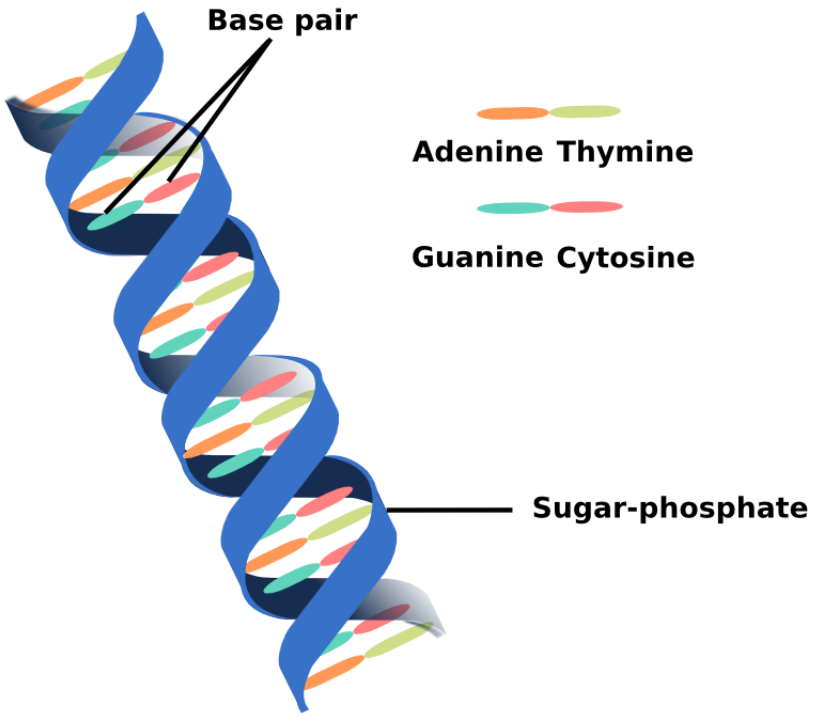

Figure 6 – Schematic representation of the DNA structure.

The bases are classified into two types, see Fig. 7:

- **purines**, A and G, consisting of two aromatic (term used in organic chemistry to describe a ring-shaped and flat molecule with resonance bonds) organic compound rings.
- **pyrimidines**, C and T or Uracil (U) in the case of the ribonucleic acid (RNA), consisting of a single aromatic organic compound ring.

The RNA is also present in the cell which acts as an intermediary between DNA and proteins. It differs from the DNA not only in the uracil base replacing thymine but

also by the sugar, ribose in this case.

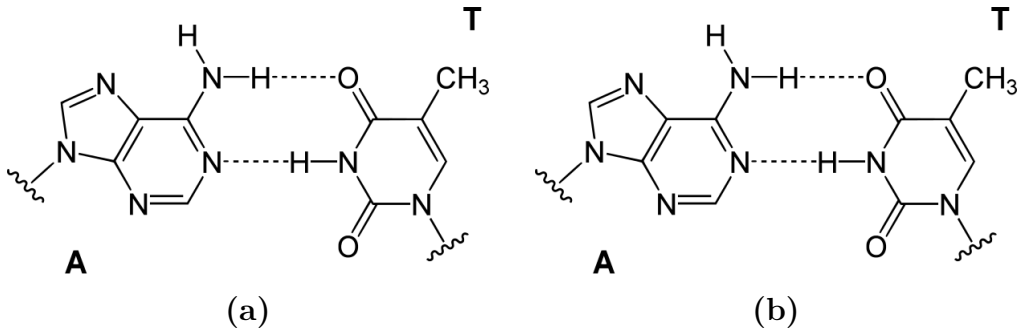

Figure 7 – Nucleobases. Chemical representation of two pairs of nucleotides: Adenine-Thymine and Guanine-Cytosine Source: [56]

The DNA packing into chromosomes is performed through proteins called histones (e.g. Fig. 8). There are several degrees of compaction of DNA:

1. Nucleosome: It is formed by the DNA double-helix wrapped around the histone, 11 nm of diameter.

2. Chromatin fiber: The sequence of nucleosomes (each containing two hundred base-pairs) is compacted in a supra helix with diameter of 30 nm.

3. Chromosomes: Chromatin can undergo different levels of compaction to form the chromosomes. DNA of human cells condenses into 46 chromosomes each of which contains about 35,000 genes (or several million base-pairs).

The degree of compaction of the DNA changes with the phases in the cell cycle, see Fig. 9. The cell cycle phases are briefly explained:

1. Interphase (G1, S, G2) the cell is growing. This phase corresponds to the 90% of the cell lifetime. The DNA during this phase is unwound and the cell will make a copy of it. The DNA, contained in the nucleus, is attached to its copy by a structure that is called a centromere.

2. Mitosis (M phase, which includes prophase, metaphase, anaphase and telophase) is a relatively short phase of the cycle and also complex and highly regulated. During this phase, the chromatin is wound into chromosomes. The nuclear envelope breaks and the chromosomes are free in the cytoplasm. Each chromosome copy is pulled by

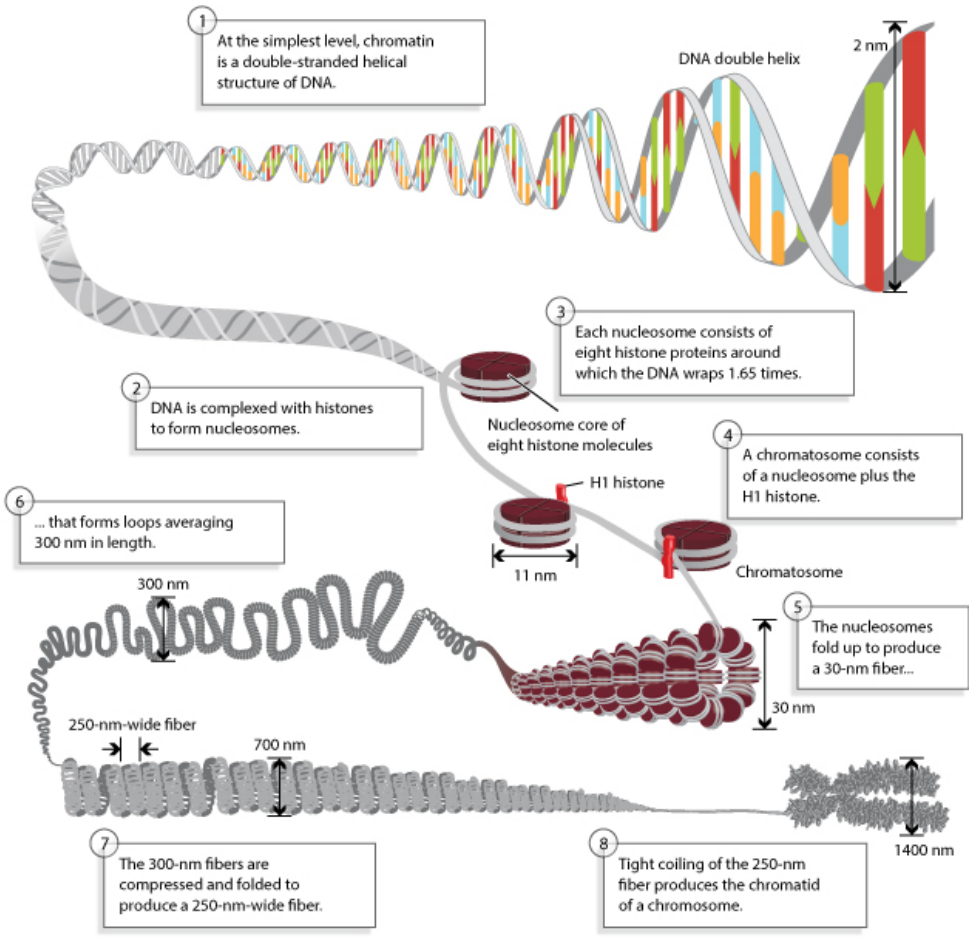

Figure 8 – DNA Packaging: Nucleosomes and Chromatin. Chromosomal DNA is packed inside the nucleus with the help of histones. Nucleosomes fold up to form the 30 nm chromatin fiber, which forms loops of about 300 nm in length. The 300 nm fibers are compressed and folded to produce a 700 nm thickness fiber, which is tightly coiled into a chromatid of a chromosome. Source: [57]

one pole or the other of the cell and the nuclear envelope is rebuild around each set of copy of the genome. The DNA collapses, the cell divides into two daughter cells.

2.2.1 DNA structure in different forms

There are three main different forms of the DNA molecule depending on the conditions of the biological environment (Fig. 10). For instance, Tomita et al. [59] have shown that when the salt content is high the DNA is more twisted. Yet, the conformation adopted by the molecule of DNA also depends on its degree of hydration and its sequence.

The ordinary form of DNA, the B-form, is the most common one in the cell under normal physiological conditions (high humidity and low salinity). B-DNA nitrogenous bases are almost perpendicular to the helix axis and it has around 10 nucleotide pairs

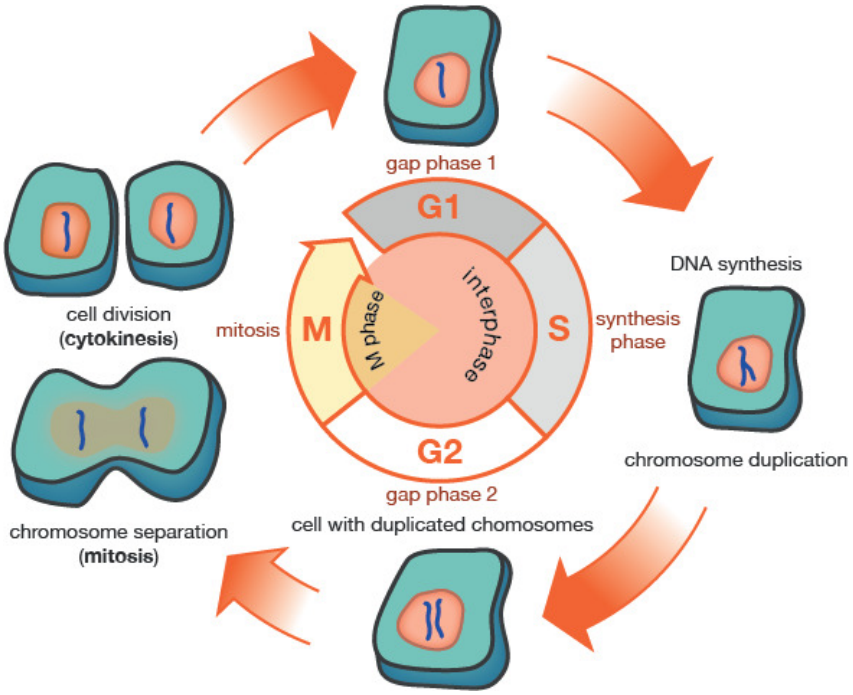

Figure 9 – Four characteristic phases of the cell cycle: Gap phase 1 (G1), DNA synthesis (S), Gap phase 2 (G2) and Mitosis (M). Source: [58]

per turn. The attachment of bases to the sugar backbone is asymmetrical resulting in the formation of two different grooves on opposite sides of the base pairs, the major and minor grooves. Although the grooves are of similar depth in B-DNA, the major groove is considerably wider than the minor groove.

When the water content is reduced or the presence of cations is important, the DNA goes from B form to the A form. In the A form, the double helix is more twisted. This conformation is shorter and more compact than B-DNA. The B to A transformation occurs not only when the relative humidity of the sample is lowered but also when the heteroduplex with RNA is formed.

Finally, the Z form has also been reported with increasing salt content and repetition of sequences purines-pyrimidines. Unlike the A and B forms, Z-DNA is a left-handed double helical structure. The Z-DNA is less twisted and longer than the B form and its sugar-phosphate skeleton forms a “zigzag”.

2.2.2 Experimental methods to measure DNA damage

The DNA molecule has already been identified as the primary target for ionizing radiation induced cell killing and mutations [60, 61, 62]. Most of this damage is repaired by

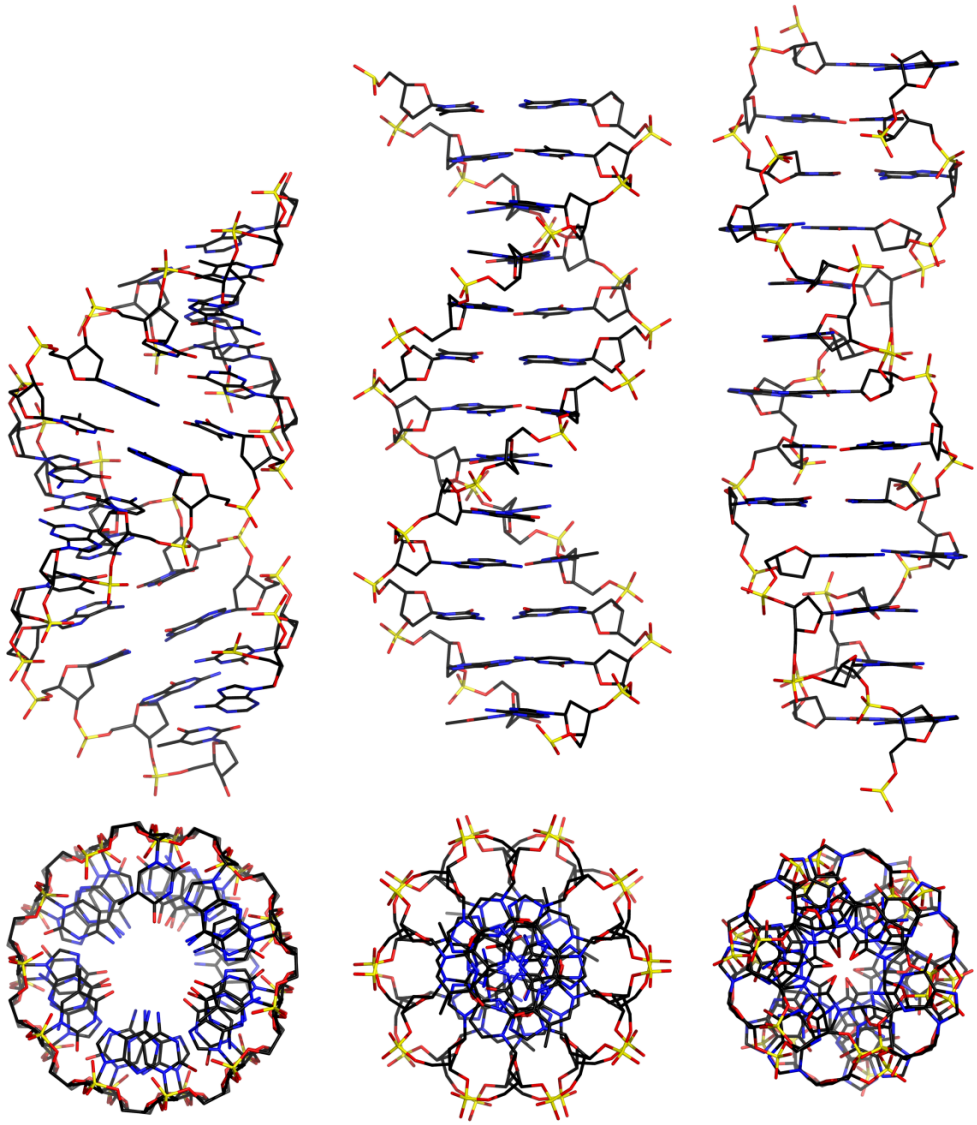

Figure 10 – The three main forms of DNA, from left to right: A, B and Z form. Source: [56]

cellular enzymatic systems, but the proportion of DNA double strand breaks not repaired may lead to chromosome aberrations that can be lethal to cells. The cell survival after radiation exposure is also dependent on the volume of total DNA contained within the nucleus. Mammalian cells containing more DNA per cell compared to yeast cells are more sensitive to radiation [60].

Several experimental techniques can be found in the literature [63]. These experiments are usually related to a specific observable effect caused by the ionizing radiation action. There are several ways by which cell killing occurs: the cell could lose the reproductive integrity, suffer apoptosis (programmed cell death), necrosis (premature accidental death, rapid depletion of ATP and mechanical breakdown of cellular components including

the cell membrane, nuclear chromatin, and organelles), senescence (biological aging, gradual deterioration of function), and autophagy (self-digestion). For each of these processes specific experiments can be performed taking advantage of some distinguished signals of them. For instance, the clonogenic assay, which is extremely important in oncology research, is able to assess the ability of a single cell to undergo multiple cell divisions to form a macroscopic colony, so it gives us a measure of cell inactivation probability after radiation exposure. In the case of apoptosis, a loss of mitochondrial potential can be detected by flow cytometry and DNA fragmentation can be detected using the deoxynucleotidyl transferase dUTP nick end labeling (TUNEL) assay or enzyme-linked immunosorbent assay (ELISA). During necrosis, the cellular membrane is degraded allowing some dyes to penetrate it and this can be detected using light microscopy or flow cytometry. One of the standard methods used to detect autophagy is through cellular morphological changes (granular cytoplasmic ubiquitin inclusions), identified using electron microscopy. However, the use of markers like microtubule-associated protein 1 light chain 3 (LC3) or p62 detected by immnuoblotting or immunofluorescent techniques have proven to be more suitable to monitor autophagy. Moreover, colorimetric assays can be also used to analyze cell growth. Colorimetric assays utilize compounds that are reduced by cells to become products that visibly change color, which can be detected by spectrophotometry.

The techniques briefly mentioned so far neither determine DNA damage yields nor are based on the study of DNA repair mechanisms. For that specific purpose other experimental methods exist. It has been mentioned that ionizing radiation induces base and nucleotide damages, cross-link with protein or DNA, and single and double strand breaks (SSB, DSB). The two main pathways responsible for repairing DSBs are non-homologous end-joining (NHEJ) and homologous recombination (HR). The NHEJ pathway repairs DSB damage throughout the cell cycle and it is the primary repair path in mammalian cells. The DSB are recognized by the Ku70/Ku80 protein complex which recruits DNA-dependent protein kinase catalytic subunits (DNA-PKcs) to the damaged site on each strand. The union of the DNA-Pkcs bounded to Ku and the DNA, it is known as the holoenzyme DNA-PK. This holoenzyme becomes activated and is able to autophosphorylate, which is imperative for enzymatic function. DNA-PK is responsible for juxtaposition of the DSB ends and recruitment of other NHEJ proteins, such as the LigaseIV/XRCC4 complex, that bind the broken DSBs together.

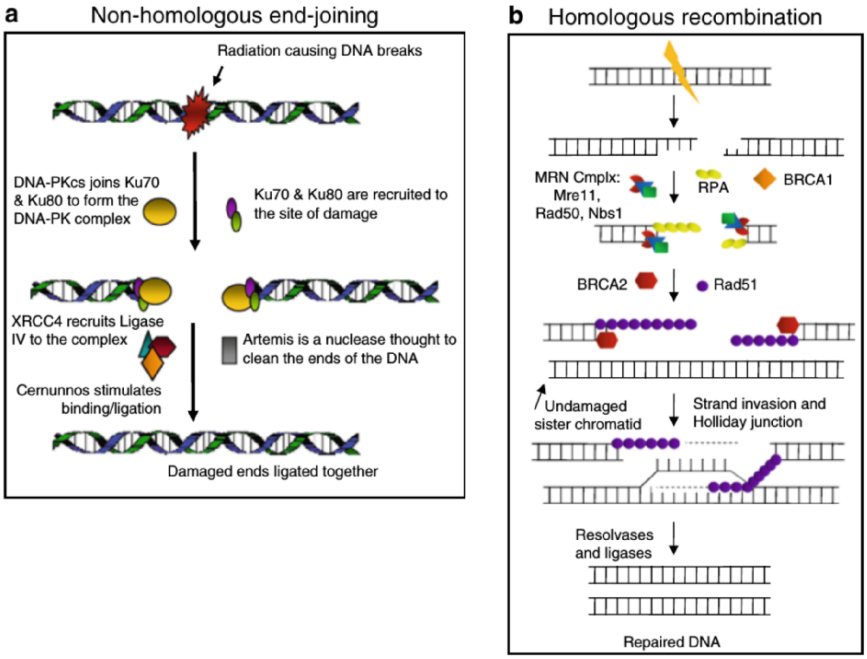

Figure 11 – The pathways involved in DSB repair in mammalian cells. (a) Nonhomologous end-joining: NHEJ and (b) Homologous recombination. Source: [63]

The HR takes place during the S/G2 phase of the cell cycle. This repair mechanism uses an undamaged sister chromatid as a template to repair the damaged chromosome arm. The protein complex Mre11/Rad50/Nbs1 (MRN) and BRCA1 recognize the DSBs. The replication protein A (RPA) recognizes the single stranded DNA (ssDNA), protects the strand and provides access for other proteins to the site. RPA is removed from the single-strand portion of the break and replaced by Rad51. BRCA2 aides in the RAD51 nucleoprotein filament strand invasion at a homologous region of the sister chromatid to form a D-loop. The second DSB end invades a portion of the D-loop to form a Holliday junction [63]. DNA polymerase is then recruited to the Holliday junction to fill in the missing sequence followed by the ligase/resolvase complex that resolves the complex and completes the repair. HR proteins are essential for cell survival, and mutations in genes coding for these proteins can greatly increase the risk of carcinogenesis.

2.2.2.1 Measurement of DNA Damage Repair

The first techniques employed to study DNA damage and repair mechanisms involved exposing cells to high doses of IR and then evaluating DNA fragments produced by sucrose gradient centrifugation techniques [64], alkaline and neutral elution [65], and gel electrophoresis (PFGE) [66, 67]. Using the PFGE technique, smaller DNA fragments

generated following IR exposure have a faster mobility and will travel further down an agarose gel compared to larger fragments. The gel usually has a vertical orientation and it is cast with a comb that generates wells in which the samples are placed. Since the DNA is going to be repaired with time, less DNA elutes from the well therefore by quantifying the DNA in the well against the DNA that elutes from the well, an accurate assessment of DNA damage repair over time can be determined. This method is good for measuring repair kinetics, but it is also laborious and requires high doses of radiation of at least several tens of Gy [108, 109].

There are other methods for analyzing DNA repair like the single-cell electrophoresis (the comet assay) [68, 69] and phosphorylation of H2AX (gH2AX)[69, 70]. In principle the comet assay is similar to PFGE, smaller DNA fragments travel faster than larger fragments through electrophoresis. The advantage of this technique is that it requires less dose, individual cells can be examined and either SSBs or DSBs can be assessed. The irradiated cells are suspended in agarose, positioned on a slide, lysed, and then exposed to electrophoresis. DNA is then stained and examined under a microscope as seen in Fig. 12. The small DNA fragments migrate outside the nucleus under electrophoresis generating an image similar to a comet (the head is the nucleus and the tail is the fragmented DNA traveling outside the nucleus). With time, the tail of the comet decreases due to the repair of DNA damage. This assay is more sensitive than PFGE and easier to perform; however, scoring of the cells can be time-consuming and tedious.

Finally, there is another technique known as quantifying gamma histone H2A (gH2AX), also used to quantify DNA damage and repair, particularly for DSBs. Actually, this is the most common method to measure DSB repair kinetics. As already mentioned, nuclear DNA is wrapped around proteins named histones: H2A, H2B, H3 and H4. Particularly, the histone H2A becomes phosphorylated and activated by damage response proteins like ATM at sites where DSB occurs. When activated, this histone is called gH2AX. Fluorescent-tagged antibodies recognizing the phosphorylated gH2AX enable individual foci generated at each DSB site to be evaluated by microscopy (Fig. 13), by flow cytometry or by western immunoblotting. This technique is very useful identifying agents that may inhibit IR repair and it is also used as a biological IR dosimeter able to quantify the response to doses of the order of few Gy [71].

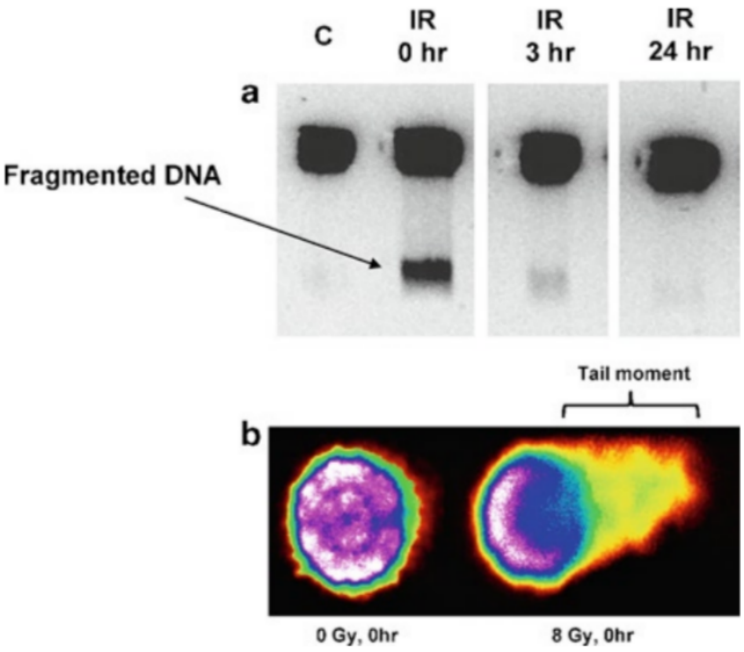

Figure 12 – (a) Pulsed-field gel electrophoresis of HT-29 cells exposed to 20 Gy and evaluated at 0, 3, 6 and 24 h post-radiation treatment. (b) Comet assay (or single-cell electrophoresis) showing an unirradiated cell versus a cell exposed to 8 Gy at 0 h. The nucleus represents the head of the comet while the fragmented DNA appears scattered like the tail moment of the comet. Source: [63]

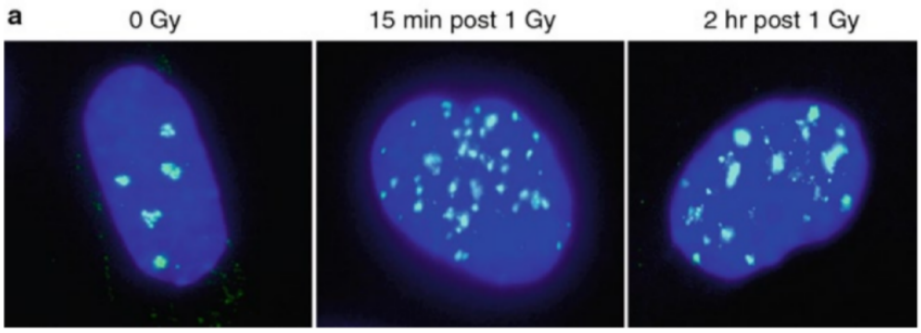

Figure 13 – Use of γH2AX to measure DNA damage and repair evaluated using microscopy. Immunofluorescence using an anti-γH2AX primary antibody to detect DSB foci in unirradiated cells (First image) and cells exposed to 1 Gy (Second and third image, at 15 min and 2 hours after exposure). Source: [63]

2.3 Monte Carlo simulation of particle transport.

The Monte Carlo (MC) method provides a numerical solution to a specific problem that models object to object or object with the environment interactions based on the knowledge of their relationships. This method uses random sampling to simulate these relationships. Its origin goes back to 1940s, with the Manhattan project for the development of the atomic bomb, but it was previously used by Laplace and Buffon in the 18^{th} century to calculate the number π [10, 100]. MC method has been extensively used in radiation transport research considering the stochastic nature of the mechanisms involved in the interaction of radiation with matter. Several MC codes have been developed for the simulation of radiation transport, some are known as general-purpose codes that could be applied in many research fields and others specialized in specific areas. In general, the available codes may be classified as those that use a condensed history technique and the ones that follow each particle step by step. In the first group, like its name suggests, a given number of scatterings along the particle path are grouped together, which reduces the processing time dramatically. A sum up of general purpose codes was presented by Nikjoo et al. [44]. The codes from the second group are known as track structure codes (TSC). Their origin is deeply related with the need of modeling in detail the track structure in water for microdosimetry or nanodosimetry applications. They evolved from initially using a single parameter (1D) to describe the radiation track to the simulation of the charged particle track step by step in a condensed medium, followed by the time evolution of the radical species generated (4D). Some of these codes are Radiobiology-oriented [26, 27, 72] or can be linked to other more specific developed for this purpose [1, 3].

2.3.1 Radiobiology oriented track structure codes

The state of art in the calculation of early DNA damage induced by ionizing radiation using Monte Carlo simulation implies the use of a TSC with a detailed geometrical DNA model inside it or the combination of this TSC with a radiobiology oriented external application which implements the biophysical model. There are several examples of TSC [44], but only a few of them allow to simulate the chemical stage. The latter are considered 4D codes since they describe the distribution of chemical events inside the geometry as a function of time. The main differences between those codes lies in the physical models

implemented and the parameters of the biophysical model.

The KURBUC TSC was introduced by Nikjoo et al. in 1993 [38], which is a suite of codes that allow to simulate electron and ion transport in water. In its first version, KURBUC simulated electron tracks in water vapor for initial electron energies between 10 eV-10 MeV. This group also developed other codes, as part of KURBUC suite, capable of simulating full slowing down tracks of low energy protons (LEPHIST code) and alpha particles (LEAHIST), with energy ranging from 1 keV to 1 MeV and 1 keV to 8 MeV, respectively. All this codes provide the coordinates of molecular interactions in water vapor, the amount of energy deposited per event, and the interaction type. The KURBUC code produces the initial yields of ionized water molecules H_2O^+, excited water molecules H_2O^*, and subexcitation electrons e^-_{sub} (E $<$ 7.4 eV) at $\sim 10^{-15}$ s. For the case of excited water molecules, three groups were used: A^1B_1, B^1A_1, Rydberg states, diffuse bands, and dissociative excitations. The pre-chemical and chemical stages of electron tracks in liquid water is described in a KURBUC extension, named CHEMKURBUC. In the pre-chemical stage (10^{-15}-10^{-12} s) the three main species, H_2O^+, H_2O^*, and e^-_{sub}, created during the physical stage are converted into molecular products [44]. During the chemical stage (10^{-12}-10^{-6} s) a step by step approach was adopted. For each time step, water radicals and their products were allowed to diffuse randomly with a diffusion coefficient D. The values of D for the 11 species considered were taken from Beaudré et al. [73]. The root mean square distance was calculated as $\lambda = (6D\tau)^{\frac{1}{2}}$, where τ is the time step and the actual distance was extracted from a Gaussian distribution. After diffusing, those species that were separated by a distance lower than their reaction radius were replaced by their reaction products. The reaction radius, similar to the other codes that simulate water radiolysis, is derived from the Smoluchowski's diffusion equation. The reaction rate constants used in CHEMKURBUC were the same as those given by Beaudré et al. [73]. In total 26 reactions were considered and the time dependent yields for the chemical stage were calculated and compared with various experimental data.

In 1997, Nikjoo et al. [72] used the CPA100 TSC [74] to calculate the yields of initial DNA damage produced by low energy electrons with energies from 100 eV to 4.5 keV. This work is an example of a parameter study of mechanistic DNA damage simulations. With the CPA100 code electrons, as primary particles, with initial energies from 10 eV to 100 keV can be simulated. The transport medium employed was liquid

water. Both primary and secondary electrons were tracked down to thermalization, at this point the position of all inelastic interactions, the amount of energy deposited, the interaction type, and the time of generation of the initial radical species were recorded. In the pre-chemical stage, ionized and excited water molecule and subexcitation electrons are produced. Finally, starting from 10^{-12} s the chemical stage begins with the generation of the water radicals. The concept of "physical track" involved all coordinates and identities of the physical interactions and the "chemical track", the information of all the radical species and molecular products. As scoring method, they placed a single electron track in both physical and chemical forms starting at the centre of a virtual sphere, sufficiently large to achieve electronic equilibrium and then placed randomly a DNA segment of canonical B-DNA. If a physical track entered the region occupied by the DNA segment, a direct hit was considered. A 4-nm cylindrical shell was built around the DNA segment, where radical diffusion and reactions are followed. The value for the cylindrical shell radius was set to mimic the average diffusion length of the radicals in a cellular environment. However, they also studied the influence of varying this parameter. All reactions during the chemical stage were considered diffusion-controlled and the rate constants and reaction radius were taken from Buxton et al. [75]. The values of the diffusion coefficient for the hydroxyl radical was the same as that used in GEANT4-DNA (taken from Refs. [76] and [77]). The diffusion of all radical species was in accordance with the Smoluchowski equation. The time interval for diffusion started at 10^{-12} s and was set to obtain a step size of about 0.1 nm. Whenever a reaction occurred inside the cylindrical shell, the reactants were removed and the products were placed randomly within the reaction radius. In the case that a radical left the cylindrical shell, it was also removed. Chemical hits were considered when a $\bullet OH$ radical reached the DNA segment. The DNA model consisted in a cylinder divided into regions as sugar-phosphate moiety and bases without an atomistic description. The sugar-phosphate chain wraps helically around the central cylinder with 1 nm diameter and the helical twist was 36°. The diameter of the DNA molecule was 2.3 nm and the double helix pith was 0.34 nm. The model structure included the first hydration shell of the DNA. Another assumption made was that the resulting charge of any reaction with the bound water layer is transferred to the DNA. The energy deposition threshold for direct action to the sugar-phosphate volume was initially set to 17.5 eV and no migration of energy along the DNA was assumed. Every time the $\bullet OH$ radical reached a sugar-phosphate or a nucleobase it was considered to react, and an activation probability

of 0.13 was set as the efficiency for strand break (SB) induction after the reaction. That activation probability takes into account a 20% probability that the $\bullet OH$ radical reaches the DNA will react with a sugar-phosphate and a 65% probability that a sugar radical produce a SB. An investigation was performed to account the dependence when varying the activation probability in the calculation from 0.002 to 0.20. A distinction was made to the DNA damage classification by complexity (patterns of breaks in one or both strands) and nature (direct or indirect). The use of liquid water instead of vapour was essential to consider indirect effects. The results showed that the threshold energy of 17.5 eV was the most probable for the production of an SB due to direct action. They also obtained that about 90% of total energy depositions was due to events with less than 60 eV but the largest DSB yields correspond to energy depositions in the 60-150 eV range for all the electron energies. DSB are most likely to be induced by lower energy electrons that have the highest ionization densities and their SB production capability is increased due to associated $\bullet OH$ radicals within the same cluster. This study showed a predominant contribution of diffusing $\bullet OH$ radicals to the DNA strand breakage when irradiated by low energy electrons. Also, the complex DSB contribution was of about 30% for almost all electron energies. Moreover, the study showed that the length of damaged sections of DNA are quite short.

A work published by Štěpán and Davídková [3] studied the DNA damage distribution along a DNA oligomer irradiated by 10 keV electrons, protons of 1, 2, 5, 10, and 20 MeV and α-particles using the simulation tool RADAMOL. RADAMOL is a modular set of several stand-alone codes written in ANSI C (standard for the C programming language). The track structure input can be provided by TRIOL [4] or GEANT4-DNA [35, 36, 37], with information of the coordinates, type of interaction, and deposited energy. The other input is the atomic level model of the DNA biomolecule in PDB format. In both cases the electrons are tracked down to energies of 7.4 eV, then are thermalized at the beginning of the physico-chemical stage. The biological target was placed at the center of virtual sphere filled with water. Individual tracks were projected over that volume, one at a time, and their time-space evolution was followed during the physico-chemical and chemical stages. They also implemented a module to take into account the effect of charge transfer along the DNA molecule. In STOCHECO the radiolytic species produced from ionized and excited water molecules diffused and underwent mutual chemical reactions. During the prechemical stage the energy deposition events in bulk water were converted to

correspondent chemical products. For excitations, two pathways were taken into account: A^1B_1, B^1A_1. Rydberg states and diffusion band superexcitations led to autoionizations. The chemical stage started from the initial spatial distribution of chemical species produced during the prechemical stage. RADAMOL considered 21 reactions between radical species and three reactions with dissolved oxygen. The model also allowed to follow the effect of other radical scavengers. The reaction radius was obtained as a solution of the Smoluchowski equation, an a reaction was considered to occur with an atom if the reaction could take place between the chemical specie and the target molecule, the surface of Van der Waals sphere of the atom was accessible from the solvent, and the separation distance was less than the sum of Van der Waals radius of the atom, radical species, and the corresponding reaction radius. The model acconte for unscavengeable and scavengeable damages. Those radicals produced in the bound water shell around DNA were assumed to react with DNA and contribute to the unscavengeable damage. The diffusion and chemical reactions of the radicals were followed step by step up to 10^{-6} s. The RADACK code considered the reactions of $\bullet OH$, e^-_{aq}, and $H\bullet$ with DNA as scavengeable radiation damage. An analysis was performed taken into account the damage complexity and all type of base damages were scored as modified bases. Charge migration was implemented considering that for ionization events in a bound water layer, both negative and positive charges migrate to the closest atom in the DNA. Those electrons localized in deoxyriboses do not migrate and were scored as SBs. From the base moiety any electron will migrate with an 80% probability to a cytosine and 20% probability thymine within a distance of $\pm$ (3-11) bp. Hole migration from sugar to base was assumed to occur with a specified probability for each base (values derived from Ref. [78]). Holes from C, A and T bases were considered to migrate to G in $\pm$(3-11) bp with 85.7% probability. The implementation of the charge transfer processes is a simple statistical model. They compared an irradiation with an electron beam of 10 keV with a 10 MeV proton beam, with comparable LET of 5 $keV/\mu m$ and the distribution of damaged bases and sugars are very similar. The profiles of base damage obtained when comparing with 2 MeV α-particles, were similar for both low and high LET particles. The difference came in terms of the overall yield of DNA damage. It was also observed that due to the higher ionization density in the track core for 2 MeV α, more chemical reactions occurred and consequently less radical attacked the DNA moiety. The yield of SSB and modified bases decreased with increasing LET for a given particle type. The clustered damage yields increased with LET up to a

maximum where the saturation effects appeared. The effects of electron and hole transfers along DNA double helix were also shown. Additionally, hole migration to guanines and relocation of electrons to cytosine and thymine bases within 3 to 11 bp were shown. The authors also concluded that when a negative charge is located in the deoxyribose, it leads to a direct DNA damage and in the case of a hole, 50% is localized in the sugar moiety and leads to a modified sugar.

One of the main references in this field is the simulation code or suite of codes PARTRAC, developed by Friedland et al. [26, 27, 79, 80] at the Helmholtz Zentrum München, German Research center for Environmental health, Institute of Radiation Protection in Germany. This TSC implemented a genetic material model with atomic resolution and also influenced the GEANT4-DNA design to simulate water radiolysis. The reaction rate constants, diffusion coefficients, and branching ratios used in GEANT4-DNA were adopted from PARTRAC. Initially, this code was created for electrons in water vapour and it has been extended for photon interactions. Afterwards, atomic resolution geometrical models for the double helix, chromatin fibers and chromosomes were introduced. The development of this code included cross sections for liquid water and stochastic chemistry calculation, track structure for heterogeneous targets, heavy charged particles interactions, and a stochastic model of DNA double strand break (DSB) repair via de NHEJ pathway. This simulation suite has a modular structure with well-defined data interfaces, each one describing an individual stage of radiation interaction and response. All these separated program codes were written in FORTRAN. The physical stage, similar to the other TSC mentioned earlier in this section, describes the track structure of primary and secondary particles in an event by event manner, from their initial energy to nearly total stopping. PARTRAC can simulate photons, electrons, protons, α-particles, and heavy ion tracks in liquid water. The ionized and excited water molecules produced during the physical stage are transient states that decay rapidly and form chemical species within about 10^{-12} s. Electrons with energies below 10 eV, including the ones produced by auto-ionization of excited states come to rest during this time interval and become solvated by attachment to water molecules that interfere immediate recombination. The physico-chemical stage is determined by the decay channels, their branching ratios, and the position of the species produced [77]. The chemistry module describes the water radiolysis in oxygen free water. The diffusion of each chemical species is performed by jumps in randomly selected directions. After each diffusion step, an analysis is performed to look

for reaction partners to separation distances compared to the reaction radius (Smoluchowski equation). Those reaction radius were calculated from the reaction rate constants assuming partially diffusion controlled reactions. After the reaction occurs, the reactants are replaced by the reaction products. The biophysical model used in PARTRAC evolved from a flexible arrangement of nucleosomes within a 30 nm chromatin fiber or a zig-zag formation and their connection by linker DNA segments, both with atomic resolution, to higher order DNA structures, always using basic fiber elements under the boundary condition that the structure reiterated identically with a certain shift along the fiber axis after a predefined number of nucleosomes. The most recent chromatin fiber developed consisted in a flat chromatin loop with 18 elements and about 100 kbp. In PARTRAC, the first steps were taken to represent whole chromosomes using virtual linkage for a study of DNA fragment distributions. A human fibroblast cell nucleus in G0/G1 phase with 46 chromosomes territories (Fig. 14) was constructed inside a cylindrical nucleus model with 15 μm diameter and 5 μm height. Other chromosome models ahve been developed, describing a human inter-phase cell nucleus of a lymphocyte with a spherical shape of 10 μm diameter and a fibroblast with an ellipsoidal shape with axis lengths of 20 μm, 10 μm, and 5 μm, both including a total genomic length of 6.6 Gbp. The study of radiation effects is performed by superposing the track structures determined in liquid water over the DNA target model. Similar to what has been described so far for other TSC, PARTRAC also simulates the physico-chemical and chemical stages. After the physical stage ends, reactive species are created outside those volumes occupied by the DNA constituents. The reaction radius used in PARTRAC were derived from reaction rate data taken from Buxton et al. [75]. The indirect strand break was assumed to occur in 13% of the interactions of $\bullet OH$ with the DNA, considering that 18.5% of the total $\bullet OH$-DNA interactions was with the sugar moiety and a breakage probability factor of 0.7. Friedland and coworkers calculated the DNA damage, in terms of SSB and DSB yields, to human fibroblasts after irradiation with photon, proton, and heavy ions beams with different beam quality. Undoubtedly, PARTRAC is the first example of TSC that simulate damage due to radiation action in a genome model with atomic resolution. Even when there is a lot of information and articles published about this work, this set of codes was not intended to be open source and publicly available.

The previously stated is not an issue in the case of the GEANT4-DNA toolkit [43, 49, 51, 81], which is an extension of the well known GEANT4 MC code. It was con-

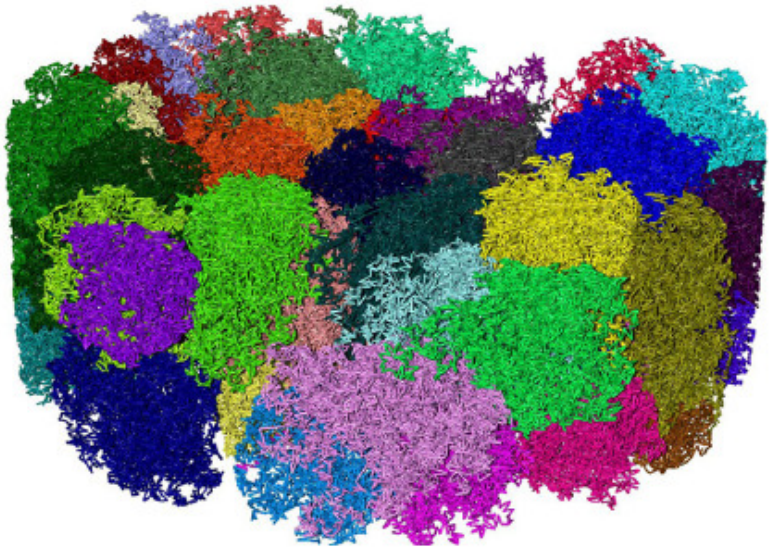

Figure 14 – Geometry used by Friedland et al. [27]: fibroblast cell nucleus in G0/G1 phase with 46 chromosomes territories.

ceived as an open source software for nanodosimetry and radiobiology research. GEANT4-DNA is a result of the collaboration group led by Sebástien Incerti and since its release has been used in many research areas like radiotherapy, targeted radionuclide therapy, and radiobiology. Since the use of this toolkit is a central part of this work, we are going to enter in more details later on. The GEANT4-DNA package in its newest version is capable of simulating the physical, physico-chemical, and chemical stages of ionizing radiation action in liquid water. The simulation of the chemistry stage in this toolkit uses the same diffusion coefficients, chemistry reactions table, and reaction rates constants as those implemented in PARTRAC, mainly because sources for the involved parameters is scarce. In order to simulate early damage to DNA, three approaches have been followed: the use of clustering algorithm, the explicit geometrical modelling of the DNA double helix and associated biological structures of interest, or a combination of the mentioned approaches. Francis et al. [82, 83, 84] were the first to propose a clustering algorithm, based on the DBSCAN algorithm [85], which was tuned in order to reproduce experimental data of double strand breaks and survival rates. This clustering algorithm was included in GEANT4-DNA in 2015 [50]. The disadvantage of the clustering approach is that it does not take into account the mechanistic simulation of physico-chemical and chemical interactions with sensitive biological targets. The other approach consists in developing a geometrical model of the biological target which can be combined with the simulation of the physical, physico-chemical and chemical stages. There are two main

categories of DNA geometrical models, a cylindrical approach where sensitive volumes are described using cylinders and a high resolution atomistic approach where each atom conforming the biological target is represented individually. As examples of cylindrical geometries there is a model proposed by Bernal et al. [25]. It consisted in a DNA double helix represented as a series of slices in the B-DNA conformation. Each slice included two sugar-phosphate groups bound by a complementary base-pair. wich was represented by a cylindrical shell. The nucleosomes are modelled as cylindrical histones surrounded by two DNA loops and were used to assemble 30 nm chromatin fibers, each containing 1200 bp per level (6 nucleosomes/level). Using this geometrical model, Incerti et al. [86] implemented in GEANT4-DNA a simplified nucleus made of randomly oriented chromatin fibers. Also the A and Z conformations of DNA were implemented using the same method. For recording both direct and indirect effects, some groups used a liquid water virtual cylinder around the DNA, known as hydration shell. As mentioned, the other option for the development of geometrical models is the high resolution atomistic approach. For this purpose, Bernal et al. [1] developed the first freely available stand-alone subroutine that implemented a full atomistic geometrical description of B-DNA. This subroutine was used in combination with GEANT4-DNA physics to compute direct damage yields [87]. A new solution to include atomistic geometries of macromolecules in GEANT4-DNA, directly from the Protein Data Bank (PDB), was described by Delage et al. [88]. Together with this application, an algorithm capable of finding the closest atom to the hit was also included, as well as strand break determination.

As said before, recent publications by several groups focused in developing more realistic cell geometries. For instance, the research performed by Meylan et al. [28] presented a new calculation chain based on the GEANT4-DNA toolkit. This work uses the capabilities that GEANT4-DNA offers to simulate the physical, physico-chemical, and chemical stages of early radiation damage. This group used the DnaFabric tool that allows the generation, editing, displaying, and reporting complex DNA geometrical models. This model does not have an atomic resolution but it describes the DNA using spherical volumes for the phosphate group, the deoxyribose, and the bases. Those geometries were used as the base unit to create a complete fibroblast nucleus (Fig. 15). Their approach allows to calculate the direct and indirect DNA strand break yields. Since the GEANT4-DNA was used in the present work, details about this toolkit are presented later.

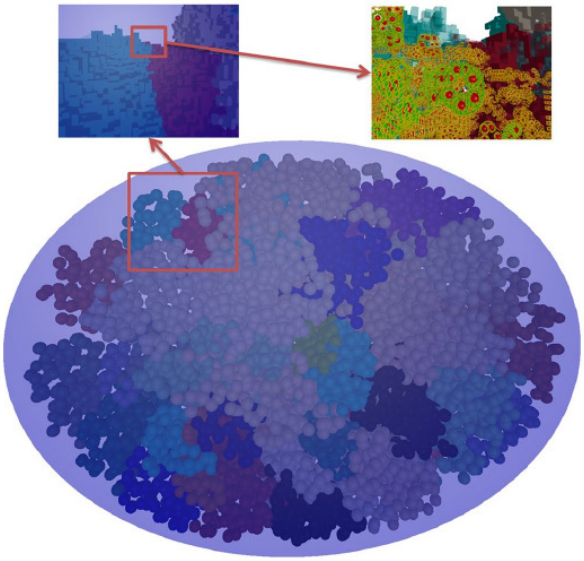

Figure 15 – Geometry used by Meylan et al. [28]: fibroblast cell nucleus.

Another work that used the GEANT4-DNA toolkit was recently published by Lampe et al [46]. They modelled a full genome of E. coli cell (Fig. 16) in GEANT4. For the nucleotide pair, similar to Meylan et al., they considered spherical volumes for the phosphate group, the sugar, and the bases, without atomic description. The results they presented showed a good agreement with simulations and some experiments. The parameter set used in this work were obtained in a previous sensitivity study performed by them [89], comparing their simulations with those performed by Nikjoo et al. [72].

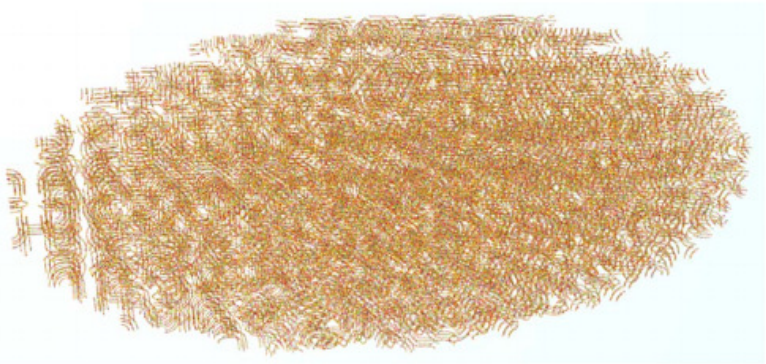

Figure 16 – Geometry used by Lampe et al. [46]: E. coli cell packed into an ellipsoid.

2.3.2 GEANT4 Monte Carlo simulation toolkit

GEANT4 simulation toolkit (GEometry ANd Tracking v10.2) [35, 36, 37] was conceived for the simulation of particle-matter interactions in high-energy physics. This code is written in C++ (object oriented programming) and it is presented as a toolkit

containing the elements needed to perform particle transport. Moreover, this toolkit is freely available to the scientific community from the project website. Nowadays, GEANT4 is considered a general purpose MC code and it is widely used in different applications like medical physics, space applications, microdosimetry, and radiation protection. GEANT4 can use both condensed history and step-by-step approaches, so it is a user task to select which one to use or combine them in the same simulation. The physical aspects of the code are explained in the physics reference manual [90].

There are some terms with a specific meaning in GEANT4. A **track** is a snapshot of a particle at a particular point along its path. A **trajectory** is a collection of track snapshots along the particle path. A **step** consist of the two endpoints which represents the fundamental propagation unit in space or time. **Process** has two meanings in GEANT4, as a term in computer science it refers to an instance of an application which is being executed. As a specific term of GEANT4, a **process** refers to a class implementing a physical interaction (ionization, multiple scattering, etc), and it is categorized by when the interaction occurs, either at the end of the step (PostStep) or during the step (AlongStep). An **event** consists of the decay or interaction of a primary particle with a target, and all subsequent interactions, produced particles and four-vectors. A **run** consists of a series of events.

GEANT4 structure is based on 17 major class categories, each one independently developed. Some categories constitute the foundation of the toolkit [35]:

- ***global*** This category covers the system of units, constants, numerics, and random number handling.
- ***materials, particles, graphical representations, geometry*** It includes the volumes for the detector description and the navigation in the geometry model.
- ***intercoms*** It provide means for interacting with GEANT4 through the user interface and a way of communicating between modules. It is also the repository of abstract interfaces for "plugins" (Visualization, etc).
- ***track*** This one contains classes for tracks and steps
- ***processes*** It evokes several physics models, which compute the total cross section and fully describe the final state of the interaction.

- ***tracking*** It manages process contribution to the evolution of track's state.
- ***event*** It manages events in terms of their tracks.
- ***run*** It manages collections of events that share a common beam and detector implementation.
- ***readout*** It allows the handling of "pile-up".
- ***visualization, persistency, interface*** These use all the mentioned categories and connect to facilities outside the toolkit.

2.3.3 GEANT4-DNA extension. Simulation of the physical, physico-chemical, and chemical stages.

The GEANT4-DNA package was released in 2007, as an extension of GEANT4 and it was developed to follow particles down to energies relevant in micro and nanodosimetry. This extension is a result of on-going research in the framework of the GEANT4-DNA project [81].

The GEANT4-DNA physics processes provide the user with a complete set of models for the interaction of electrons and the different charge states of hydrogen (H^0, H^+) and helium atoms (He^0, He^+, He^{2+}) with liquid water. This includes several models of cross sections down to the eV scale for electron interactions and they transport all particles on an event-by-event basis. These features allow a more detailed simulation of the structure of elementary energy deposits and secondary particle production in liquid water down to track lengths of the order of a few nanometers. The cross sections models take into account a detailed description of the water molecule, with five electronic excitation modes as well as five ionization channels (4 valence shells and the oxygen K shell). In GEANT4-DNA, each physical interaction is described by a physics process that can evoke several models. Those models can be complementary, in the way that they can be applied to different energy ranges, or alternative which means they cab be used in the same energy range.

GEANT4-DNA groups three types of physical models:

1. Theoretical models:

- Ionization by electrons with energy greater than 100 eV.
- Ionization and excitation by protons whose energy is greater than 500 keV and less than 100 MeV.
- Elastic scattering of electrons.

2. Semi-empirical models:

- Ionization by protons whose energy is less than 500 keV.
- Excitation by protons whose energy is less than 500 keV.
- Charge change of the proton. (decrease H^+ or increase H)
- Ionization by electrons whose energy is less than 100 eV (correction of theoretical model).

3. Models provided by experimental data:

- Molecular attachment of electrons.
- Vibrational and rotational excitation by electrons.

The theoretical models that are going to be used in a track structure code must be applicable over the full slowing down energy range (down to 10 eV for liquid water). In the version 10.2.p03 of GEANT4-DNA, electron processes include ionisation, electronic excitation, elastic scattering, vibrational excitation, and molecular attachment; for the different charge states of hydrogen and helium atoms, ionisation, electronic excitation, electron loss or capture, and elastic scattering are considered. A summary of the main models implemented in GEANT4-DNA (v10.02.03) is presented in Table. 1. For heavy ions of high charge and energy (HZE) (ex. C, N, O, and Fe), an effective charge scaling is performed from the same models as for proton.

The processes and models can be either specified by the user separately or a physics constructor like *G4EmDNAPhysics* can be used, which is a pre-built C++ class containing all particles, processes, and models [43]. There are other alternative constructors: G4EmDNAPhysics_option1, G4EmDNAPhysics_option2, G4EmDNAPhysics_option4, G4EmDNAPhysics_option5, G4EmDNAPhysics_option6
each one with improvements in the models implemented in GEANT4-DNA.

Among the improvements introduced in the inelastic cross sections to extend the range of validity to very low energies, the complex dielectric response of the material, $\varepsilon(E, \vec{q})$, was used, where E and $\vec{q}$ are the energy and momentum transfer, respectively. An exact calculation of this function is not possible for real materials so several approximations have been used. For the specific case of GEANT4-DNA, the determination of the dielectric response function is based on a semiempirical optical-data model described by Emfietzoglou [91] and some modifications to this model have been introduced by Incerti et al. [51] and Kyriakou et al. [92, 93]. The latter accounts for the binding energy threshold through a redistribution of the oscillator strength and refinements in the exchange and perturbation corrections to the Born approximations. This model is implemented by two classes: G4DNAEmfietzoglouExcitationModel and G4DNAEmfietzoglouIonisationModel, both included in the "option4" physics constructor.

For the description of elastic scattering, the screened Rutherford uses a screening parameter based on experimental data on nitrogen gas [94]. This model is the default option in GEANT4-DNA. A new model, based on experimental measurements in water, has been implemented using the screening parameter of Uehara et al. [38]. This class, named as G4DNAUeharaScreenedRutherfordElasticModel, is also available in the physics constructor "option4".

The inclusion of the Monte Carlo code CPA100 into GEANT4-DNA allows the use of alternative models for the simulation of atomic excitation, ionization, and elastic scattering: G4DNACPA100ExcitationModel, G4DNACPA100IonisationModel, and G4-DNACPA100ElasticModel as part of the physics constructor "option6" [50]. In order to take into account the molecular structure of targets, the CPA100 code implemented the Binary Encounter Bethe (BEB) model for the evaluation of the ionization single differential and total cross sections for each molecular sub-shell [95]. The BEB analytical expression does not need empirical parameters, only the binding and mean kinetic energies for each orbital and the number of electrons in each sub-shell. For elastic scattering, differential cross sections above 50 eV are determined using the independent atom model (IAM) by taking into account the scattering amplitude of each atom and their separation [96]. Below this energy threshold, the differential elastic scattering cross section is taken directly from experimental measurements of a solid water target [97].

Table 1 – List of GEANT4-DNA v10.02.03 physical processes and models. Source: [81]

Physical processes	**Process class**	**Model class**	**Energy range**
		e^-	
Elastic Scattering	G4DNAElastic	G4DNAChampionElasticModel	7.4 eV-1 MeV
Excitation	G4DNAExcitation	G4DNABornExcitationModel	9 eV-1 MeV
Ionization	G4DNAIonisation	G4DNABornIonisationModel	11 eV-1 MeV
Vibrational excitation	G4DNAVibExcitation	G4DNASancheExcitationModel (Michaud et al.)	2 eV-100 eV
Molecular attachment	G4DNAAttachment	G4DNAMeltonAttachmentModel (Melton et al.)	4 eV-13 eV
		H^+	
Elastic Scattering	G4DNAElastic	G4DNAIonElasticModel	100 eV-1 MeV
Excitation	G4DNAExcitation	G4DNAMillerGreenExcitationModel	10 eV-500 keV
		G4DNABornExcitationModel	500 keV-100 MeV
Ionization	G4DNAIonisation	G4DNARuddIonisationModel	100 eV-500 keV
		G4DNABornIonisationModel	500 keV-100 MeV
Charge change	G4DNAChargeDecrease	G4DNADingfelderChargeDecreaseModel	100 eV-100 MeV
		H	
Elastic Scattering	G4DNAElastic	G4DNAIonElasticModel	100 eV-1 MeV
Excitation	G4DNAExcitation	G4DNAMillerGreenExcitationModel	10 eV-500 keV
Ionization	G4DNAIonisation	G4DNARuddIonisationModel	100 eV-100 MeV
Charge change	G4DNAChargeIncrease	G4DNADingfelderChargeIncreaseModel	100 eV-100 MeV
		He^0, He^+, He^{2+}	
Elastic Scattering	G4DNAElastic	G4DNAIonElasticModel	100 eV-1 MeV
Excitation	G4DNAExcitation	G4DNAMillerGreenExcitationModel	1 keV-400 MeV
Ionization	G4DNAIonisation	G4DNARuddIonisationModel	
Charge change	G4DNAChargeDecrease (He^{2+})	G4DNADingfelderChargeDecreaseModel	1keV-400MeV
Charge change	G4DNAChargeIncrease (He^0,He^+)	G4DNADingfelderChargeIncreaseModel	1keV-400MeV
		C, N, O and Fe	
Ionization	G4DNAIonisation	G4DNARuddIonisationExtendedModel	0.5 MeV/u -10^6 MeV/u

2.3.4 Simulation of water radiolysis

The term "radiolysis" in general refers to the breakage of chemical bonds caused by the ionizing radiation action, which leads to the dissociation of molecules. The study of which chemical species are generated after the passage of the ionizing radiation through a biological system is an active research field, known as radiation chemistry. The importance of water in biological systems (its abundance in the cell) made it one of the first molecules to be studied. Then, water radiolysis refers to the process by which the dissociation of the water molecule occurs after being exposed to ionizing radiation. The origins of radiation chemistry dates back to the 19^{th} century [98] and the André Debierne's work, performed in 1914, showing that the water decomposition due to ionizing radiation action generates chemical species called radiolytic products, among which are free radicals. In the previous section §2.1, the first three stages that follow the irradiation of a biological system were briefly explained. These processes are not independent of each other. They are closely related and superposed in time scale. As was mentioned, the secondary electrons produced by ionizing radiation in the medium slow down to energies of the order of a few eV, which is nearly the threshold to produce electronic excitations in water, and during this slowing down process several discrete events occur according to the energy transferred to the medium. Depending on the energy transferred by the electrons, the molecule can undergo:

- Ionization (threshold in water $\sim$ 11 eV)
- Excitation (threshold in water $\sim$ 8 eV)
- Thermal transfer (vibrations, rotation, translation).

Thermal transfer is insignificant, a large dose would be necessary before the thermal effects became large enough to affect cellular biochemistry. The initial event is the energy transfer of about 7 - 100 eV, energy enough to produce ionizations and excitations. It was said that through ionizations and excitations three main chemical species are created in the local vicinity of the particle track: the ionized water molecule (a radical ion $H_2O\bullet^+$), an excited water molecule (H_2O^*) and a sub-excitation electron.

The chemical processes or reactions that occur during water radiolysis are complex and very dependent of the concentration of the different chemical species produced.

In addition, they also can vary greatly when introducing external elements. Among these elements, it is important to point out that the presence of oxygen increases the effect of the ionizing radiation, acting as a radiosensitizer. Indeed, in 1909 Schwarz recognized that applying pressure to skin and thereby decreasing blow flow and oxygen supply caused a reduction in radiation induced skin reactions. For low LET radiation in the absence of oxygen (hypoxic cells), the dose delivered to the tumor must be increased by a factor of 2.5-3 to obtain the same effect as in the oxygen-saturated situation. This is known as the Oxygen Enhancement Ratio (OER) and it measures the differential radiosensitivity of poorly versus well oxygenated cells.

The simulation of the water radiolysis in GEANT4-DNA is possible due to the development of a chemical module by Karamitros et al. [49, 99, 100]. In the following sections, a general description about how the physico-chemical and chemical stages are simulated in GEANT4-DNA is presented.

2.3.4.1 Simulation of physico-chemical stage in GEANT4-DNA

The physico-chemical stage considers all the electronic and atomic modifications resulting from readjustments of the medium to return to equilibrium. During this stage, ionized and excited water molecules can recombine or dissociate into new chemical species (see Figure 17). Two competitive processes for the ionized water molecule can be mentioned:

- Recombination with the electron that has just been ejected:

$$H_2O^+ + e^- \longrightarrow H_2O^* \tag{2.4}$$

- Dissociation into:

$$\begin{aligned}
&H_2O^* \longrightarrow H_2O + \gamma \\
&H_2O^* \longrightarrow H\bullet + \bullet OH \\
&H_2O^* \longrightarrow H_3O^+ + \bullet OH + e^-{}_{aq} \\
&H_2O^* \longrightarrow H_2 + \bullet OH + \bullet OH \\
&H_2O^+ \longrightarrow H^+ + \bullet OH \xrightarrow{H_2O} H_3O^+ + \bullet OH(H_2O \quad as \quad a \quad solvent)
\end{aligned} \tag{2.5}$$

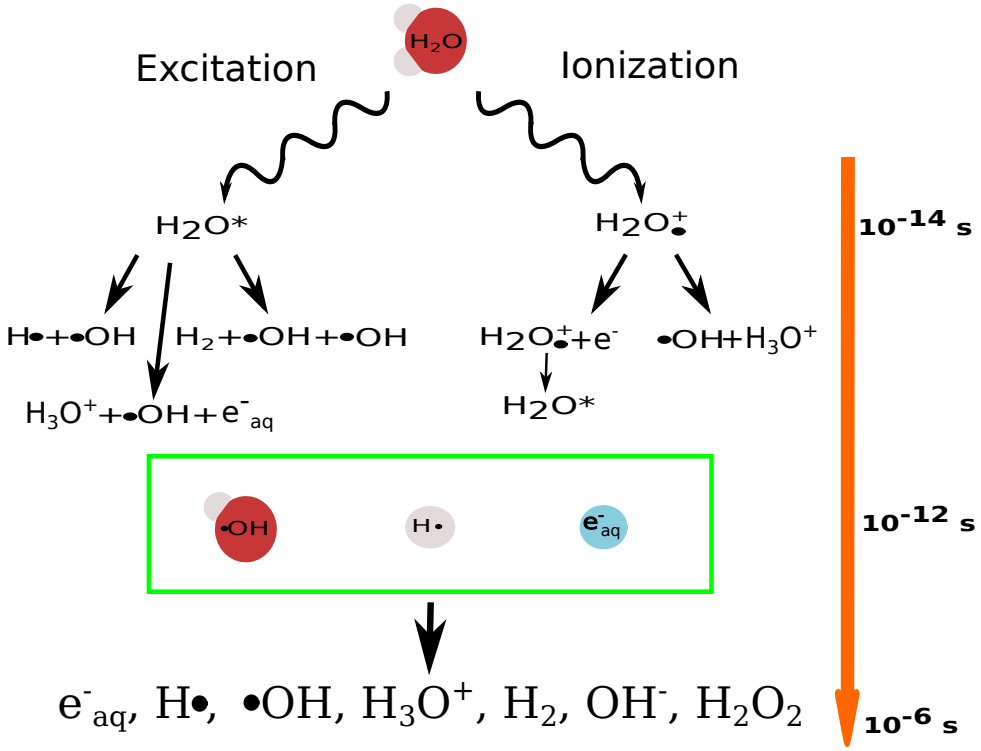

Figure 17 – After the physical stage the water molecule is left ionized or excited and began to dissociate or recombine into new chemical species during the physico-chemical stage which elapses between 10^{-14}-10^{-12} s after exposure approximately. During the chemical stage, 10^{-12}-10^{-6} s after exposure, the chemical species created began to diffuse and react.

In Table. 2 the different dissociation channels of the water molecule and their branching ratios implemented in GEANT4-DNA are presented. These values are the same as those used by PARTRAC [76].

The energy of the first excitation level is 8 eV. That is the reason why electrons with energies lower than this value are called sub-excitation electrons. They mainly suffer elastic interactions but also lose energy through vibrational and rotational interaction modes. The sub-excitation electrons are thermalized and captured by the surrounding water molecules becoming solvated:

$$e^- + nH_20 \longrightarrow e^-{}_{aq} \quad 10^{-12} \quad s \tag{2.6}$$

This capture is the result of the interaction between the negative charge of the electron and the electric dipole moment of the water molecules that surround it. The thermalization process is supposed to finish within one picosecond after irradiation.

Table 2 – Branching ratios of a water molecule at 1 ps. The values are the same as those used by the PARTRAC code, see [76]

Electronic State	**Dissociation Channel**	**Fraction(%)**
	Ionisation	
All ionisation states	$H_3O^+ + \bullet OH$	100
	Excitation	
A1B1 (1b1)$\longrightarrow$ (4a1/3s)	$\bullet OH + H\bullet$	65
	$H_2O + \Delta E$	35
B1A1 (3a1)$\longrightarrow$ (4a1/3s)	$H_3O^+ + \bullet OH + e^-{}_{aq}$	55
	$H_2 + 2 \bullet OH$	15
	$H_2O + \Delta E$	30
Rydberg, diffusion bands	$H_3O^+ + \bullet OH + e^-{}_{aq}$	50
	$H_2O + \Delta E$	50

There are two processes competing with thermalization that could also capture the electron. One is dissociative attachment that occurs when the electron is captured by one water molecule and the other is geminated recombination through which the electron is captured by an ionized water molecule. The term geminated means that this process occurs before any separation by diffusion occurs, just after the ionized water molecule is created.

2.3.4.2 Simulation of chemical stage in GEANT4-DNA

Approximately 10^{-12} s or 10^{-6} s after exposure the chemical species produced during the physico-chemical stage are going to diffuse randomly in the medium and can react each other. Initially, the radiolytic products are located around the energy deposits, with the passage of time they distribute more evenly in the medium. That is the reason why during the chemical stage two phases can be distinguished: a heterogeneous one at the beginning of the stage (10^{-12} s) when all the products are distributed in the form of clusters around the deposits and the homogeneous phase, starting from about 10^{-7} s.

The initial spatial distribution depends on the LET of the incident particle [26, 76]. For low LET high energy electrons, the initial distribution of chemical species in the form of clusters will be sparse. This is not the case for protons and α-particles, whose LET is higher and therefore the spatial distribution of species is dense. Normally, most of

the reactions between different clusters is complete after 1 μs. Moreover, during the first instants of diffusion the probability of encounter of the chemical species is high and as time passes and they diffuse in the medium, the chance of encounter decreases.

The simulation of the chemical stage in GEANT4-DNA is based on the particle-continuum (or particle based) representation. In this representation, the molecules are considered "particles" evolving in a "continuum" which is liquid water. The particle continuum approach computes the positions of the individual molecules as a function of time. This representation is associated with a stochastic treatment of the diffusion of molecules based on the Smoluchowski theory and the Ermak-McCammon equation. This choice is adapted to systems whose reagents are distributed in a heterogeneous manner and whose life times do not exceed the order of microseconds.

Some assumptions are made in the code:

- The medium is isotropic and consists of liquid water.
- Each pair of molecules diffuse independently of each other.
- The molecules are spherical.

The diffusion is simulated by means of a succession of time steps during which all molecules are moved simultaneously. For each time step, an analysis of the position of each molecule is performed and if a reaction partner is found within the reaction radius, the molecules react. If a reaction occurs, the reactants are replaced by the reaction products.

The transportation method implemented in GEANT4-DNA uses the following equation:

$$x(t_0 + \tau) = x_0 + \frac{F(x_0)}{m\gamma}|_{t_0}\tau + \sqrt{2D\tau}N(0,1) + O(\tau^2), \tag{2.7}$$

where t_0 is the initial time; τ, the time step; x_0, the initial position; $F(x_0)$, a force applied to the molecule; m, the mass of the molecule; $\gamma = \frac{\alpha}{m}$, the friction coefficient; α, a positive constant related to the viscosity expressed in $[mass][time]^{-1}$; D, the diffusion coefficient; and $N(0,1)$, a random number described by a Gaussian distribution of mean 0 and variance 1. This equation is known as Ermak-McCammon method [101]. With this expression the Brownian character of the molecule's displacement can be reproduced.

Since in GEANT4-DNA no external forces are considered, then the previous equation can be simplified:

$$\frac{F(x_0)}{m\gamma}|_{t_0} = 0 \tag{2.8}$$

As a result, the equation 2.7 can be rewrite as follows:

$$x(t_0 + \tau) = x_0 + \sqrt{2D\tau}N(0,1) + O(\tau^2) \tag{2.9}$$

Replacing τ by dt and using equation 2.9 in three dimensions:

$$\left.\begin{cases} dx = \sqrt{2Ddt}\ N_x(0,1) \\ dy = \sqrt{2Ddt}\ N_y(0,1) \\ dz = \sqrt{2Ddt}\ N_z(0,1) \end{cases}\right\} \Longrightarrow \vec{r}_{t+dt} = \vec{r}_t + \begin{pmatrix} dx \\ dy \\ dz \end{pmatrix} \tag{2.10}$$

The molecule position at $\vec{r}_{t+dt}$, after a time step dt, is calculated using equation 2.10 and sampling the random numbers $N_x(0,1)$, $N_y(0,1)$ and $N_z(0,1)$. To evaluate the new position of the molecule, we need the diffusion coefficient and the time step used in the simulation. Table. 3 shows the diffusion coefficients used in GEANT4-DNA. The time step is calculated at the beginning of each step.

Table 3 – Diffusion coefficients for the species in question as described by Kreipl et al. [76]

Chemical Species	**Diffusion coefficients D** $(10^{-9}m^2s^{-1})$
e^-_{aq}	4.9
$OH\bullet$	2.8
$H\bullet$	7.0
H_3O^+	9.0
H_2	4.8
OH^-	5.0
H_2O_2	2.3

There are two methods commonly used to calculate the time step. One is to select a fixed time step value and the other uses dynamic time steps. In GEANT4-DNA, a combination of both methods was implemented. Kreipl et al. [76] proposed to predefine static time steps as a function of the virtual time. As a result, some predefined time steps

are set accounting for the spatial distribution of chemical species at the beginning of the chemical simulation (at 1 ps) and according to the time evolution (see Table 4).

Table 4 – Time steps Δt with respect to the physical time, as described by Kreipl et al. [76]

Time interval (s)	$\boldsymbol{\Delta t}$ (ps)
Until 10^{-11}	0.1
10^{-11}-10^{-10}	1
10^{-10}-10^{-9}	3
10^{-9}-10^{-8}	10
Above 10^{-8}	100

Implementing only a fixed time step method has some disadvantages. For instance, if a large time step is chosen some reaction could be missed and if it is too small, this results in a waste of computing time for simulating Brownian motions. A solution to this was to include a dynamic time step method, which defines time steps for a given pair of reactants during which no reaction happens, starting from the 1D solution of the Smoluchowski equation. A confidence interval of 95% is used. The time step calculated with this method can be seen as "maximum diffusion time without reaction" or as a minimum encounter time t_D during which a chemical species may encounter a reactant with about 5% confidence interval [49]. Then, this method to calculate the time step will allow the species to diffuse larger distances before reacting.

The Smoluchowski equation gives us the probability density function $p(x,t)$ for finding a molecule "A" with respect to a molecule "B" at the distance x at time t. In 1D it reads:

$$\frac{\partial p(x,t)}{\partial t} = -\frac{\partial}{\partial x}\left(\frac{F}{m\gamma}\, p(x,t)\right) + \frac{\partial^2}{\partial x^2}\Big(D\; p(x,t)\Big), \tag{2.11}$$

where F is an external force; m, the mass of the molecule; γ, the friction coefficient; and D, the diffusion coefficient. Equation 2.11 is valid only for time scales larger than the speed's relaxation time, when the speed (v) has time to relax and the acceleration (a) can be neglected:

$$\gamma v \gg a \tag{2.12}$$

Equation 2.11 in three dimensions is given by:

$$\frac{\partial p(\vec{r},t)}{\partial t} = \vec{\nabla} \cdot [D[\vec{\nabla} p(\vec{r},t) - \beta \vec{F}(\vec{r}) p(\vec{r},t)]], \tag{2.13}$$

where $\beta = \frac{1}{k_B T}$, k_B is the Boltzmann constant, and T is the temperature. To obtain this equation the Einstein-Smoluchowski relation has been used (applicable when D is space independent):

$$D = \frac{k_B T}{m\gamma} \tag{2.14}$$

The solution to this equation 2.13 is:

$$p(\vec{r},t) = \frac{4\pi r^2}{(4\pi D t)^{3/2}} e^{-\frac{r^2}{4Dt}}. \tag{2.15}$$

The minimum encounter time t_D is determined by the separation distance of the pair of reactants minus the reaction radius. The solution to the Smoluchowski equation in 1D without an external force and with D independent of the position, is given by:

$$p(x,t) = \frac{1}{\sqrt{4\pi D t}} e^{-\frac{x^2}{4Dt}}. \tag{2.16}$$

The cumulative distribution function is:

$$\begin{aligned} P(x \leq r) &= \int_0^r p(x,t)dx \\ &= \left[erf\left(\frac{x}{2\sqrt{Dt}}\right)\right]_0^r \\ &= erf\left(\frac{r}{2\sqrt{Dt}}\right), \end{aligned} \tag{2.17}$$

knowing that:

$$erf(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-s^2} ds. \tag{2.18}$$

Defining the initial separation distance between molecule "A" and "B" as d_0, r_{AB} as the reaction radius, and d_{Amin} and d_{Bmin} as the diffusion lengths of molecule "A" and "B", then:

$$d_0 - r_{AB} = d_{Amin} + d_{Bmin} \tag{2.19}$$

By selecting a 95% of confidence interval that the molecule diffuse without reactions:

$$P(x \leqslant d_{min}) = 0.95 \tag{2.20}$$

and since $erf(\sqrt{2}) \simeq 0.95$, then using eq. 2.17, it can be found that:

$$\begin{aligned} erf\left(\tfrac{d_{min}}{2\sqrt{Dt_D}}\right) &\simeq 0.95 \\ \frac{d_{min}}{2\sqrt{Dt_D}} &= \sqrt{2} \\ d_{min} &= 2\sqrt{2Dt_D} \end{aligned} \tag{2.21}$$

By inserting eq. 2.21 in eq. 2.19, we obtain:

$$\begin{aligned} d_0 - r_{AB} &= 2\sqrt{2D_A t_D} + 2\sqrt{2D_B t_D} \\ t_D &= \frac{(d_0 - r_{AB})^2}{8(\sqrt{D_A} + \sqrt{D_B})^2} \end{aligned} \tag{2.22}$$

Using eq. 2.22, it is possible to estimate an approximate minimum diffusion time before the encounter of each reactant for each step. It is important to notice that for each reaction the minimum encounter time is given by the closest pair of reactants. Then, by searching for each reaction, for the closest pair of reactants, and finally for the corresponding diffusion time, it is possible to predict which reaction is going to occur next.

The dynamic time step has a disadvantage, which is that before the reaction occur, the pair of reactants will have to make smaller steps increasing the computation time. In order to avoid this issue, a constraint t_{limit} in the computed time step is introduced for each reactant pair. If t_D is smaller than t_{limit}, t_{limit} is chosen as the next time step.

Any couple that meets the condition $t_D < t_{limit}$ is saved and each molecule is diffused with t_{limit}. The values used for t_{limit} were the same proposed by Kreipl et al. [76] (see Table 4). At the end of the step, the separation distance of all recorded couples is checked. If this distance is greater than the reaction radius the probability of encounter during t_{limit} is determined using the Brownian bridge technique.

In order to determine the minimum diffusion time from eq. 2.22, it is still necessary to calculate r_{AB}. The reactions modeled in GEANT4-DNA are diffusion controlled. For a bimolecular reaction, of second order, like:

$$A + B \longrightarrow C,$$

the temporal evolution of the concentrations of each molecule is given by the following equation:

$$\frac{d[A]}{dt} = \frac{d[B]}{dt} = -k[A][B], \tag{2.23}$$

where k is the reaction rate constant of the complex formation and depends on the environment, the temperature, and the pressure. The reaction rate constant is related to the reaction radius through the following expression:

$$\begin{aligned} k &= 4\pi N a D r_{AB} \\ r_{AB} &= \frac{k}{4\pi N_A D}, \end{aligned} \tag{2.24}$$

where D is the sum of the diffusion coefficients of the two reactants and N_A is the Avogadro's number. In Table 5 the reaction rate constants used in GEANT4-DNA are presented, as taken from reference [76].

Brownian bridge technique

As mentioned above, GEANT4-DNA introduces several restrictions to the time step selected, using both static and dynamic methods. In general, if two unique molecules are close enough a very small time step is imposed and for this reason, the limit time is used. When the global simulation time is in one of the intervals shown in Table 4, the time step chosen is the corresponding t_{limit}. The t_{limit} values adopted in this work for

Table 5 – Reaction rates as described by Kreipl et al. [76]

Reactions	**Reaction rate** $(10^{10} M^{-1} s^{-1})$
$H\bullet + e^-_{aq} + H_2O \longrightarrow OH^- + H_2$	2.65
$H\bullet + OH\bullet \longrightarrow H_2O$	1.44
$H\bullet + H\bullet \longrightarrow H_2$	1.20
$H_2 + OH\bullet \longrightarrow H\bullet + H_2O$	4.17×10^{-3}
$H_2O_2 + e^-_{aq} \longrightarrow OH^- + OH\bullet$	1.41
$H_3O^+ + e^-_{aq} \longrightarrow H\bullet + H_2O$	2.11
$H_3O^+ + OH^- \longrightarrow 2H_2O$	14.3
$OH\bullet + e^-_{aq} \longrightarrow OH^-$	2.95
$OH\bullet + OH\bullet \longrightarrow H_2O_2$	0.44
$e^-_{aq} + e^-_{aq} + 2H_2O \longrightarrow 2OH^- + H_2$	0.50

each interval of the global simulation time were those proposed by Kreipl at al. [76]. The previous approach does not account for the fact that those molecules could react during the small time step "ignored" and replaced by the t_{limit}. The Brownian bridge technique [102] takes this into account. It determines the probability for finding two reactants separated by a distance less than r_{AB} at a time t, given that their pre-step and post-step separation distances are both greater than r_{AB}. However, that probability is approximate because the problem is reduced to one dimension [49]:

$$P_{Br}(dt) = exp\frac{-(x_f - r_{AB})(x_0 - r_{AB})}{D_{A+B}dt}, \tag{2.25}$$

where D_{A+B} is the sum of the diffusion coefficients of the two molecules considered; r_{AB}, the reaction radius; x_0, the separation distance in the pre-step position; and x_f, the separation distance in the post-step position.

3 Methods

3.1 GEANT4-DNA application

In this work an application named "phychem" was developed using the GEANT4-DNA extension of the general purpose GEANT4 code on its version 10.2.P03, which provides a full set of models for the physical, physicochemical, and chemical stages of the interaction of ionizing radiation with liquid water.

In summary, the following classes were implemented in the application:

- ***RunAction*** This is a class normally used in GEANT4 applications that manages actions at the beginning and end of every run.
- ***PrimaryGeneratorAction*** A mandatory class used for specifying the radiation source.
- ***DetectorConstruction*** A mandatory class where the detector geometry and material are specified. Also, a *W*orld geometry that encompasses all the simulation geometries is defined.
- ***PhysicsList*** It is also a mandatory class with specifications about all the particles, physics processes, and cut-off parameters.
- ***StackingAction*** It customizes the access to the track stacks.
- ***SteppingAction*** This class is where physical steps are managed and also recorded to the output file for the physical stage.
- ***TimeStepAction*** This class is where chemical diffusion steps are managed and also recorded to the output file for the chemical stage. It also allows to access the information related with the reactions taking place during this stage.
- ***EventAction*** It handles actions at the beginning and end of every event.
- ***TrackingAction*** It handles actions at the creation and completion of every track.

In the *phychem* application, the user has to provide an input file containing the information about the particle beam, energy, number of initial events, and time cut for the chemical stage. In the case that the user does not provide this input file, that data should be updated in each corresponding class. Any change to the detector geometry or the source needs to be performed in the *DetectorConstruction* and *PrimaryGeneratorAction* classes, respectively. The output was classified according to the stage, one for the physical stage or direct effects and another for the chemical stage or indirect effects. Initially, the output files created at the end of each run had an ASCII (human-readable) format, but the amount of information generated during the chemical stage made necessary the use of a binary format. At the end of each run, according to the number of computer processes used during the simulation, several physical and chemical output files are generated. All physics and chemical output files are merged into one final binary file per stage (physics and chemistry phase space files) using the external routine *rankfile*. The event number in these files is ordered in an increasing and consecutive manner.

In the physics output file, the information about the primary event, the primary or secondary particle that created the energy deposit, physical process (physical interaction in GEANT4-DNA), position of the particle (x,y,z), particle energy, energy deposit, and energy transfer inside the target volume are recorded. For the chemical output file, data containing the process rank, primary event, chemical species type, chemical species identification number, current position (x,y,z), and the global time (time since the event was created) are registered.

3.2 MPI interface and parallel programming in GEANT4

GEANT4-DNA supports parallel programming using an interface called *G4MPI* with the Message Passing Interface (MPI) libraries. This feature has been extended to provide hybrid applications that use MPI and multithread (MT). MPI can be implemented on a wide variety of parallel computing architectures. The standard defines the syntax and semantics of a core of library routines writing portable message-passing programs in C, C++, and Fortran. To efficiently reduce the memory consumption it is possible to schedule a single MPI job for each node and use GEANT4 multithreading capabilities to scale across the CPUs and cores available on the node.

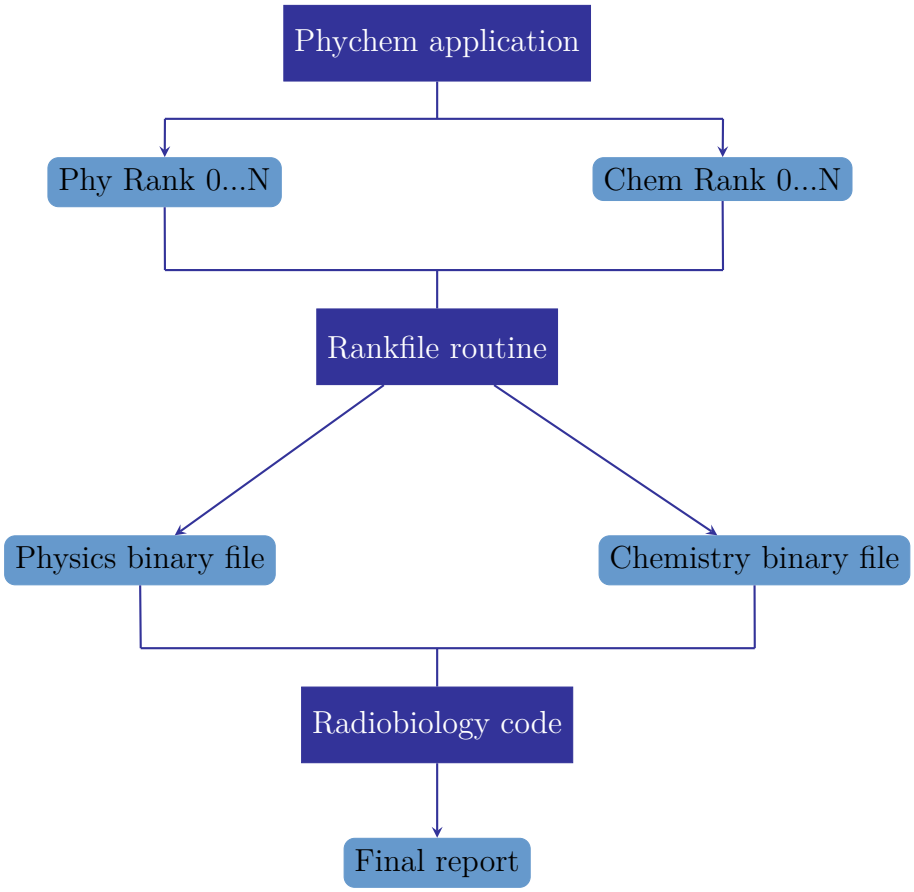

Figure 18 – Workflow of this thesis work. The phychem application creates for each process a physics and a chemistry output file. The physical outputs are merged in a final file for the physical phase space and the chemical outputs are merged into one file that contains the information related to the chemical stage. Both files are analyzed by the radiobiological code which gives the final report.

In order to activate MPI in GEANT4, two instances of the classes *G4MPImanager* and *G4MPIsession* must be created. The user interaction is performed through the usual GEANT4 UI commands: /run/beamOn/, this command will trigger the MPI processes to cooperatively perform the simulation. MPI performs the steering of the job. The *rank0* process accepts all the UI commands set by the user, distributes the work among all available processes, and controls the random number generator seeds. Currently, the MPI interface also supports user analysis tools like ROOT, AIDA, XML, and CSV). This g4analysis package is thread-safety and it is also able to automatically produce histograms at the end of the job. The implementation of parallel programming allows to reduce the execution time as compared to serial applications.

3.2.1 Cluster specifications

The simulations performed were carried out in the *mc*simulations computer cluster of the Física Radiológica Médica (GFRMd) research group, located at the Department of Applied Physics of the Gelb Wataghin Insitute of Physics, Unicamp. This cluster runs over a openSUSE 13.1 (Bottle) operating system of 64 bits and has a total of 124 CPU cores and 576 GiB of toal RAM. There are 9 nodes available, four of them are Intel(R) Xeon(R) CPU E5-2650 0 @ 2.00GHz and the others are Intel(R) Xeon(R) CPU E5-2620 v3 @ 2.40GHz. The primary node or server has 3 Tb of storage capacity. All nodes are linked by a 1 Gb/s Ethernet network. The utilization, scheduling, and administration of the cluster is managed by the TORQUE v6.0.1 system. This system controls the job submission process and manages the use of the available computer nodes in the cluster.

For the purpose of this work only GEANT4-DNA MPI capabilities were implemented. The main reason why we did not use the MPI combined with MT capabilities is related to the format of the output files. Indeed, when using g4analysis tools only a few possibilities for the file format were available (normally ASCII files or CSV) for the GEANT-DNA version used in this work. On the one hand, the huge size of the chemical output demanded the use of the binary file format for all the output files. On the other hand, taking into account that the inclusion of the simulation of the chemical stage introduces a great degree of complexity in the simulations and that this module is under development in GEANT4-DNA, it was better considered to generate a different output file for each node and not to use the merging capabilities that g4analysis offers. All output files were tagged using the parallel process rank (identification number). The argument of the */run/beamOn/* command in the input file corresponds to the initial number of events to be run by each separate process.

3.3 GEANT4-DNA. Simulation parameters

For the purpose of this work, the physics constructor *G4EmDNAPhysics*, a prebuilt C++ class, containing all particles, processes, and models was employed [43]. In this class, a complete set of models are provided for describing step by step electromagnetic interactions of electrons, all charged states of hydrogen and helium ions, and some other heavy ions with liquid water. Some of these particles, like electrons, can be tracked down

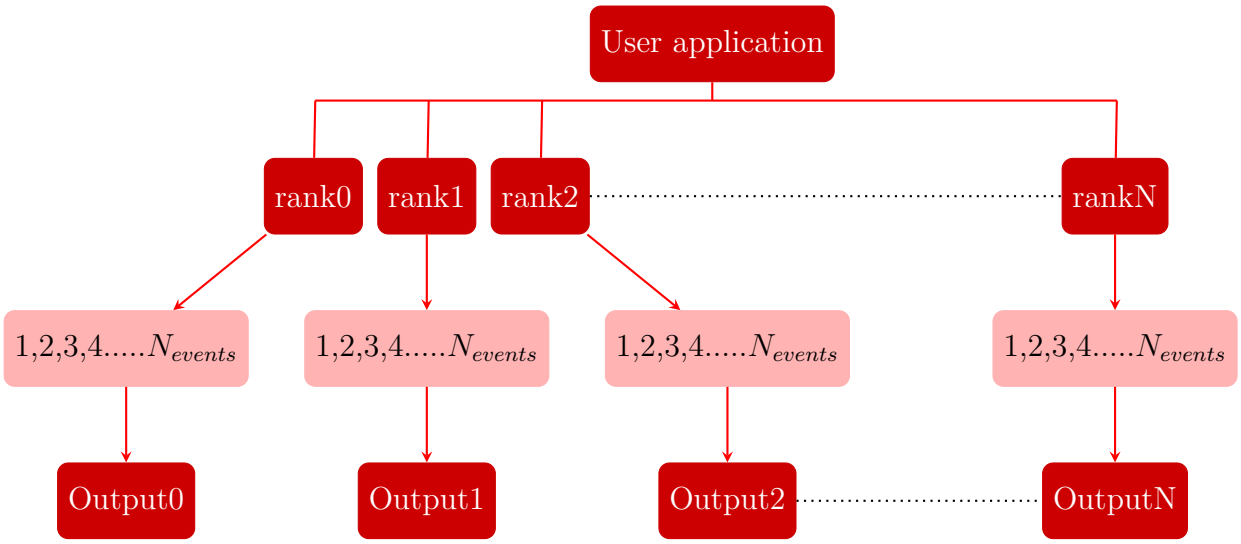

Figure 19 – MPI implementation in GEANT4-DNA. The User application is distributed in the available nodes and processes. Each process is going to manage N_{events} which is the argument in the UI command /run/beamOn. The rank is the identification number associated to the computer process. The "Out" refers to the binary output files for the physical and chemical stage.

to the eV scale, for more details see Section. §2.3.3.

The models for the physicochemical and chemical stages in GEANT4-DNA use a particle-continuum representation, where only the species in which we are interested are explicitly modeled as point-like objects diffusing in the medium through Brownian motion, while the rest of the medium content is treated as a continuum [49]. At first, the dissociation or relaxation of water ions and molecules is processed selecting a dissociation channel according to the corresponding branching ratio. It is considered that the chemical stage starts 1 picosecond after shooting the primary event. All dissociation and relaxation products diffuse using the Smoluchowski model, which assumes that those Brownian particles have reached a thermal equilibrium. After selecting a time step, each species is diffused and placed in a new position so that species "jump" from one position to another. If two chemical species are separated by a distance smaller than a reaction radius the corresponding chemical reaction occurs. Such reactions are known as diffusion-controlled reactions. Both pre-chemical and chemical stages use the chemistry constructor *G4EmDNAChemistry* in which all the necessary parameters for the simulation of water radiolysis are included, see Section. §2.3.3. The global time limit for the duration of the chemical stage was set to 1 ns.

It is important to point out that when all the chemical output files are merged

into one file, the identification number of the chemical species could be repeated. This is related to the way in which MPI deals with the loop over all the chemical tracks for each process. Moreover, in some cases the same identification number could be assigned to a different chemical species type. To avoid this the process rank number was saved in the final binary file for the chemistry stage, in order to correctly manage the chemical data in the radiobiological code.

3.3.1 Simulation geometry

The simulation geometry consisted in a water cell nucleus of 3.085 μm $\times$ 3.085 μm $\times$ 3.383 μm placed inside a *World* region also full of water with 10.0 μm $\times$ 10.0 μm $\times$ 12.0 μm side lengths, see Figure 20. Both geometries were defined using the G4Box constructor built in GEANT4. The separation distance between the source (irradiation in Z and X direction) and the detector was equal to the thickness of a Mylar foil.

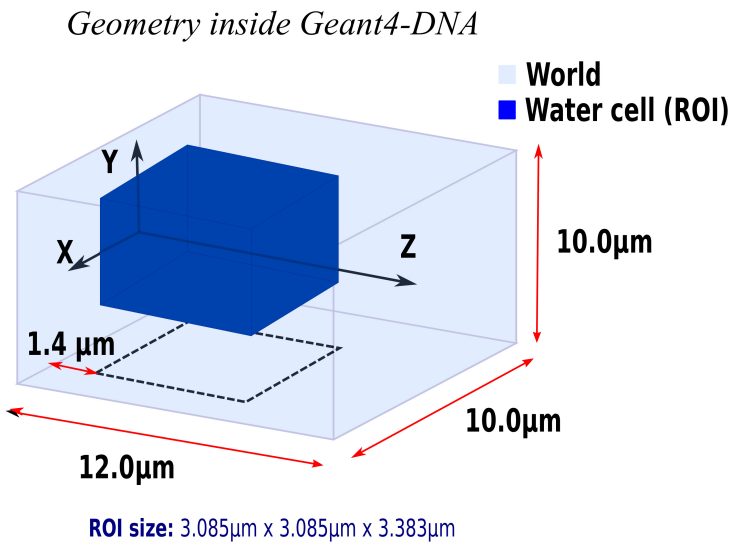

Figure 20 – Simulation geometry using GEANT4-DNA: the dark and light blue boxes represent the cell nucleus and the world region, respectively. The separation distance for the hypothetical Mylar foil (1.4 μm) is left between the source plane and the nucleus.

3.3.2 Geometry Consideration in the GEANT4-DNA code

The target geometry in the GEANT4-DNA code was represented as a parallelepiped cell full of water, since the biophysical model is described in an external code,

then it was necessary to put some restrictions to the regions that are occupied by the atoms of the DNA model. These geometrical restrictions consisted in avoiding that chemical species were produced inside the region occupied by the chromatin fiber. For each chemical species at its production time, the initial positions inside the ROI were transformed to the chromatin fiber reference system. The cylindrical shell occupied by the chromatin fiber was delimited from the one that only contains water and the former was considered as the "forbidden or restricted region" (see Fig. 21). All the chemical species that are created inside forbidden regions are discarded at the first time step and no longer followed. Afterwards, all chemical species are allowed to diffuse throughout the whole simulation world.

- The index of the closest chromatin fiber to the chemical species position was determined:

$$\begin{aligned} IX &= 1 + (posX - X_{min})/\Delta X \\ IY &= 1 + (posY - Y_{min})/\Delta Y. \end{aligned} \tag{3.1}$$

- The new position in the chromatin fiber reference system is:

$$\begin{aligned} X &= posX - X_{min} - (IX - 0.5)\Delta X \\ Y &= posY - Y_{min} - (IY - 0.5)\Delta Y. \end{aligned} \tag{3.2}$$

- Finally the following condition to check if the first step was inside the cylindrical shell is tested:

$$R_{int} < \sqrt{X^2 + Y^2} < R_{ext}. \tag{3.3}$$

In these equations, posX and posY are the global coordinates for each chemical track diffusion step, X_{min} and Y_{min} are the minimum X and Y coordinates for the grid cell geometry, ΔX and ΔY are equal and they match the external diameter of the chromatin fiber, and R_{int} and R_{ext} are the internal and external radius of the cylindrical shell.

This geometrical condition must be combined with the time step consideration, since we only wish to discard the species created inside that region. Initially, only the *Local time* (in GEANT4 this is the time since the track was created) was selected. However, the first test runs setting 0.1 ps as the first time step were not successful because the reaction products had an associated initial time of 0 ps. The solution for this was to add a status

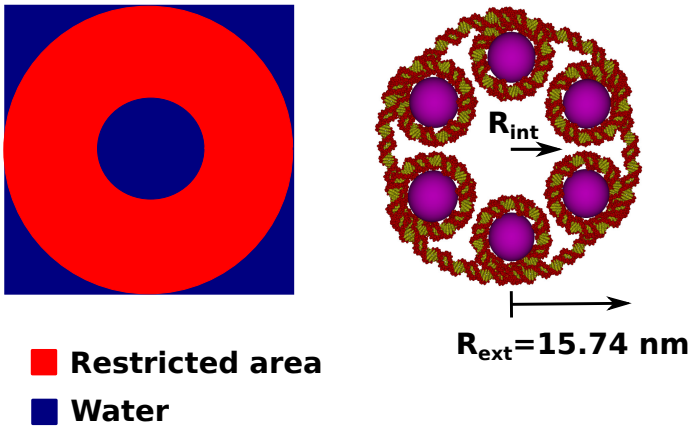

Figure 21 – Schematic representation of the allowed regions (Blue one) and the restricted areas or forbidden regions (red one) in the chromatin fiber system for the first diffusion time step of each chemical species. The dimension of the internal radius R_{int} of the cylindrical shell is 5.28 nm and of the external radius R_{ext} is 15.74 nm Only a grid cell corresponding to a single chromatin fiber is shown. This grid cell is repeated throughout the cross sectional area of the ROI.

variable that changes whenever a chemical species shows up, in order to evaluate whether the first appearance was inside the cylindrical shell or not.

This consideration allowed to obtain damage yields close to those reported in the literature for the LET range studied and also reduced the output files size in a way that less storage space was needed.

3.3.3 Irradiation setup

It is important to mention that considering the models included in GEANT4-DNA, the user could define any primary generator using the traditional GEANT4 source definition. For the purposes of this work, track structures for protons with energies of 0.5, 1, 2, 3, 4, 5, 7, 10, 15, 20, and 30 MeV and α-particles with energies of 2, 4, and 10 MeV were generated in liquid water using the *'G4_WATER'* NIST material with a density scaled to 1.0 g/cm^3. Those beam qualities were selected because this research is particularly focused to the hadrontherapy application. The LET for each beam was estimated at the center of the cell nucleus, those values are shown in Table 6.

Two irradiation setups were employed: one with the particle beam direction parallel to Z-axis and the initial position in the XY plane randomly sampled and the other with the particle beam direction parallel to the X-axis sampling the initial position in the YZ plane. In both cases, the separation distance between the source plane and the target

Table 6 – LET values for each particle beam simulated in this work. The LET for each beam was estimated at the center of the cell nucleus.

Energy (MeV)	**LET** ($keV/\mu m$)
Protons	
0.5	50
1	28
2	16
3	12
4	10
5	8
7	6
10	5
15	3
20	3
30	2
α-particles	
2	193
4	110
10	54

volume was 1.4 μm to mimic a Mylar foil commonly used in experiments. Using these irradiation setups, the two possible extreme cases were considered in our simulations, one irradiation oriented parallel to the chromatin fiber axis and the other normal to the fiber axis.

3.4 Radiobiological code

This is a code written in FORTRAN [1] for radiobiological purposes. The first version allowed the calculation of direct damage by superposing a phase space file containing the information of all the energy depositions onto the geometrical model. If the deposition event occurred inside the cell nucleus and the energy transfer was greater than 8 eV, the corresponding position was transformed to the chromatin fiber coordinate system. Then, using the "find the closest atom algorithm", a SB was counted if the deposition took place within one van der Waals radius around any of the sugar-phosphate group atoms. For each SB detected, the corresponding bp index, hit position, event number (history) that generated it, among other information, were saved.

For the statistical analysis, the total number of simulated events was equally

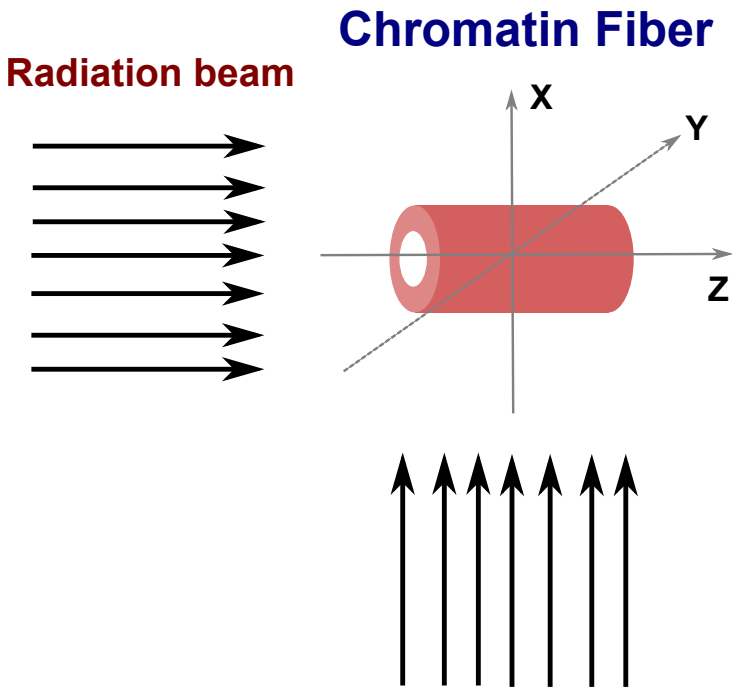

Figure 22 – Irradiation setups: parallel to the chromatin fiber axis (Z orientation) and perpendicular to the chromatin fiber axis (X orientation).

distributed into batches with the same deposited dose inside the cell geometry. For each quantity, the average value over all batches was reported.

3.4.1 DNA model

This code uses an atomic-resolution geometrical model for the B-DNA which has been described in detail in Ref. [1]. It has a nucleotide pair as the basic unit to build up the whole genome structure. In total, there are 63 atoms per nucleotide pair and their correspondent positions were reported elsewhere [1, 103]. By twisting 36 degrees and displacing 0.33 nm each bp, the double helix is built up. Each nucleosome is formed by two DNA loops containing 154 bp around a sphere that emulates a histone. The chromatin fiber is formed by 6 nucleosomes per level, disposed in a supra-helix with 7.11 nm pitch and an external diameter of 31.38 nm. Two adjacent nucleosomes are bound by a DNA fragment containing 46 bp. This fragment, when projected on the plane normal to the chromatin axis, forms an arc with angular aperture of 60°and 14.5 nm radius.

The whole cell nucleus has the same dimensions as the target cell described before. This nucleus was divided into 49 chromosome domains, each one containing 14x14 parallel chromatin fibers. The axis of the chromatin fibers is parallel to the Z direction. The total amount of DNA base pairs contained in the cell nucleus is 5.47 Gbp. See Figure

23 for details.

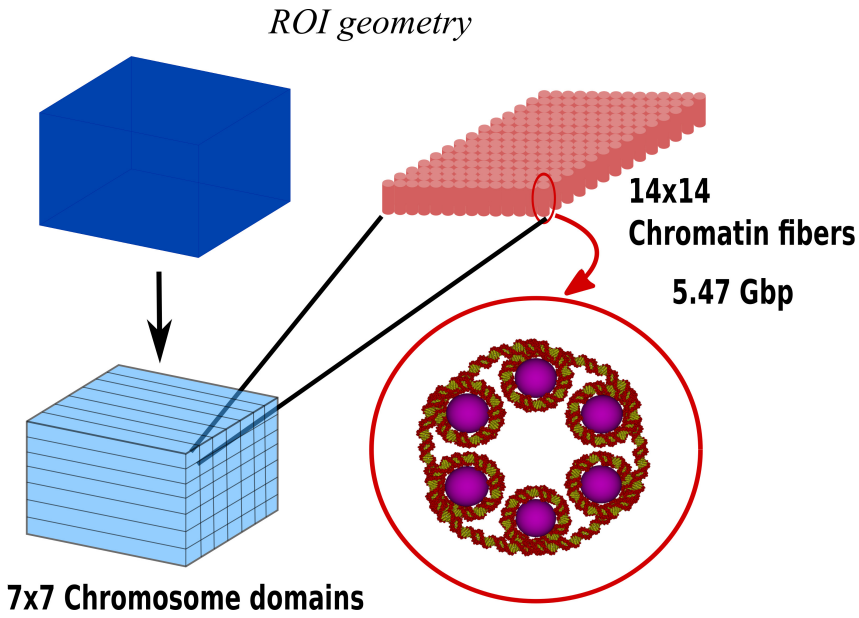

Figure 23 – Geometrical model implemented in the radiobiological code. The cell nucleus is divided in 7×7 chromosome domains, each one containing 14×14 chromatin fibers in the B-DNA conformation [1].

3.4.2 Indirect damage calculation

In order to account for the indirect damage, the existing FORTRAN code for radiobiological purposes was updated. The first step is to provide the code with the two ouput files obtained with GEANT4-DNA, for the physical and chemical stages. This new version was divided in two main blocks, the first one for direct damage analysis followed by a second one for indirect damage analysis. It is important to note that a correspondance should be made between the information contained in the physics and chemistry file. To accomplished this, the event or history number is used to match one stage with the other for each batch. The second information that needs to be entered in the code is the number of histories or events per batch in order to allow divide the total information for statistical analysis.

Following a similar procedure used in the physical stage, the output file for the chemical stage (chemistry phase space file) was superposed onto the geometrical model. There are several species that can be generated during the chemical stage in GEANT4-DNA see Table 3 in Section §2.3.4.2. All types of chemical species were included in the chemical phase space file, but so far only the DNA damage produced by the hydroxyl radical ($\bullet OH$) was taken into account since this is the most active species that induces

damage in the DNA [104]. To select only the $\bullet OH$ radical the tag for chemical species type was used.

The subroutine "Find the Closest Atom" was adapted to consider the indirect action. The critical parameter for this stage is the reaction radius. It was calculated following the Smoluchowski formalism used by the physico-chemical and chemical modules in GEANT4-DNA [49], which consider diffusion controlled reactions (see Section. §2.3.4.2). Water radicals can interact with any reactive site they encounter in their diffusion path separated by a distance lower or equal to the reaction radius. From the Equation. 2.24 we have:

$$k = 4\pi Na(D_A + D_B)r_{AB} \tag{3.4}$$

Considering that the diffusion coefficient for the atom of the DNA molecule is zero [3, 26, 27] since the target is stationary, the reaction radius r_{AB} is obtained as follows:

$$r_{AB} = \frac{k}{4\pi N_A D_{\bullet OH}} \tag{3.5}$$

where k is the reaction rate constant taken from Buxton et al. [75], $D_{\bullet OH}$ is the diffusion coefficient for the hydroxyl radical in this specific case, and N_A is Avogadro's number.

For the diffusion coefficient, the same values used in PARTRAC [26, 27] and by Karamitros et al. [49] for the chemical module in GEANT4-DNA were adopted, see Table. 3.

Hydrogen abstraction in the sugar-phosphate backbone (sites 5'H,4'H,3'H,2'H,1'H) was the reaction considered for the $\bullet OH$ radical. The reaction radius obtained was 0.12 nm, using the reaction rate constant for the deoxyribose site (see Table. 7). The same value was reported by Tartier et al. [104] and Aydogan et al. [105].

In the chemical section of the code, if the diffusion step for the track is inside the cell nucleus and this track is an $\bullet OH$ radical, the subroutine to find the closest atom is called. The subroutine now takes two new input parameter besides the coordinates of the diffusion step in the chromatin fiber system, one serves to identify in which stage the subroutine was called and the other, to enter the chemical species ID type in order to

Table 7 – Reaction constants taken from Buxton et al. [75] for the reactions taken place between DNA bases and sugar-phosphate with water radicals

DNA Site	**Reaction Rate constant** $(10^9 M^{-1} s^{-1})$
Hydroxyl radical	
2-deoxyribose	2.5
Adenine	6.10
Cytosine	6.10
Guanine	9.20
Thymine	6.40
Hydrogen radical	
2-deoxyribose	0.03
Adenine	0.09
Cytosine	0.50
Guanine	0.50
Thymine	0.10
Hydrated electron	
2-deoxyribose	0.01
Adenine	9.00
Cytosine	13.00
Guanine	14.00
Thymine	17.00

associate the correct reaction radius in the look-up table inside the subroutine. In this work, only one chemical species was processed but the use of this "type ID" as input parameter allows a further extension of the code to study the DNA damage caused by other chemical species.

In the subroutine, two major changes were introduced, the first one discards those diffusion steps that are inside the histone region and therefore are assumed to be scavenged. If this happens, no analysis of the distance to the nearest atom in the sugar-phosphate group is performed and the subroutine gives back a parameter that allows save the track ID, the current rank, and event in order to not consider the following steps. The other change performed is related to the consideration of the reaction radius and the reactive sites of interest in case a hit is found. At the beginning of the subroutine the reaction radius is assigned using the type ID of the chemical species.

First, it is necessary to locate if the hit occurred in the nucleosome or in the linker fragment. A SB due to the indirect action was assumed to take place if the distance between the $\bullet OH$ radical and the closest hydrogen atom is less than the reaction radius calculated (see Fig. 24). Then, the information about the vertical level index, the closest nucleosome index, the impacted bp index, and the number of the impacted atom is returned.

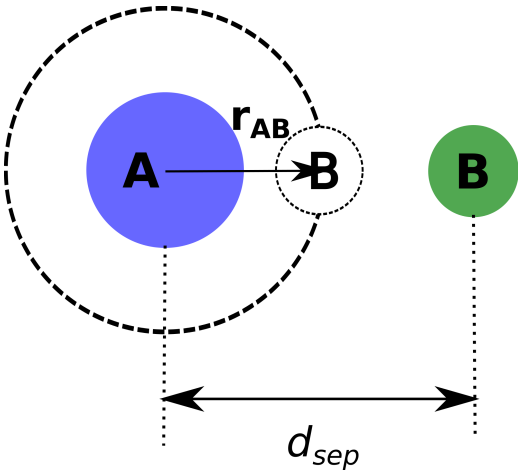

Figure 24 – A SB due to indirect action was considered if the separation distance (d_{sep}) is less than or equal to the reaction radius (r_{AB}).

With this new version of the code, strand breaks caused by direct and indirect actions are reported and for this reason it was necessary to create a new array to save the nature of the damage and where it occurred. This array, called *ISSB*, specifies the base pair, the DNA molecule chain, and the chromosomal domain. The role of this array is crucial in the clustering analysis to count the number of DSBs and classify them according to the nature and complexity.

Finally, each time a chemical species hit succeeded (a reaction occurs and leads to a SB) the next diffusion steps for that specific track need to be discarded since the reaction occurs and this species needs to be killed. Three cases were considered for the "killing" algorithm:

1. The site was already damaged by a primary or secondary particle during the physical stage.

2. The site was already damaged by another chemical species ($\bullet OH$).

3. The site was reached for the first time by a chemical species ($\bullet OH$).

Whenever the first case occurs, the track ID and the corresponding rank and event are saved in a stack in order to discard the next diffusion steps. In the second case, a probability of 0.85 was introduced to decide whether or not the track is going to be discarded. As explained before, the array *ISSB* identifies the base pair and the strand but not the specific atom impacted (for simplicity and to safe memory allocation). The previous probability of 0.85 takes into account that if another SB occurs in the same place due to indirect action there are other 6 hydrogen atoms (6 of 7 hydrogen atoms) per strand that could be targeted. To introduce this probability, a random number is generated between 0 and 1. If it is lower than 0.85 the track is killed and the next diffusion steps are no longer considered. Otherwise it is allowed to live. It should be pointed out that these two cases are not saved as a new value in the array *ISSB*, so they are not accounted as new damage in the clustering analysis.

If the damage site is attacked for the first time, then a breakage probability of 0.65 is introduced to consider that not all $\bullet OH$ interactions with the sugar-phosphate group necessarily lead to a SB. This value can be derived from the assumption that SBs due to the indirect action occur in 13% of the interactions of the $\bullet OH$ with the DNA and considering a 20% contribution of the interaction between the $\bullet OH$ and the sugar-phosphate group [26, 27, 72]. Following the same procedure just explained, a random number is generated and if this number is lower than the breakage probability, the track is killed and the following diffusion steps are discarded. If the track successfully induces a break then it is saved in the *ISSB* array and it is also killed.

Similar to the physical stage, The additional information of the SBs recorded is saved in the array *ISSBCHRO*, this includes the number of sequential base pair, the event associated, the position of the hit, the track ID and strand damaged. It is also possible, as for the physics stage, identify if one hit could occur in overlapping atoms of consecutive base pairs.

The simulation parameters employed in this work are presented in Table. 8. The selection of those parameters was based in an exhaustive analysis of the values reported in the literature together with some theoretical facts that support the selection of some specific values in this work. For instance, the threshold energy of 8 eV to account for a SB is related to the possibility in GEANT4-DNA of tracking secondary electrons in water down to energies around 7.8 eV and the fisrt excitation level of the water molecule is

Table 8 – Simulation parameters set in this work

Parameter	
SB treshold energy	8 eV
Reaction radius r_{AB} $\bullet OH$ - H reaction	0.12 nm
Chemical stage time cut	1 ns
Physics List	G4EmDNAPhysics
Chemistry List	G4EmDNAChemistry
Distance between SBs to induce a DSB	10 bp
Breakage probability for $\bullet OH$ - H reaction in the sugar-phosphate	0.65

approximately 8.2 eV. For the reaction radius, the value of 0.12 nm was calculated using the equation 2.24 with the adequate reaction rate constant and diffusion coefficient. Other parameters like the chemical stage time cut and the separation distance between two SBs to induce a DSB were reported by other authors [46, 72]. Furthermore, the breakage probability of 0.65 was previously used by other research groups and details about how to obtain this value can be found in §3.4.2. Some authors [72, 89] performed a parameter study to obtain the final set of parameters but that is not the case of this work.

3.4.3 Classification of DNA strand breaks

DNA strand breaks were classified according to the complexity of the damage (single and double) and also considering the nature of the break, due to direct or indirect action [72]. Once we have the information of the position in which the strand break occurs, it is possible to analyze the complexity of the damage. So far, three combinations of strand breaks were considered, see Figure 25:

- Case 1: Two breaks in opposite strands separated by less than 10 bp (DSB).
- Case 2: Two breaks in the same strand separated by less than 10 bp (SSB+)

- Case 3: “L shape” break, one SSB separated by less than 10 bp of two breaks in opposite strands at the same bp level (DSB).

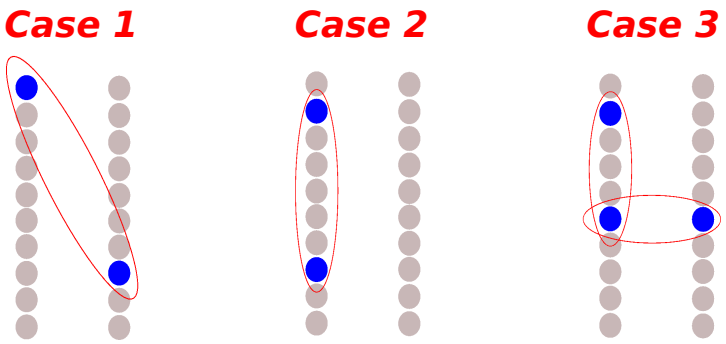

Figure 25 – Classification of DNA strand breaks according to complexity.

For all the cases considered here, damage classification in terms of its origin was taken as purely physical, purely chemical, or a mixture of both. The identification of the nature of each break was performed using the *ISSB* array as reference. More details about the verification of this algorithm can be found in the section (3.7). All damage yields were determined per unit absorbed dose and number of bp ($Gy^{-1}Gbp^{-1}$).

3.5 Site-hit probability analysis

In order to test the subroutine for indirect damage, we proceeded to analyze the value of the Site-Hit Probability (SHP). The SHP is defined as the ratio between the number of events that succeeded and the total number of hits processed. The theoretical SHP value estimated using the effective volume of 7 hydrogen atoms in each sugar-phosphate group was 0.015. If the track position coordinates are uniformly sampled over the entire cell volume, as in the GEANT4-DNA application, the SHP should remains constant independently of the LET value since this is a ratio of volumes (the effective volume occupied by the hydrogen atoms divided by the total volume).

The SHP is going to be reported as a function of LET of the incident particle in the Results chapter. The theoretical value is set as reference, but it should be noted that if overlapping of the hydrogen atoms occurs then this value is an upper limit for the SHP and the calculated using the radiobiological code could be lower. Still, it is expected that the SHP remains constant for the entire LET range studied.

3.6 Chemical species yield: *G* values determination

In radiation chemistry, the yield of chemical species is quantified by the time dependent radiolytic yield *G(t)*. The *G(t)* values are defined as the number of a particular chemical species produced per 100 eV of energy deposited at a given time (t):

$$G(t) = \frac{N(t)}{(E/100eV)} \tag{3.6}$$

The number of chemical species could increase or decrease with time, which depends on the particular species. In this work the G-values for the $\bullet OH$ at 1 ns were determined for both proton and alpha particles. The G-values obtained were compared with those found in the literature for protons and alpha particles at the same time.

3.7 Quality control tests

Due to the complexity of the algorithms for classifying the DNA damage, 7 quality tests (see Figure 26) were designed to verify its correct performance. The model used for this purpose was simplified, with each chromatin fiber containing only 40 bp. Predefined SB patterns were distributed along both strands increasing the complexity level and then they were processed with the corresponding algorithm.

The number of simple, complex, and all possible combinations of damage according to the nature of the break was verified. All quality control tests were successful.

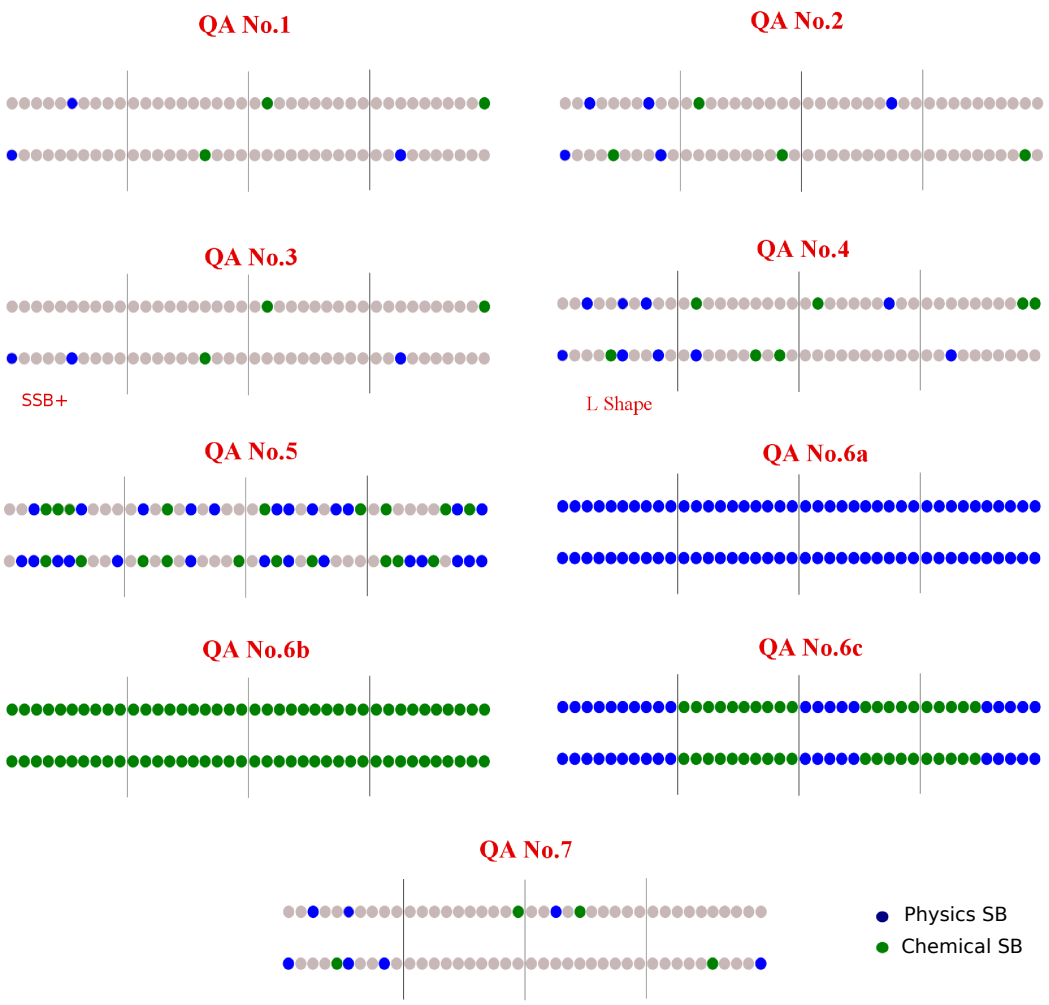

Figure 26 – Seven quality tests developed to mimic damage to the DNA strand.

4 Results and Discussion

4.1 Computational resources and storage capacity

At the end of each job or simulation, the information about the CPU time, walltime, memory, and virtual memory used is summarized in an email to the user if this option is activated in the submitted Portable Batch System (PBS) file. In this section, the four energies that demanded more of computing resources were selected. Two of the selected energies correspond to proton beams (28 and 50 $keV/\mu m$) and the other two, to α-particle beams (54 and 193 $keV/\mu m$). The total number of histories run for the proton beams selected were 192 for 28 $keV/\mu m$, 128 for 50 $keV/\mu m$ and for the α-particle beams were 128 for 54 $keV/\mu m$ and 64 for 193 $keV/\mu m$. From the total number of histories simulated only the required amount per batch to deposit the same absorbed dose for all the beam qualities studied was selected (see §4.5).

Figure 27 shows the CPU time and walltime for each LET value. The walltime corresponds to the real time the parallel job lasts and CPU time is the total execution time or runtime along which the CPU was dedicated to a process. Here, only the values for the irradiation oriented parallel to the X -axis were selected. The reason for doing this is that no remarkable differences were found between the two irradiation setups. Also, when comparing the values of CPU time and walltime with the geometry restriction on and off, no important differences were observed. This is related to the fact that those chemical species that were created inside the region of the chromatin fiber occupied by DNA were not killed, they simply were not save in the chemical output file.

For the memory and virtual memory we also obtained similar values when comparing different irradiation setups and the use or not of the geometrical restriction in our model (see Figure 28). In terms of memory, the resources needed in this work were acceptable taking into account the cluster capability (see §3.2.1).

The main advantage of implementing the geometrical consideration in the GEANT4-DNA was evident when comparing the storage capacity needed for each beam with and without activating the geometrical restriction. Figure 29 shows the space needed for each LET value in Gb, when the geometrical consideration is on and off. Whenever the geom-

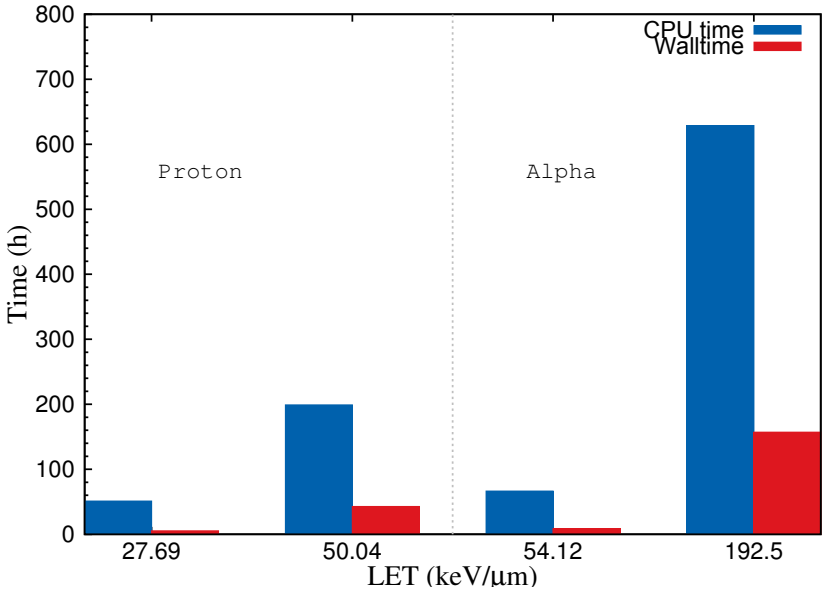

Figure 27 – CPU time and walltime for two protons (0.5 MeV and 1 MeV) and two alpha particles (2 MeV and 10 MeV). The data for each particle is divided by a dotted line.

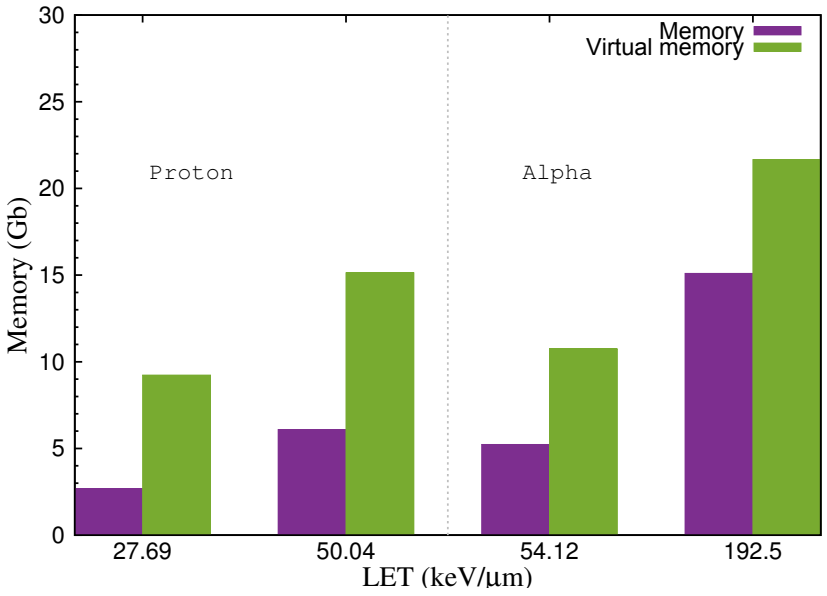

Figure 28 – Real and virtual memory for two protons (0.5 MeV and 1 MeV) and two alpha particles (2 MeV and 10 MeV). The data for each particle type is divided by a dotted line.

etry restriction was used, a reduction in the storage space of about 68% for LET values of 28 and 50 $keV/\mu m$. For LET of 54 and 193 $keV/\mu m$, the reduction were 38% and 64%, respectively. This gain in terms of the storage capacity reduction was crucial for the good performance of the cluster and a determining factor to perform this geometric consideration within the *phychem* application.

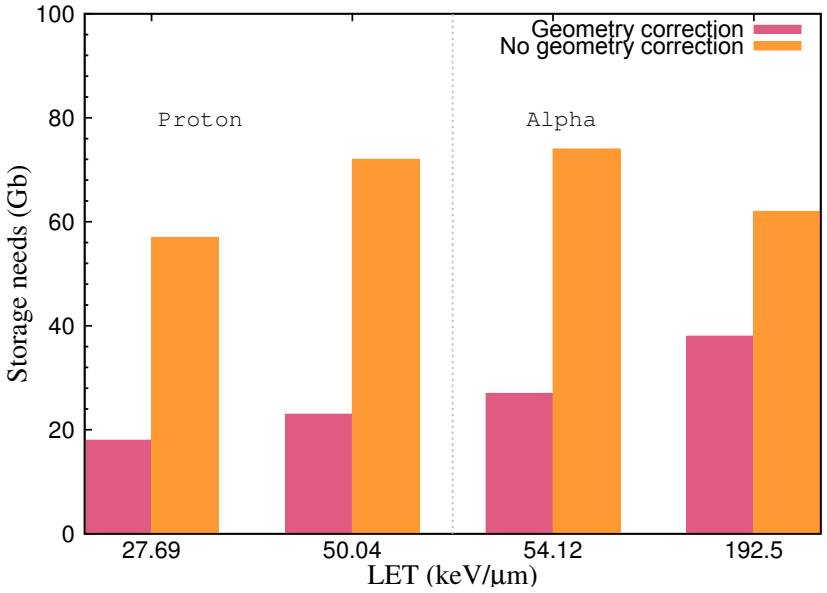

Figure 29 – Storage space occupied by the two protons (0.5 MeV and 1 MeV) and two alpha particles (2 MeV and 10 MeV). For each LET value the storage space is shown when the geometry correction is activated and when it is not used. The data for each particle is divided by a dotted line.

Fig. 30 shows the percentage of the total server capacity occupied by the output of each irradiation setup with and without geometry restriction. Each sector of the figure gathers all the LET range considered in this work. For each beam energy, the total storage needed to convert the individual output files supplied by the *phychem* application into the final output for each stage was taking into account.

4.2 Geometry Consideration in the GEANT4-DNA code

After implementing the mentioned geometry restriction in the GEANT4-DNA application (see §3.3.2), all chemical species produced inside the region occupied by the chromatin fiber were disregarded. This consideration reduced the indirect single strand break yield (SSBY) value by 62% and the indirect double strand break yield (DSBY) by 87%. These results were obtained by shooting 480 protons with 10 MeV energy. For 1 MeV protons, those values were 72% and 94%, respectively. Fig. 31 shows a snapshot

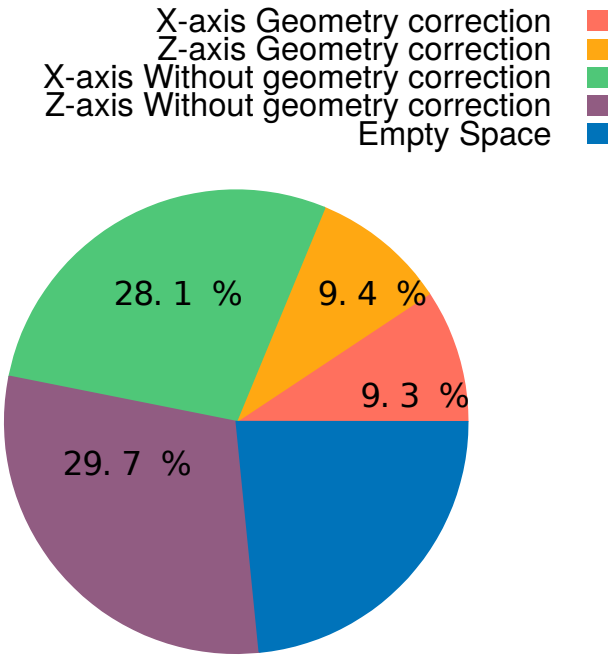

Figure 30 – Storage space occupied by all the simulations performed for each irradiation setup (X and Z axis) with and without geometry restriction. The percentage of space dedicated for each purpose was represented with respect to the total storage capacity available in the server (3 Tb).

taken during the chemical stage for a single grid cell containing one chromatin fiber. This grid cell is repeated throughout the cross sectional area of the ROI. To show that the geometrical constraint is fulfilled, for the first local time step (0.1 ps or 0 ps for reaction products during the chemical stage), the positions of those chemical species that survived to this constraint in the chromatin fiber reference system generated by a proton beam of 1 MeV are plotted. As previously mentioned in §3.3.2, this geometric restriction is only applied at the first local time step for every original chemical species, that is, at its production time. All chemical species that survive this restriction are allowed to diffuse throughout the whole simulation "World". The implementation of this restriction is an approximation that tries to improve the estimation of damage yields in the cell geometry and its limitation is related to those small regions inside the cylindrical shell occupied by water in the chromatin fiber model. Accounting for those regions without DNA atoms could increase the damage yields. Despite this limitation, not implementing any geometry restriction would lead to damage yields higher that those reported in the literature.

Figure 32 shows the positions (X and Y) of chemical species in the chromatin fiber reference system that survive to the restriction for four different local time steps: 0, 0.1, 50.1 and 100.1 ps. Initially, for the time steps at 0 ps (this is the first time step assigned

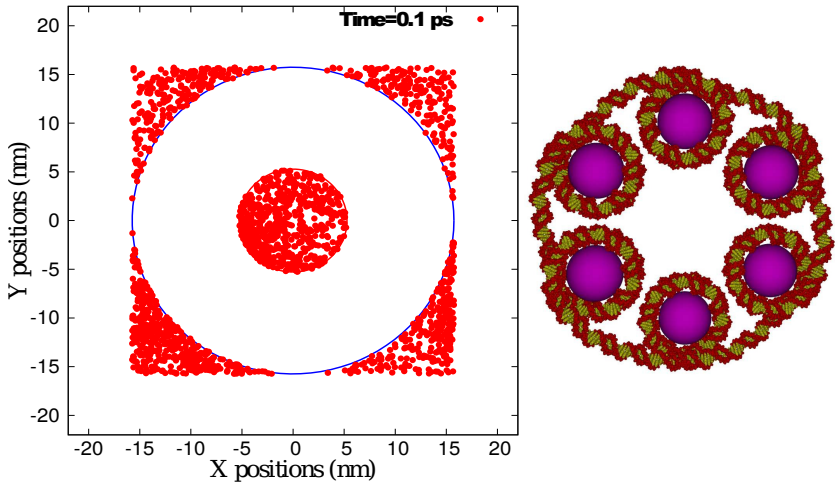

Figure 31 – Snapshot showing the position of all chemical species created at the first local time step (T=0.1 ps) in the chromatin fiber reference system for 1 MeV proton beam. Each X and Y value is converted to the chromatin fiber reference system using the index of the corresponding chromatin fiber in the cell nucleus. Only a grid cell corresponding to a single chromatin fiber is shown.

to reaction products) and 0.1 ps there are no species inside the forbidden region. As time passes, the species are allowed to freely diffuse everywhere including those restricted regions.

Fig. 33 depicts a 1 MeV proton track with the chemical species generated around it for three different diffusion time steps. Since the region traversed by the track is near a forbidden region, the number of chemical species is dramatically reduced when the geometry restriction is activated (bottom) when compared with the case when this restriction is turned off (top).

4.3 Site Hit probability analysis

Among all tests performed in order to know if the new version of the subroutine "Find the Closest Atom" of the radiobiological code was working properly, the comparison of the SHP for the entire LET range studied was one of the most important. In this work we focused our attention in the SHP for the chemical stage since the physical stage was previously tested.

The logical workflow followed in this work begins with the phychem application with the geometry restriction activated, followed by the damage analysis with the radiobiological code. The theoretical value for the chemical SHP, considering the seven hydrogen

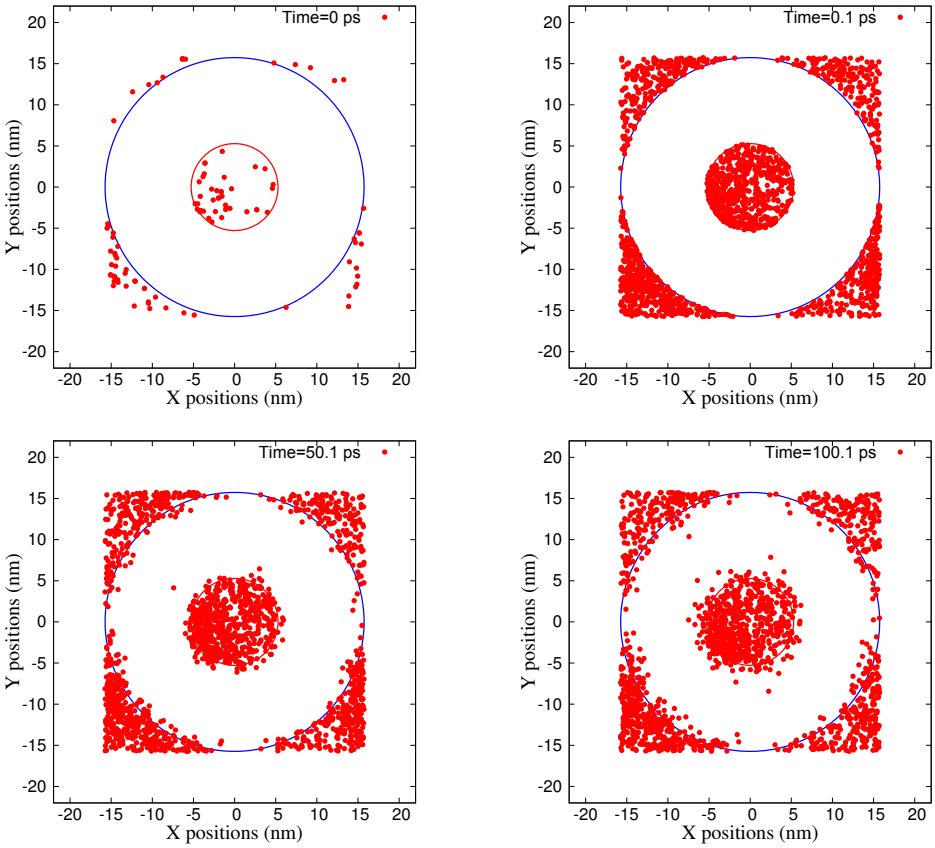

Figure 32 – Geometry consideration in "phychem" application. Snapshots are presented for the time steps 0, 0.1, 50.1, and 100.1 ps. At the production time the chemical species can only occupy the region outside the cylindrical shell. Those chemical species that survive this restriction are allowed to diffuse throughout the whole simulation "World".

atoms in each sugar-phosphate group as targets, was 0.015. Since the subroutine takes into account that overlap may occur for two atoms from consecutive base-pairs, we expect that this value should be of the same order but lower. So, the theoretical value is an upper limit for the SHP for each LET.

The first results obtained show a decreasing behaviour in the SHP plot as function of LET (see the red curve in Fig. 34). In order to discard possible sources of error, the positions of all the chemical species processed in the chemistry output file were overwritten changing its real position by a new one sorted randomly. The latter was done for each particle beam studied to see what was the behaviour of this sorted SHP with LET. This returned a constant value around 0.0140 ± 0.0002 on average for the entire LET range, as shown in Fig. 34 (blue curve).

At this point, it was necessary to study other possible explanations for this problem since when randomness is forced, the expected behaviour for the SHP was obtained.

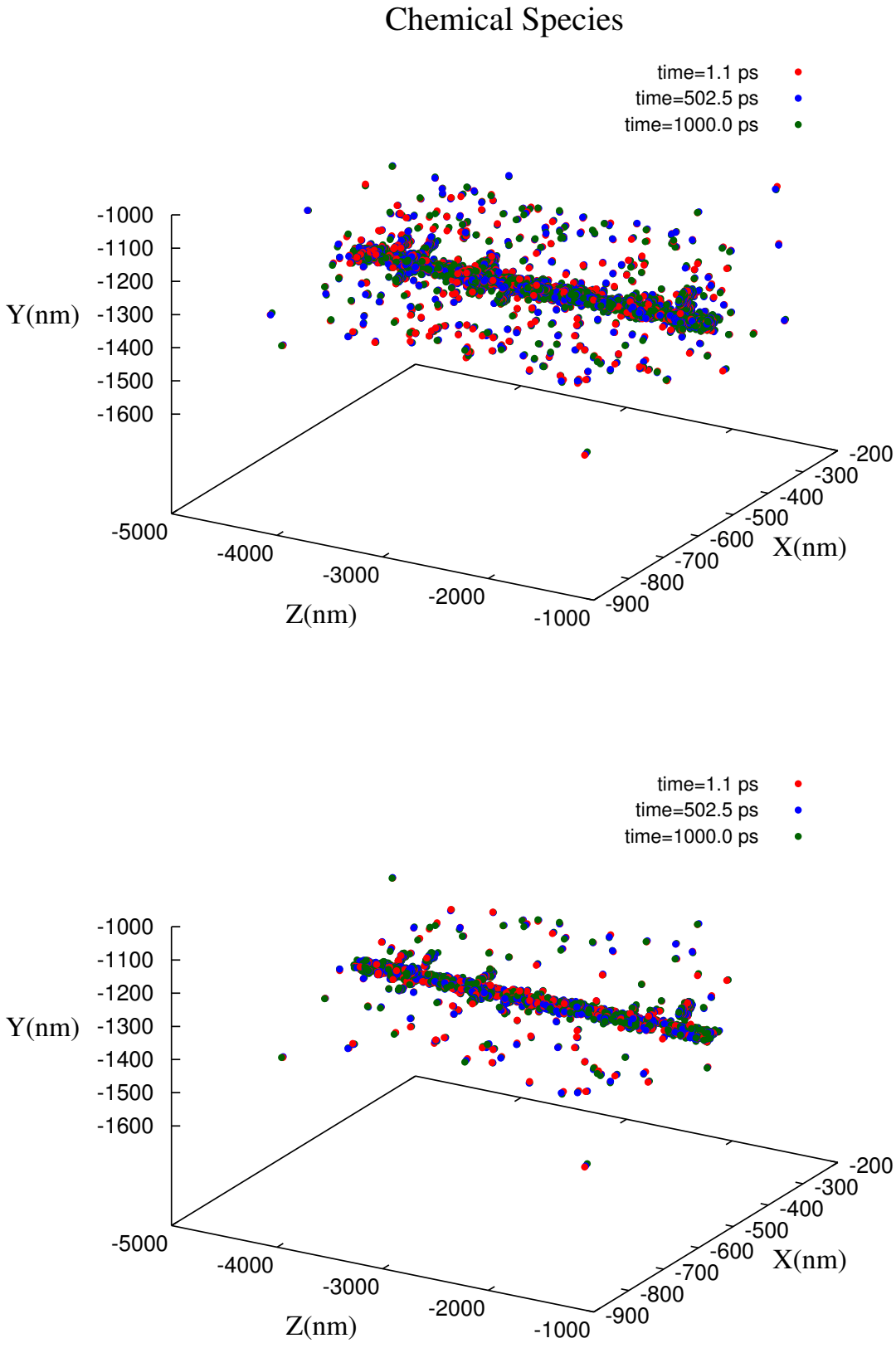

Figure 33 – Representation of a proton track of 1 MeV with the chemical species generated without (top) and with (bottom) the geometry restriction activated for the global time steps: 1.1 ps (red circles), 502.5 ps (blue circles) and 1 ns (green circles)

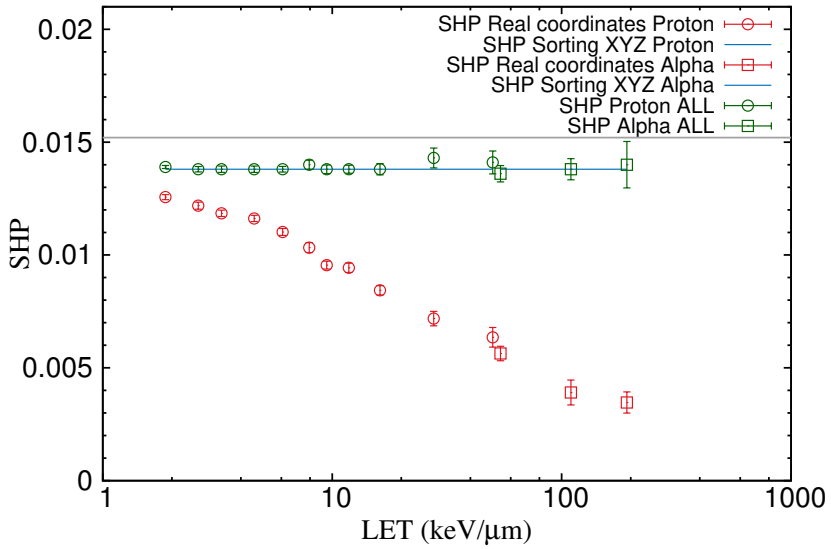

Figure 34 – SHP values: theoretical (gray line), randomly sorted (blue line), and the actual with geometrical restriction (red point) and without geometrical restriction (green points). Uncertainties represent one standard deviation of the mean.

Indeed, the trend showed by the SHP when we used the real coordinates for each chemical species was to decrease with increasing LET. When this happens, the track structure of the beam also changes. The density of reactive species around the track increases with LET, mainly for the α-particles. This takes us back to the geometrical restriction performed in the phychem application because the fact that we were discarding species created inside the cylindrical shell could break the spatial uniformity of the species distribution and, as a consequence, the final chemical output file will contain non-random X, Y, and Z positions. The only way to prove this is to put the geometrical restriction in GEANT4-DNA "off" and calculate the resultant SHP for this case as a function of LET. According to the data shown in §4.1, more storage space was necessary for running these tests. The final SHP for the chemical stage without activating the geometry consideration and accounting the number of successes over the number of chemical hits processed by the subroutine "Find the Closest Atom" is also shown in Fig. 34 (green values). The average value for the entire LET range considered was 0.0141 ± 0.0002. This final result confirmed that a lack of uniformity was present in the chemistry output file which led to a non-constant SHP in the first place. After this was confirmed, the geometry restriction was activated again since it is very difficult to manage the data storage in the case it is swithed "off".

4.4 Chemical species yield: *G* values

Another method employed to test our results was to determine the G-values as a function of LET for $\bullet OH$ at 1 ns and our results were compared with those reported by Kreipl et al. [76] for PARTRAC and Karamitros et al. [49]. In a similar fashion to what happened with the SHP, the first results obtained for the G-values were underestimated. In this case, the source of this problem came from the fact that many hydroxyl radicals were discarded because they were produced in one of those "forbidden regions", which led to fewer radicals per 100 eV of energy deposited at 1 ns.

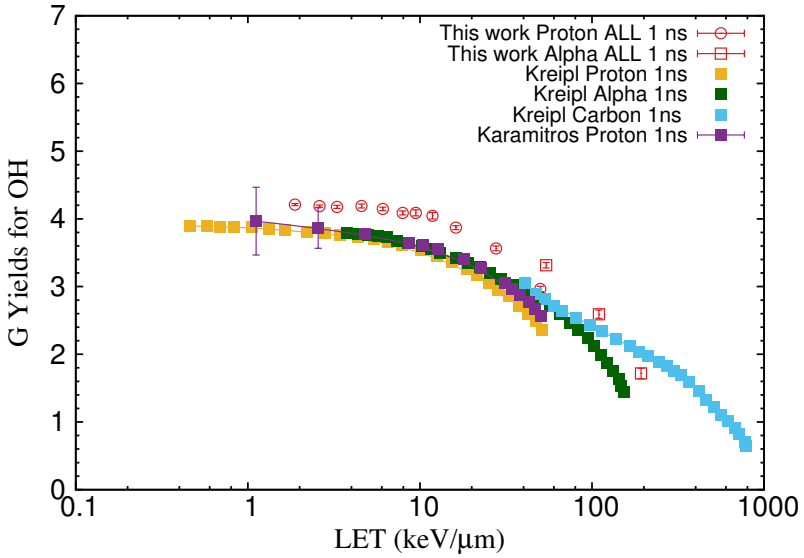

Figure 35 – G-values at 1 ns obtained in this work (red open circles for protons and red open squares for alphas), and reported by Kreipl et al. [76] for protons, α-particles, and carbon ions (orange full squares, green full squares and cyan full squares, respectively) and Karamitros et al. [49] for protons (purple full squares). Uncertainties represent one standard deviation of the mean.

Then, the chemistry output files obtained without activating the geometry restriction were also used to calculate the G-values. Furthermore, no restriction when the hydroxyl enters the histone region was taken into account in the radiobiological code. Figure 35 shows the G-values for both proton and α-particles for the entire LET range studied.

4.5 Single, double, and total strand break yields determination

The geometry used in this work, described in §3.3.2 and §3.4.1, has all chromatin fibers oriented with their axes parallel to each other. This orientation can influence the damage yields depending on the irradiation incidence angle. In the case the irradiation axis is parallel to the Z axis, there is a chance that more ionizing radiation tracks traverse the cell in a position in the XY plane that is inside those restricted areas were DNA is present, and in this case there is a bigger chance for a chemical species generated there being discarded along all the particle track. If the irradiation incidence is perpendicular to the chromatin fiber axis then each particle track traverse a combination of regions occupied by DNA and water. In order to account for this anisotropy effect, all damage yields were obtained as the average of two irradiation directions. The selected incidence angles were 0 and 90 degrees, that is the primary beam was oriented along the Z and X-axis direction, respectively. Before presenting the final results, the differences in the damage yields obtained for both irradiation setups are shown.

Fig. 36 shows the total strand break yield (TSBY) together with the direct and indirect contributions obtained for both irradiation setups. It can be observed that below 10 $keV/\mu m$ the total strand break yield for the X -axis irradiation is more stable when compared to the Z -axis irradiation. This is mainly due to the indirect contribution to the total yield. For the direct component, there is an agreement between both irradiation orientations for LET values lower than 16 $keV/\mu m$. The error bars were removed to better distinguish between both setups. In all cases the relative standard deviation (RSD) was lower than 5%. In this work, all damage yields values are reported with 68,2% confidence interval (one standard deviation).

The total number of simulated events for each LET value was equally distributed in batches in order to deliver approximately (1.46 $\pm$ 0.05) Gy per batch to the target volume for proton beams and (2.52 $\pm$ 0.03) Gy for α-particles. For α-particles, the deposited dose selected per batch was the same produced by one track of the beam with higher LET value (2 MeV α-particle). Since simulations using proton beams were performed before those for α-particles, the deposited dose per batch was lower than the one obtained for one track of 2 MeV α-particle.

The single strand break yield (SSBY) also shows a continuous decrease similar

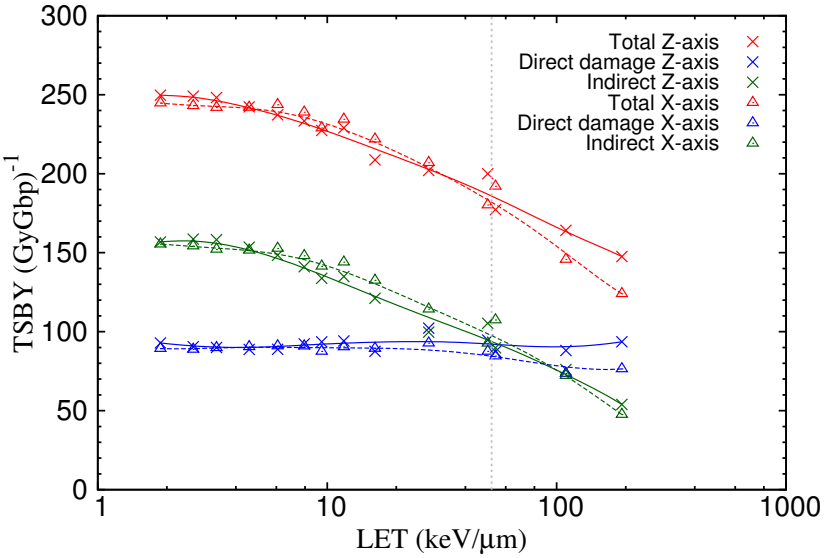

Figure 36 – Total strand break yields for proton and α-particles for the Z and X -axis irradiation. The dotted gray line separate proton and alpha values with geometrical restriction. The direct, indirect and total component are shown. The error bars were removed to better distinguish between both setups. In all cases the RSD was lower than 5%.

to the TSBY for the Z-axis irradiation (see Fig. 37). The differences are not so evident as in the TSBY curves. Each value obtained for an specific irradiation orientation differ from the other within the associated uncertainty. Only for the direct component for the LET range from 10 to 110 $keV/\mu m$ the differences are larger. The complex SSB yield (SSB+) agreed for the entire LET range studied.

However, major differences are observed for the double strand break (DSBY) curves, see Fig. 38. Indeed, the clustered damage accounted in this work seems to have more angular dependence for high LET proton and alpha particle beams. This behavior for high LET values is mainly related to the mixed contribution. It can be observed that the total DSBY for the Z -axis irradiation continuously increase while the X -axis remains nearly constant below 16 $keV/\mu m$. The total DSBY for α-particles for the X -axis irradiation seems to remain constant or presents a slight decrease and for the Z-axis irradiation an increase with LET can be observed.

From now on, the damage yields for each LET are presented as the average value for both irradiation setups in the X and Z axis direction.

Figure 39 shows TSBY as function of LET for all the proton and α-particle energies considered in this work. They are discriminated according to the origin of the damage: direct and indirect. The LET for each beam was estimated at the center of the

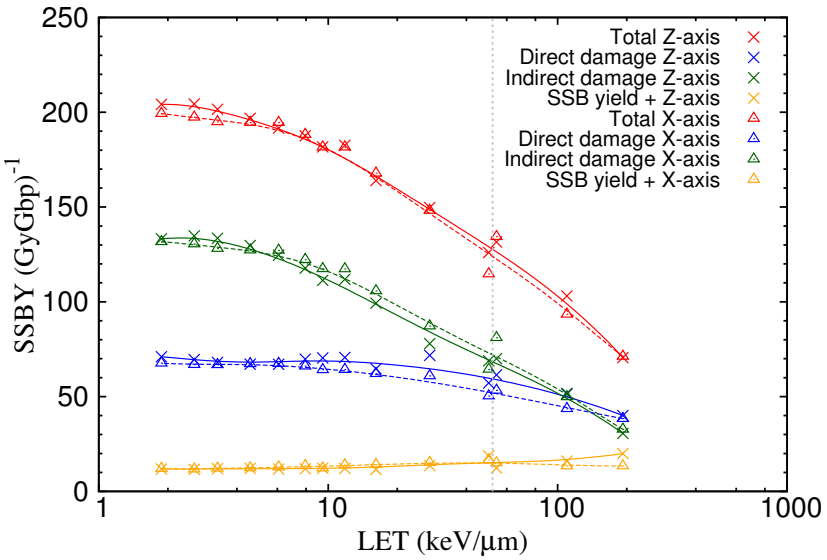

Figure 37 – Single strand break yields for proton and α-particles for the Z and X -axis irradiation. The dotted gray line separates proton and alpha values with geometrical restriction. The direct, indirect, SSB+ and total component are shown. The error bars were removed to better distinguish between both setups. In all cases the RSD was lower than 5%.

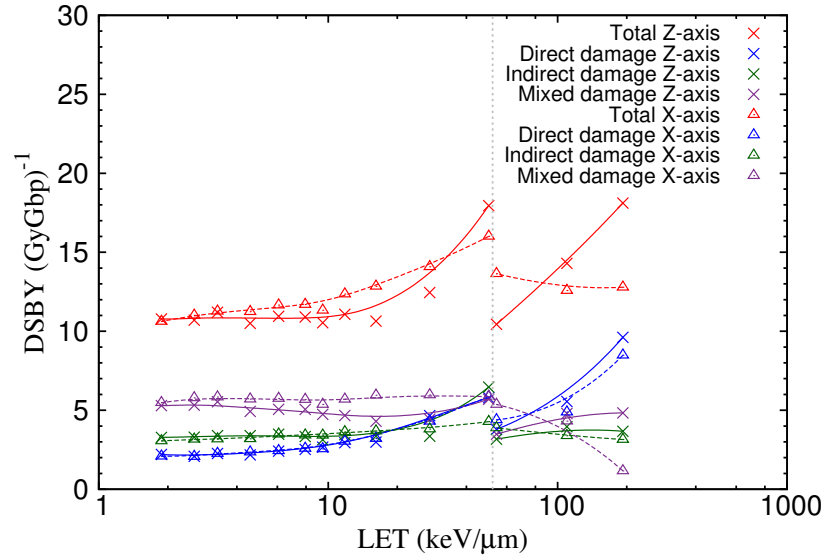

Figure 38 – Double strand break yields for proton and α-particles for the Z and X -axis irradiation. The dotted gray line separate proton and alpha values with geometrical restriction. The direct, indirect, mixed and total component are shown. The error bars were removed to better distinguish between both setups. In all cases the RSD was lower than 5%.

cell nucleus. These yields were obtained accounting for all strand breaks, which may lead further to single or double strand breaks. Each damage component is represented by the same line color. The reader should notice that open circles and open squares were used for identifying proton and α-particles, respectively. This convention is used throughout this work. For LET values ranging from 2 to 12 $keV/\mu m$, the TSBY fluctuates around an average value of (240 ± 7) $(GyGbp)^{-1}$ with a RSD of 3%. Above 12 $keV/\mu m$, a decrease can be observed going from (232 ± 6) $(GyGbp)^{-1}$ to (136 ± 12) $(GyGbp)^{-1}$, which represents a 41% decrease. The indirect component determines the overall behaviour of the TSBY. The direct component remains constant around an average value of (90 ± 4) $(GyGbp)^{-1}$ for all LET values and particles types studied.

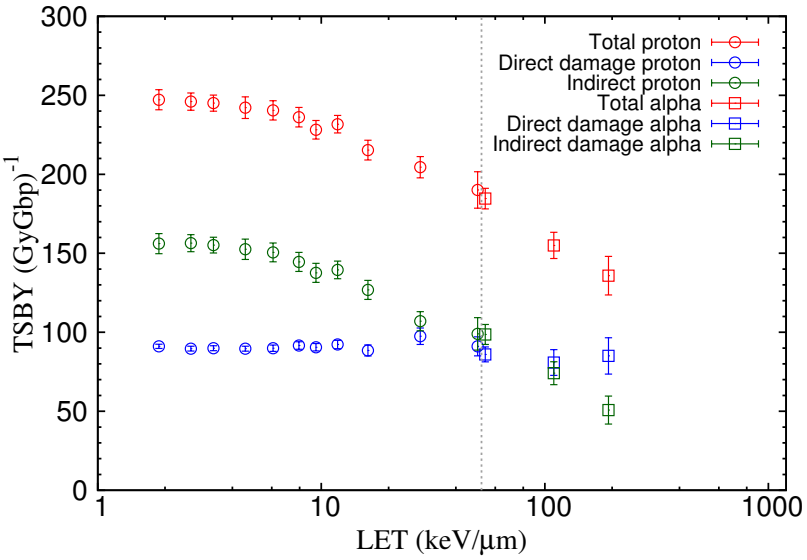

Figure 39 – Total strand break yields for proton and α-particles: direct, indirect, and total components with geometrical restriction. Uncertainties represent one standard deviation of the mean.

SSBY as a function of LET for protons and α-particles are plotted in Fig. 40. Again, these yields are discriminated according to the corresponding damage mechanism. The SSB+ yield includes direct, indirect, and mixed mechanisms but they were not discriminated in this plot for the sake of simplicity. In the LET range studied, the total SSBY decreased by 61%. For lower LET values a stability in the total SSBY curve can be observed. For 2-12 $keV/\mu m$ LET values, the SSBY values have a RSD of 4% relative to the averaged SSBY. A pronounced decrease is observed for LET values above 12 $keV/\mu m$. Direct SSBY seems to be stable up to about 28 $keV/\mu m$ and then decreases with LET. Indirect SSBY follows a similar behavior but the decrease begins at about 16 $keV/\mu m$.

Additionally, the SSB+ yield remains nearly constant up to the LET value of 28 $keV/\mu m$ and then an increment of 43% for protons and 21% for α-particles is observed.

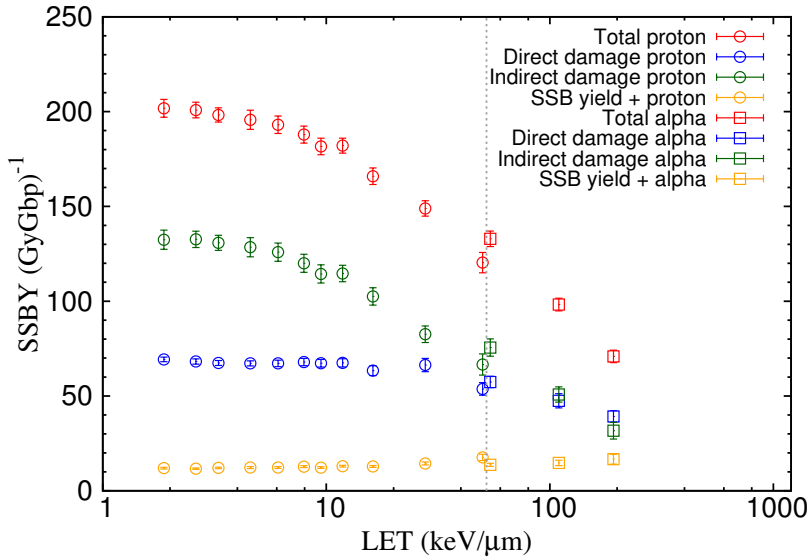

Figure 40 – Single strand break yields for proton and α-particles: direct, indirect, SSB+, and total component with geometrical restriction. Uncertainties represent one standard deviation of the mean.

Figure 41 displays the results obtained for the DSBY. Since there are at least two SSB involved in the formation of a DSB, this damage was classified as direct, indirect, or mixed. The latter represents a DSB induced by direct and indirect single breaks. The tendency observed in the total DSBY is to increase with LET, for both protons and α-particles. For protons, total DSBY goes from (11 ± 1) $(GyGbp)^{-1}$ up to (17 ± 2) $(GyGbp)^{-1}$; for α-particles it goes from (12 ± 1) $(GyGbp)^{-1}$ to (16 ± 2) $(GyGbp)^{-1}$. The direct and indirect components for protons increase with LET while for α-particles the direct component increases and the indirect one remains nearly constant with LET. For the LET range studied here, the mixed component of the DSBY decreases with LET for both particles.

Figure 42 shows the contribution of each damage component to the total number of DSB as a function of LET. The major contribution to the total DSB up to 28 $keV/\mu m$ comes from the mixed mechanism, which decreases very fast with LET. For LET values over 28 $keV/\mu m$ the direct component gains more importance. The indirect contribution remains constant for LET values lower than 28 $keV/\mu m$ with a RSD lower than 2%, then decreases with the LET. The direct component shows an increasing trend for the LET range studied in this work. For LET values greater than 28 $keV/\mu m$, a switch in the

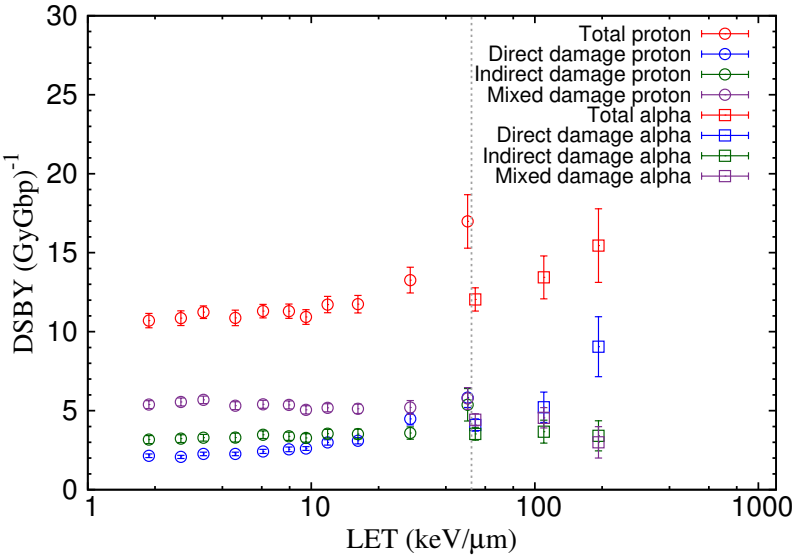

Figure 41 – Double strand break yields for proton and α-particles: direct, indirect, mixed and total components with geometrical restriction. Uncertainties represent one standard deviation of the mean.

relative importance of each component is observed.

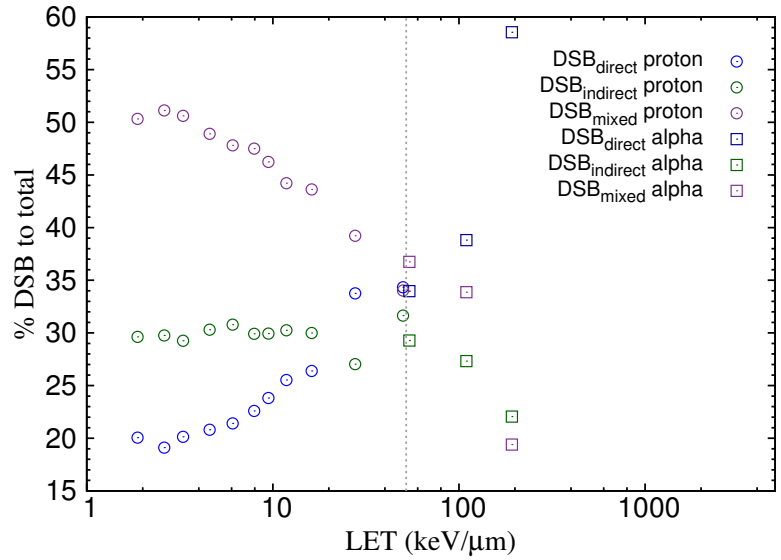

Figure 42 – Ratio of DSB_{direct}, $DSB_{indirect}$, and DSB_{mixed} relative to DSB_{total} with geometrical restriction.

As shown in Fig. 39, total TSBY decreases with LET due to the decreasing trend of the indirect contribution while the direct contribution remains nearly constant. The decrease of the indirect yield is related to the structure of the chemical track [26]. For high LET values, the ionization density increases and so the initial chemical species yield, so that the recombination probability of $\bullet OH$ radical with other chemical species like

$H\bullet$, other $OH\bullet$, H_2, and e^-_{aq} increases. As a consequence, the amount of $\bullet OH$ radicals that remain alive and are able to react with the sugar-phosphate group is reduced. The constant behaviour of the direct TSBY was discussed in detail by Bernal et al. [25, 106]. They found that the direct TSBY mainly depends on the target volume and the number of energy depositions per unit absorbed dose, which depends very weakly on the particle type and energy. This figure also shows that the direct and indirect curves cross each other at about 70 $keV/\mu m$, which agrees with the results reported by Friedland et al. (see Fig. 2 in Ref. [26]). Figure 43 shows the direct contribution to TSBY as a function of LET for proton and α-particles, the difference between those values in terms of the RSD was less than 5%. Additionally, in the same figure, the direct DSBY is also showed for protons and α-particles. These values are lower compared to those reported in Ref. [25] (for protons of 10 $keV/\mu m$ 63% lower and for α-particles of 100 $keV/\mu m$ 70% lower), which could be related to differences in the geometry specification (atomic resolution). However, both direct TSBY and DSBY show the same behaviour as that previous work. It can be also observed that, within the uncertainty, there is continuity of the TSBY when going from protons to α-particles.

The decrease in the SSBY shown in Fig. 40 is related to the behavior of the indirect TSBY commented just above. For the direct component a decrease is observed for high LET values due to the constant behaviour of the direct TSBY and the increase of the direct DSBY with LET. Unlike in the TSBY curve, a discontinuity in the SSBY when going from proton to α-particles can be observed. This behavior can be explained by the discontinuity in the DSBY and the continuity of the TSBY. That is, SB tend to cluster into DSB, so there are less SSB.

In Fig. 41 it is observed that total DSBY increases with LET, for both proton and α-particles. This tendency is consistent with that found in previous works [26, 28, 72] and is related to the increase of the clustering of energy depositions and chemical species production. The reader should also notice that the indirect DSBY for α-particles is lower than that for protons at the same LET, which was already observed and explained previously for direct damage [83], but now it is also seen for the indirect case. Since energy depositions are responsible for chemical species production, less clustered energy depositions should also lead to less clustered species production.

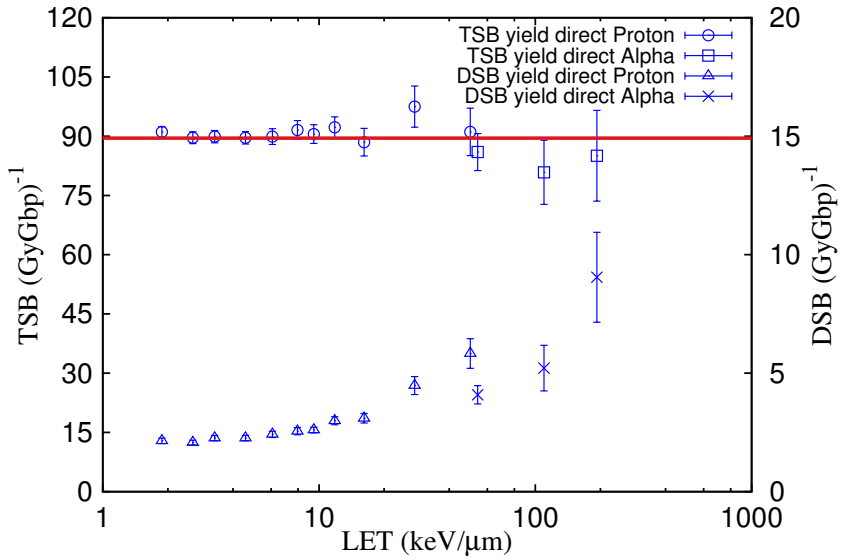

Figure 43 – Total and double strand break yields taking into account direct contributions for protons and alpha particles with geometrical restriction. To keep the same color for the direct component the data points symbol in the DSBY curve was changed in order to distinguish the particle type: triangles (protons) and cross (α-particles). The average value corresponds to 89.53 $(GyGbp)^{-1}$ (red solid line). Uncertainties represent one standard deviation of the mean.

4.6 Comparison of DSBY results with experimental data and simulations found in the literature

A comparison between the results presented in this work and other simulations and experimental data found in the literature was performed. Results from simulations are influenced by the physical models implemented in Monte Carlo codes and damage model parameters of the biophysical model. Experiments are dependent on the ability to determine small DNA fragments, among other factors. Figures 44 and 45 show this comparison for protons and α-particles, respectively.

The results for proton and α-particles DSBY presented in this work follow the same behaviour as in other simulations [26, 27, 28, 46, 47, 110]. In terms of absolute values, there are high discrepancies between these works. Different MC codes and biophysical models were used. Our results for protons are consistent with those reported by Friedland et al. [26] and the experimental result of Belli et al. [108]. The simulation performed by Nikjoo et al. [110] used the PITS code for primary ion interaction and the CPA100 code for the transport of secondary electrons in liquid water and the chemical stage simulation. There are a few examples of parameters study of mechanistic DNA damage simulations,

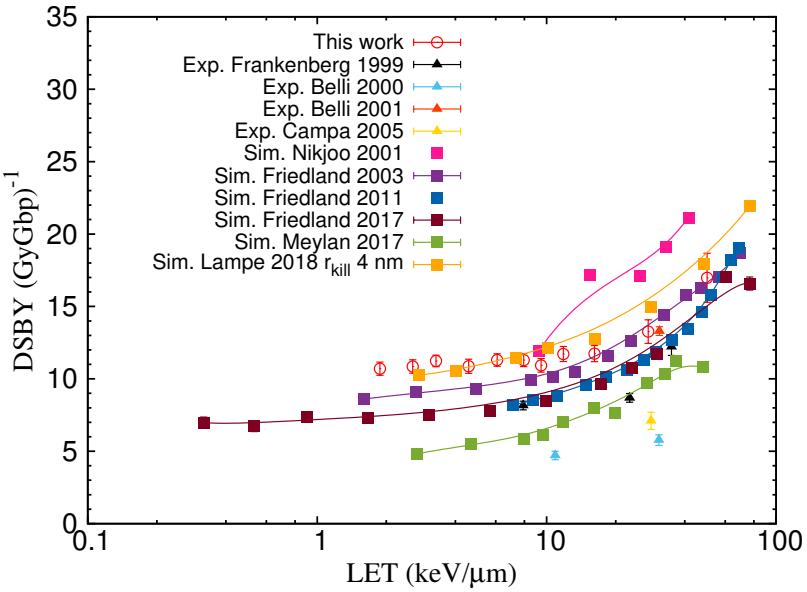

Figure 44 – Comparison of results for double strand break yields for protons. Experimental data used as reference corresponds to the work performed by: Frankenberg et al. [19], Belli et al. [107, 108], Campa et al. [109] using the PFGE technique (see §2.2.2.1). Other simulations results are also plotted: Nikjoo et al. [110], Friedland et al. [26, 27, 47], Meylan et al. [28], Lampe et al. [46]. Available uncertainties are shown. Uncertainties for our results represent one standard deviation of the mean.

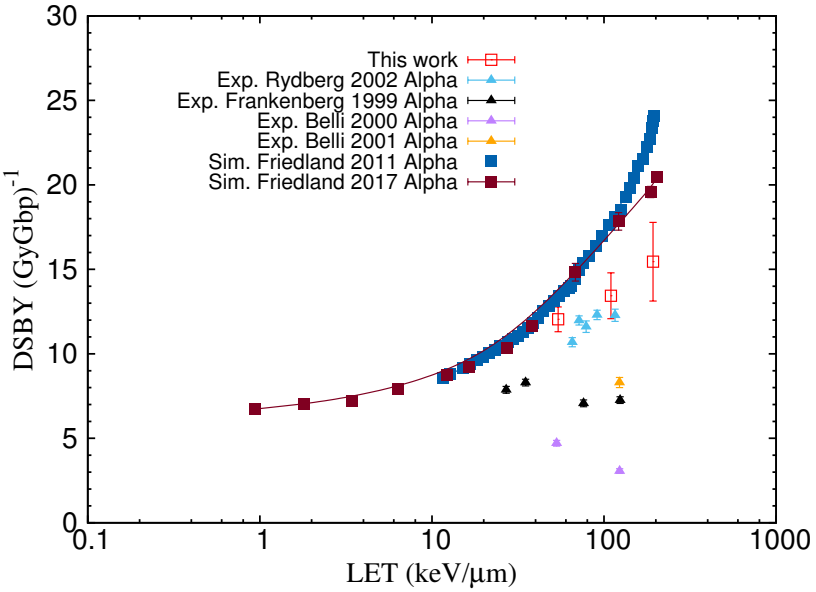

Figure 45 – Comparison of results for double strand break yields for α-particles. Experimental data presented corresponds to: Frankenberg et al. [19], Belli et al. [107, 108], Rydberg et al. [111] using the PFGE technique (see §2.2.2.1). Simulations results are also plotted: Friedland et al. [27, 47]. Available uncertainties are shown. Uncertainties for our results represent one standard deviation of the mean.

one of them published by Nikjoo in 1997 [72]. In this study, they used the CPA100 code and electrons as primary particles. The geometry was a sphere filled with DNA segments generated using the μ-randomness method to determine a set of parameters. From this sensitivity study, they chose a threshold energy of 17.5 eV for the SB production, a breakage probability of 0.65, and a time limit for the chemical stage of 1 ns. Their results reported in Ref. [110] were higher than those presented in this work, the differences go up to 32%. Recently, Bordage et al. [112] compared the default cross sections of the GEANT4-DNA toolkit with the CPA100 code developed by Terrisol et al. [74] and with an implementation of this code in GEANT4-DNA. They encountered significant differences, specially in the electron travel distances in liquid water. The total inelastic cross sections in the CPA100 code are larger than the default GEANT4-DNA cross sections.

Nowadays, most of the simulations try to use realistic cell geometries [28, 46, 47, 113]. Particularly, the work performed by Zein et al. [113] implemented in GEANT4-DNA a realistic phantom of mitochondrial networks of fibroblasts cells developed from fluorescent microscopic images.

Meylan et al. [28] used the default physics list of the GEANT4-DNA v10.1 package but with a different biophysical model. They built a realistic phantom of a fibroblast cell nucleus with a DNA model that can be imported into GEANT4-DNA using the DnaFabric software [48]. This DNA model does not have atomic-resolution, so the reaction between the •OH radical and the hydrogen atoms of the deoxyribose was not accounted for explicitly. In addition, they used a higher threshold energy (17.5 eV) for single break induction, a probability to consider a SB of 0.4 and a cut off time for the chemical stage of 2.5 ns. Relative to Meylan et al., our results overestimate the DSBY for protons even when they used the default option for the physics in GEANT4-DNA. This could be related to the different energy threshold, the breakage probability, and the geometrical model, that is, a different biophysical model.

The most recent study on mechanistic DNA damage simulation using the GEANT4-DNA extension as the MC track structure code was performed by Lampe et al. [46, 89]. As in Meylan et al. [48], they also developed a realistic geometry of an *Escherichia coli* bacterial cell but they also described each DNA base pair with independent volumes representing the sugar, phosphate, and the nucleobases, without specifying the atomic composition. Lampe et al. [89] also performed a parameter study testing the influence of

three different sets of physical models available in GEANT4-DNA (default, option4 and option6) and compared their results to those presented by Nikjoo et al. [72]. The default physics list available in GEANT4-DNA uses a partial wave model available in the energy range 7.4 eV-1 MeV for elastic scattering of electrons and the first Born approximation with semi-empirical corrections of the dielectric function at low energies for inelastic cross sections [112]. The alternative option 4 for the physics models in GEANT4-DNA uses models based on the Emfietzoglou's parameterisation of the dielectric response function of liquid water for inelastic interactions and an improved Screened Rutherford model with a screening factor obtained from vapour water data. Finally, the option 6 is the implementation of the CPA100 code [112]. Lampe et al. [46, 89] found that likelihoods for inducing SSB or DSB were higher for the option 6 (CPA100), which has higher interaction cross sections. Also, two different threshold energy models were tested: a linearly varying probability from 5 eV-37.5 eV and a fixed threshold of 17.5 eV. Even when the cross sections for the default models were lower when compared to the other two options, more damage was measured using the PARTRAC's linear probability than the using the fixed value. The fixed threshold of 17.5 eV was also used by Nikjoo et al. [72]. In their work using a complex cell geometry, Lampe et al. used the option 4, which provides intermediate results but higher than those obtained with the default models. The set of parameters selected by Lampe et al. [46] were a fixed energy threshold of 17.5 eV, a time limit for the chemical stage of 1 ns, a radius r_{kill} of 4 nm from the DNA site beyond which the species is no longer followed, and a breakage probability of 0.4. It is important to point out that the reason for selecting a breakage probability of 0.4 by Lampe's group came from the sensitivity study performed fitting their results to those obtained by Nikjoo et al. [72]. This reason differs from the one stated by Meylan et al. [48], which considered that two of every five reactive sites are reached by the $\bullet OH$ radical. The DSBY for protons obtained by Lampe et al. for high LET are higher than those obtained in this work, reaching 12% of relative percentage difference. This could be related to differences in the physical models implemented, the option 4 should led to higher damage yields values than the default one but the fact that we select a lower threshold energy of 8 eV increased our results specially at lower LET values (below 10 $keV/\mu m$) where the differences did not exceed 6%.

For α-particles, DSBY consistent with the experimental values reported by Rydberg et al. [111] was obtained. The group of Friedland reported two main data series for the proton case. Our results are similar to their first results [26] above about 10 keV/μm.

In that work, they introduced a linear probability for break induction after a given energy deposit, going from 0 to 1 when the deposited energy goes from 5 to 37.5 eV. Friedland et al. used this approach for adjusting the break yield direct/indirect percentage ratio to 35:65 for low LET radiations, taking into account that breaks may be induced by electron with energies below 10 eV. According to the experimental works published by Boudaiffa et al. [114, 116], single break cross section for DNA break by electron impact seems to depend weakly on the incident electron energies in the range 10-50 eV. Moreover, the results from Panajatovic et al. [115] show that this cross section does not show a well defined trend, at least up to 20 eV. Thus we decided to keep a step-like probability for single break induction, accounting for the fact that electron transport is carried out down to about 8 eV. Another important difference between the PARTRAC model and ours is that it uses two van der Waals radii around each sugar-phosphate atom to define the target volume that must be hit for inducing direct and quasi-direct (non-scavengable) breaks. We used only one van der Waal radius. On the one hand, this should lead PARTRAC to overestimate site hit probabilities and so break yields. On the other hand, their linear probability as a function of the deposited energy should compensate this effect, at least partially, when compared with our step-like function. From the chemical point of view, PARTRAC and GEANT4-DNA basically follow the same production and transport model. However, this model is not implemented in the same manner in these codes. PARTRAC accounts for the production of chemical species within a 25 nm cylinder around the chromatin axis, excluding the space occupied by the DNA and histone atoms. In our approach, chemical species are produced in the whole space, except that occupied by the DNA and histones. Furthermore, they have used a probability for chemical break after •OH reaction with the sugar moiety of 0.7 [26] and 0.65 [47], whilst we use 0.65. There are also several differences between the physical models used for cross section determination but such comparison is out of the scope of this work. PARTRAC commonly follows electrons down to 10 eV and we do it down to 8 eV, which may lead to an overestimation of our results. It should be pointed out that Friedland et al. have presented several PARTRAC flavors along the development of their biophysical model. For instance, different values for the upper energy in the linear probability function for break induction have been used: 37.5 eV in Ref. [26], 57.6 eV in Ref. [117], 60 eV in Ref. [16], and 37.5 eV in their most recent work [47]. The damage yields reported in the latter work are virtually the same as those published in Ref. [27] (see Figs. 44 and 45). It is not clear for us why the yields reported by Friedland

et al. in 2017 [47] are lower than those published in 2003 [26]. They attribute this decrease to the use of a lower water density (1.0.vs.1.06 g/cm^3), a different irradiation setup, and a more complex DNA geometrical model in their latter work. As shown in Fig. 45, our DSB yields for α-particles are closest to the experiments than those reported by Friedland et al.

4.7 Comparison of TSBY values, TSB direct to TSB indirect and SSB to DSB ratios

A comparison was also performed for the TSBY obtained in this work and others found in the literature [26, 46] for protons. Fig. 46, shows the total, direct, and indirect contribution to the TSBY. The results obtained in this work differ in terms of absolute values from those published by Friedland [26] using a geometrical model with atomic resolution. The major differences encountered were in the direct contribution reaching up to 33%, however the behaviour is constant in all cases, as expected. This considerable difference in the direct TSBY is associated to the energy threshold used in PARTRAC (linearly varying probability from 5 eV to 37.5 eV) and in this work (8 eV). The above stated has an impact in the total TSBY which deviates from the PARTRAC results up to 20%. Furthermore, the indirect component for high LET values shows differences below 10% and for the lower LET the deviation reached 17%. The other work used as reference [46], does not use a geometry with atomic resolution and the energy threshold to account for a direct SB was 17.5 eV. In addition, the use of different physical models explains why they had the lowest TSBY for the direct contribution of the three different simulations showed. For the indirect damage, their values and the ones presented in this work differ by 20% and their trend does not follow the one obtained with the PARTRAC suite and in this work.

Fig. 47 shows a comparison between the direct (TSB_d) to indirect (TSB_i) TSBY ratio obtained in this work and those obtained by Friedland et al. [26], Meylan et al. [28] and Lampe et al. [46]. Our results agree very well with those calculated from Friedland's results, the higher discrepancy was 20%. The simulation results obtained by Meylan et al. and Lampe et al. seems to underestimate this ratio, but again the geometrical model and the parameters used by them differs from the ones selected in this work and used by

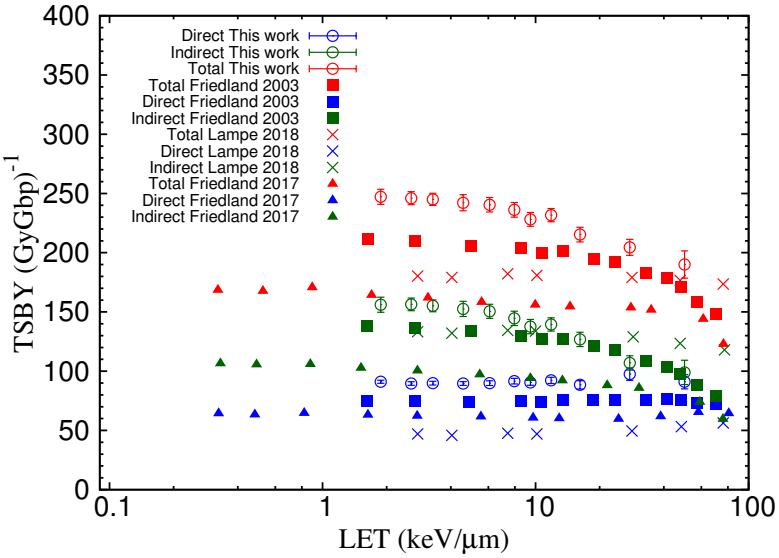

Figure 46 – Total, direct and indirect strand breaks for protons with geometrical restriction compared to other simulation results found in literature [26, 46]

Friedland et al.

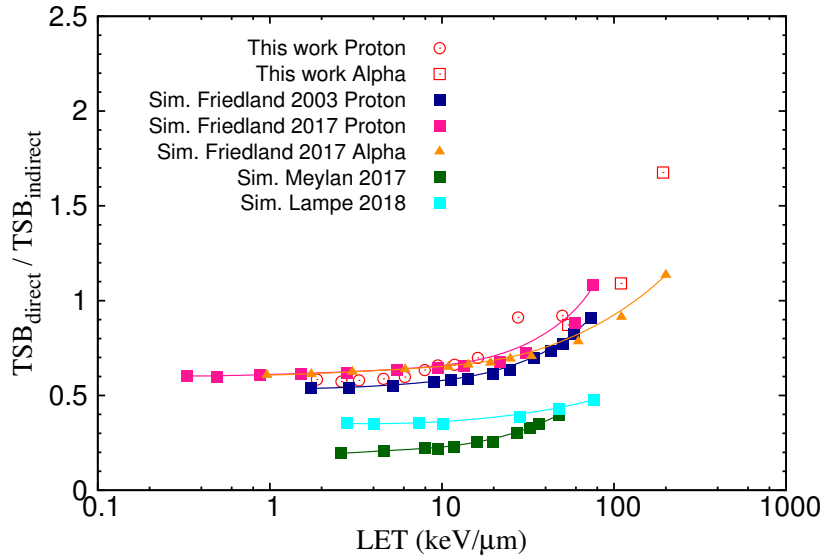

Figure 47 – Ratio of TSB_{direct} to $TSB_{indirect}$ with geometrical restriction. The data used as reference was calculated from the direct and indirect strand break yields reported by Friedland et al. [26] (Fig. 2), [47] (Fig. 3a), Meylan et al. [28] (Fig. 8) and Lampe et al. [46] (Fig. 6).

Finally, the ratio of SSB to DSB is presented in Fig. 48. This ratio should be independent of the differences between experimental and simulation geometries. As in the previous ratio (Fig. 47), the values of SSB/DSB obtained in this work are in agreement with the PARTRAC results for the LET range studied. The highest deviation is below 18%. All simulation results plotted in this figure are very consistent for high LET values.

For low LET, Lampe et al. underestimate this ratio using a distance from the DNA chain for killing the species of 4 nm and 1 nm. For the entire LET range studied, the experimental measurements in plasmid DNA using proton beams [118, 119, 120] have a larger SSB to DSB ratio when compared to all the simulation results. It is possible that more chemical damage due to distant radicals is induced in plasmid DNA under proton irradiation, probably due to the lack of folding proteins [46].

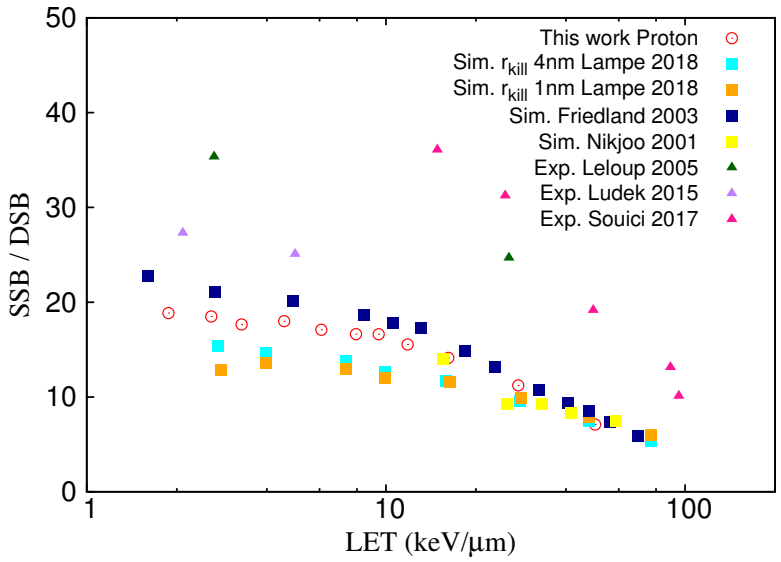

Figure 48 – Ratio of SSB to DSB obtained in this work with geometrical restriction and compared to other simulations [26, 46] and measurements in plasmid DNA irradiation with proton beams using gel electrophoresis [118, 119, 120] (see §2.2.2.1).

Using a reaction radius of 0.12 nm (derived from equation 2.24), a good agreement of the SSBs and DSBs with experimental results [19, 108] and simulations [26] it is observed. Also, the direct to indirect TSBY and SSB to DSB ratios for protons are very consistent between different simulations, specially when these ratios are compared with the PARTRAC code results. For the lowest proton LET (2 $keV/\mu m$), we have obtained a SSB to DSB ratio of approximately 19:1 and a direct to indirect number of breaks ratio of 37:63. For ^{60}Co radiation, which is in the same LET region, the literature reports similar ratios of 20:1 and 35:65 respectively [26]. This excellent results show the high quality of our biophysical model for the LET range studied.

Conclusions

The research conducted and presented here is part of the efforts been performed to relate the physical properties of the ionizing radiation track to observed biological effects. The main contribution of this work is a new biophysical model capable to account for the indirect damage induced by chemical species resulting from water radiolysis. To the best of our knowledge, it is the first time the GEANT4-DNA code is used together with a geometrical model with atomic resolution for this purpose. Moreover, previous to this work the only group that conducted this type of research using an atomic description of the DNA was that led by Friedland [26], which developed the PARTRAC code. Unlike the work done by Friedland et al., the procedure followed in this work does not introduce any fitting process. That is, the involved parameters were extracted from previous related works and introduced into the biophysical model with no further adjustments. This is an effort for obtaining a model based on first principles, as much as possible.

To take advantage of the capabilities that the GEANT4-DNA toolkit offers for simulating the physics, physico-chemical, and chemical stages some considerations need to be introduced to optimize the simulations. Among these optimizations, the most important are the use of parallel programming and binary format for the output files generated. In addition, if the biophysical model is not implemented in the GEANT4-DNA application then some restrictions should be considered for creating chemical species in those regions occupied by the DNA molecule.

A radiobiological code was developed that considers the scavenging capacity of the hydroxyl radical by proteins (histones). Furthermore, the approach followed is able to distinguish the specific hydrogen atom in the sugar phosphate group that reacts with the hydroxyl radical. The selection of the parameters set in the simulations is critical and can dramatically influence the final results.

Several procedures should be performed to test the correct performance of the code. The site-hit probability and G-value were used as testing parameters. Furthermore, seven other quality control tests for the clustering algorithm were carried out.

SSBY, DSBY, and TSBY were computed as a function of LET for protons and

α-particles. Their behaviors are consistent with those reported by simulations [26, 28, 46, 110] and experimental [19, 107, 108, 109] works. The use of two irradiation setups was determinant for the final results, specially for the DSBY.

The use of complex geometrical models with atomic description of the DNA molecule is mandatory to adequately evaluate the damage yields, mainly through the indirect mechanism.

The nearly constant value for the direct total TSBY with LET was confirmed regardless of the particle type. For low LET values the total TSBY is stable and tends to decrease beyond 10 $keV/\mu m$. At some point the chemical contribution became lower than the direct one. The total DSBY increases with LET for both proton and alpha particles. At the same LET,α-particles have a lower DSBY compared to lighter particles, which is a well known behavior. The mixed contribution was also reported and it tends to decrease with LET for both protons and α-particles.

The evaluated ratios of the TSB_d to TSB_i and the SSBY to DSBY agree with the corresponding ratios obtained from the results reported by the Friedland's group. Those values for low LET were very close to the ones reported in the literature for ^{60}Co.

The comparison with other simulations led to some discrepancies in terms of absolute values, which could be explained by different parameter selections: the threshold energy for a SB induction, the break probability after the chemical reaction with the deoxyribose, the time duration of the chemical stage, the reaction radius and the complexity of the damage considered in the simulation. Nevertheless, our results are in good agreement with those presented by Friedland et al. [26], even when some of the previously mentioned parameters were not identically selected.

The high dispersion and scarcity of experimental results makes difficult to conclude what biophysical approach is closest to reality. Moreover, it would be appropriate to perform an analysis of fragment size distributions to adequately compare our simulations with the experimental data, since it has been reported that small fragments can not be resolved experimentally, leading to lower experimental values of DSB yields.

The current version of the radiobiological code and the GEANT4-DNA application were developed considering future updates. Among the foreseen improvements, we could mention the inclusion of more complex damages in the clustering analysis, the con-

sideration of base damage, other radicals action in the DNA molecule like e_{aq} and $H\bullet$, and the estimation of RBE using a reference radiation quality. The latter is challenging since the simulation time needs to be increased considerably in order to account for enough number of photon tracks traversing the small cell nucleus geometry and depositing energy inside it.

Bibliography

1. Bernal, M. *et al.* An atomistic geometrical model of the B-DNA configuration for DNA-radiation interaction simulations. *Comput. Phys. Commun.* **184**, 2840–2847 (2013). Cited 8 times on pages 11, 25, 26, 44, 52, 78, 79, and 80.

2. Alloni, D., Campa, A., Friedland, W., Mariotti, L. & Ottolenghi, A. Track structure, radiation quality and initial radiobiological events: Considerations based on the PARTRAC code experience. *Comput. Phys. Commun.* **88**, 77–86 (2012). No citations in the text.

3. Štěpán, V. & Davídková, M. RADAMOL tool: Role of radiation quality and charge transfer in damage distribution along DNA oligomer. *Eur. Phys. J. D.* **68**, 240 (2014). Cited 3 times on pages 44, 47, and 81.

4. Bigildeev, E. A. & Michalik, V. Charged particle tracks in water of different phases. Monte Carlo simulation of electron tracks. *Radiat. Phys. Chem.* **47**, 197–207 (1996). Cited on page 47.

5. PUBLICATION, U. N. *Sources, effects and risks of ionizing radiation, United Nations Scientific Committee on the Effects of Atomic Radiation (UNSCEAR) 2016 Report.* (UNITED NATIONS, New York, 2017). Cited on page 22.

6. Goodhead, D. Understanding and characterisation of the risks to human health from exposure to low levels of radiation. *Radiat. Prot. Dosim.* **137**, 109–117 (2009). Cited on page 23.

7. Boice, J. D., Jr. The linear nonthreshold (LNT) model as used in radiation protection: an NCRP update. *Int J Radiat Biol.* **93**, 1079–1092 (2017). Cited on page 23.

8. Kreuzer, M. *et al.* Multidisciplinary European Low Dose Initiative (MELODI): strategic research agenda for low dose radiation risk research. *Radiat. Environ. Biophys* **57**, 5–15 (2018). Cited on page 23.

9. Goodhead, D. & Cucinotta, F. NASA space radiation summer school for research. In *Proceedings of the 63rd International Astronautical Congress 2012 (IAC 2012)*, 237–242 (Curran Associates, Inc., 2012). Cited on page 23.

10. Laplace, P. S. Théorie analytique des probabilités, 3rd rev. ed. (Paris: Veuve Courcier, 1820). Cited on page 44.

11. Palmans, H. *et al.* Future development of biologically relevant dosimetry. *Br J Radiol.* **88**, 20140392 (2015). Cited on page 23.

12. INTERNATIONAL ATOMIC ENERGY AGENCY. *Relative Biological Effectiveness in Ion Beam Therapy.* No. 461 in Technical Reports Series (INTERNATIONAL ATOMIC ENERGY AGENCY, Vienna, 2008). Cited on page 23.

13. Mestres, M., Caballín, M., Barrios, L., M., R. & Barquinero, J. RBE of X rays of different energies: A cytogenetic evaluation by FISH. *Radiat. Res.* **170**, 93–100 (2008). Cited 2 times on pages 23 and 24.

14. ICRU. Report 86: Quantification and reporting of low-dose and other heterogeneous exposures. *J ICRU* **11** (2011). Cited on page 23.

15. Hall, E. J. *Radiobiology for the radiologist, 7th edition* (Philadelphia : Wolters Kluwer Health/Lippincott Williams & Wilkins, 2012). Cited on page 24.

16. Friedland, W., Jacob, P. & Kundrát, P. Mechanistic simulation of radiation damage to DNA and its repair: on the track towards systems radiation biology modelling. *Radiat. Prot. Dosim* **143**, 542–548 (2011). Cited 2 times on pages 24 and 109.

17. Friedland, W. & Kundrát, P. Track structure based modelling of chromosome aberrations after photon and alpha-particle irradiation. *Mutat Res Genet Toxicol Environ Mutagen* **756**, 213–223 (2013). Cited on page 24.

18. 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. *Radiat. Phys. Chem.* **128**, 3–10 (2016). Cited on page 24.

19. Frankenberg, D. *et al.* Induction of DNA double-strand breaks by ^{1}H and ^{4}He ions in primary human skin fibroblasts in the LET range of 8 to 124 $keV/\mu m$. *Radiat. Res.* **151**, 540–549 (1999). Cited 5 times on pages 14, 24, 106, 112, and 114.

20. Göggelmann, W. *et al.* Re-evaluation of the RBE of 29 kV x-rays (mammography x-rays) relative to 220 kV x-rays using neoplastic transformation of human CGL1-hybrid cells. *Radiat. Environ. Biophys* **42**, 175–182 (2003). Cited on page 24.

21. Schmid, E., Krumrey, M., Ulm, G., Roos, H. & Regulla, D. The maximum low-dose RBE of 17.4 and 40 keV monochromatic X rays for the induction of dicentric chromosomes in human peripheral lymphocytes. *Radiat. Res.* **160**, 499–504 (2003). Cited on page 24.

22. Pomplun, E. A new DNA target model for track structure calculations and its first application to I-125 auger electrons. *Int. J. Radiat. Biol.* **59**, 625–642 (1991). Cited on page 24.

23. Nikjoo, H., O'Neil, P., Terrissol, M. & Goodhead, D. T. Quantitative modelling of DNA damage using monte carlo track structure method. *Radiat. Environ. Biophys.* **38**, 31–38 (1999). Cited on page 24.

24. Bernal, M. A. & Liendo., J. A. An investigation on the capabilities of the PENELOPE MC code in nanodosimetry. *Med. Phys.* **36**, 620–625 (2009). Cited 2 times on pages 24 and 25.

25. Bernal, M. *et al.* The invariance of the total direct DNA strand break yield. *Med. Phys.* **38**, 4147–4153 (2011). Cited 4 times on pages 24, 25, 52, and 104.

26. Friedland, W., Jacob, P., Bernhardt, P., Paretzke, H. & M., D. Simulation of DNA damage after proton irradiation. *Radiat. Res.* **159**, 401–410 (2003). Cited 19 times on pages 14, 24, 26, 44, 49, 62, 81, 84, 103, 104, 105, 106, 108, 109, 110, 111, 112, 113, and 114.

27. Friedland, W., Dingfelder, M., Kundrát, P. & Jacob, P. Track structures, DNA targets and radiation effects in the biophysical Monte Carlo simulation code PARTRAC. *Mutat. Res.* **711**, 28–40 (2011). Cited 12 times on pages 10, 14, 24, 26, 44, 49, 51, 81, 84, 105, 106, and 109.

28. Meylan, S. *et al.* Simulation of early DNA damage after the irradiation of a fibroblast cell nucleus using Geant4-DNA. *Sci Rep.* **7** (2017). Cited 13 times on pages 10, 14, 24, 26, 52, 53, 104, 105, 106, 107, 110, 111, and 114.

29. Lea, D. E. & Catcheside, D. G. The mechanism of the induction by radiation of chromosome aberrations in tradescantia. *J. Genet.* **44**, 216–245 (1942). Cited on page 24.

30. Kellerer, A. M. & Rossi, H. H. The theory of dual radiation action. *Curr Top Radiat Res* **8**, 85–158 (1972). Cited on page 24.

31. Guerrero, M. & Carlone, M. Mechanistic formulation of a lineal-quadratic-linear (LQL) model: Split-dose experiments and exponentially decaying sources. *Med. Phys.* **37**, 4173–4181 (2010). Cited on page 24.

32. Nikjoo, H., Uehara, S., Emfietzoglou, D. & Brahme, A. Heavy charged particles in radiation biology and biophysics. *New J. Phys* **10** (2008). Cited on page 24.

33. Nelson, W., Hirayama, H. & Rogers, D. EGS4 code system. (1985). Cited on page 24.

34. Baró, J., Sempau, J., Fernández-Varea, J. M. & Salvat, F. PENELOPE: An algorithm for Monte Carlo simulation of the penetration and energy loss of electrons and positrons in matter. *Nucl Instrum Methods Phys Res B* **100**, 31–46 (1995). Cited on page 24.

35. Agostinelli, S. *et al.* GEANT4 - a simulation toolkit. *Nucl. Instr. Meth. Phys. Res.* **506**, 250–303 (2003). Cited 4 times on pages 24, 47, 53, and 54.

36. Allison, J. *et al.* GEANT4 developments and applications. *IEEE Trans Nucl* **56**, 270–278 (2006). Cited 3 times on pages 24, 47, and 53.

37. Allison, J. *et al.* Recent developments in GEANT4. *Nucl. Instr. Meth. Phys. Res.* **835**, 186–225 (2016). Cited 3 times on pages 24, 47, and 53.

38. Uehara, S., Nikjoo, H. & Goodhead, D. Cross-sections for water vapour for the Monte Carlo electron track structure code from 10 eV to the MeV region. *Phys. Med. Biol* **38**, 1841–1858 (1993). Cited 3 times on pages 24, 45, and 57.

39. Grosswendt, B. Formation of ionization clusters in nanometric structures of propane-based tissue-equivalent gas or liquid water by electrons and α-particles. *Radiat. Environ. Biophys* **41**, 103–112 (2002). Cited on page 24.

40. Plante, I. Monte-Carlo simulation of ionizing radiation tracks. In Mode, C. J. (ed.) *Applications of Monte Carlo Methods in Biology, Medicine and Other Fields of Science*, chap. 18 (InTech, Rijeka, 2011). Cited on page 24.

41. Wälzlein, C., Scifoni, E., Krämer, M. & Durante, M. Simulations of dose enhancement for heavy atom nanoparticles irradiated by protons. *Phys. Med. Biol* **59**, 1441 (2014). Cited on page 24.

42. Chauvie, S. *et al.* GEANT4 physics processes for microdosimetry simulation: design foundation and implementation of the first set of models. *Nucl Instrum Methods Phys Res B* **54**, 2619–2628 (2007). Cited on page 25.

43. Incerti, S., Douglass, M., Penfold, S., Guatelli, S. & Bezak, E. Review of Geant4-DNA applications for micro and nanoscale simulations. *Phys. Med.* **32**, 1187–1200 (2016). Cited 5 times on pages 25, 26, 50, 56, and 73.

44. Nikjoo, H., S. Uehara, S., Emfietzoglou, D. & Cucinotta, F. A. Track-structure codes in radiation research. *Radiat. Meas* **41**, 1052–1074 (2006). Cited 3 times on pages 26, 44, and 45.

45. Nikjoo, H. & Girard, P. A model of the cell nucleus for DNA damage calculations. *Int. J. Radiat. Biol.* **88**, 87–97 (2012). Cited on page 26.

46. Lampe, N. *et al.* Mechanistic DNA damage simulations in Geant4-DNA part 2: Electron and proton damage in a bacterial cell. *Phys. Med.* **48**, 146–155 (2018). Cited 13 times on pages 10, 14, 26, 53, 85, 105, 106, 107, 108, 110, 111, 112, and 114.

47. Friedland, W. *et al.* Comprehensive track-structure based evaluation of DNA damage by light ions from radiotherapy-relevant energies down to stopping. *Sci. Rep.* **7**, 45161 (2017). Cited 8 times on pages 14, 26, 105, 106, 107, 109, 110, and 111.

48. Meylan, S., Vimont, U., Incerti, S., Clairand, I. & Villagrasa, C. GEANT4-DNA simulations using complex DNA geometries generated by the DnaFabric tool. *Comput. Phys. Commun* **204**, 159 – 169 (2016). Cited 3 times on pages 26, 107, and 108.

49. Karamitros, M. *et al.* Diffusion-controlled reactions modeling in Geant4-DNA. *J. Comput. Phys.* **274**, 841–882 (2014). Cited 9 times on pages 12, 26, 50, 60, 65, 69, 74, 81, and 97.

50. 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. *Phys. Med.* **31**, 861–874 (2015). Cited 3 times on pages 26, 51, and 57.

51. Incerti, S. *et al.* Comparison of GEANT4 very low energy cross section models with experimental data in water. *Med. Phys* **37**, 4692–4708 (2010). Cited 3 times on pages 26, 50, and 57.

52. Gunderson, L. & Tepper, J. *Clinical Radiation Oncology, 4th edition* (Elsevier, 2015). Cited 4 times on pages 9, 29, 33, and 34.

53. ICRU Report 16: Linear Energy Transfer *J ICRU* (1970). Cited on page 29.

54. Attix, F. H. *Introduction to Radiological Physics and Radiation Dosimetry* (WILEY-VCH Verlag GmbH & Co.KGaA, 2007). Cited 2 times on pages 9 and 30.

55. Joiner, M. & Van der Kogel, A. *Basic clinical radiobiology, 4th edition* (Hodder Arnold, 2009). Cited 3 times on pages 9, 32, and 33.

56. Contributors Wikipedia. Wikipedia, the free encyclopedia (2018). URL https://en.wikipedia.org/. [Online; accessed 10-February-2018]. Cited 3 times on pages 9, 36, and 39.

57. Annunziato, A. DNA packaging: Nucleosomes and chromatin. (2008). Cited 3 times on pages 9, 35, and 37.

58. Shmoop Editorial Team, DNA Structure, Replication, and Technology - Shmoop (2018). URL https://www.shmoop.com/dna/dna-replication.html. [Online; accessed 05-March-2018]. Cited 2 times on pages 9 and 38.

59. Tomita, H., Kai, M., Kusama, T. & Ito, A. Monte Carlo simulation of DNA strand-break induction in supercoiled plasmid pBR322 DNA from indirect effects. *Radiat. Environ. Biophys* **36**, 235–241 (1998). Cited on page 37.

60. Savage, J. Update on target theory as applied to chromosomal aberrations. *Environ Mol Mutagen* **22**, 198–207 (1993). Cited 2 times on pages 38 and 39.

61. Olive, PL. The role of DNA single and double-strand breaks in cell killing by ionizing radiation.. *Radiat Res* **150**, 42–51 (1998). Cited on page 38.

62. Calini, V., Urani, C. & Camatini, M., Comet assay evaluation of DNA single- and double-strand breaks induction and repair in C3H10T1/2 cells.. *Cell Biol Toxicol* **18**, 369–79 (2002). Cited on page 38.

63. Basu, S. & Wiklund, L. *Studies on Experimental Models* (Humana Press, c/o Springer Science+Business Media, 2011). Cited 5 times on pages 9, 10, 39, 41, and 43.

64. Elkind, M. & Kamper, C. Two forms of repair of DNA in mammalian cells following irradiation. *Biophys J* **10**, 237–245 (1970). Cited on page 41.

65. Zwelling, L. e. a. Protein-associated deoxyribonucleic acid strand breaks in L1210 cells treated with the deoxyribonucleic acid intercalating agents 4'-(9-acridinylamino) methanesulfon-m-anisidide and adriamycin. *Biochemistry* **20**, 6553–6563 (1981). Cited on page 41.

66. Geigl, E. & Eckardt-Schupp, F. The repair of double-strand breaks and S1 nuclease-sensitive sites can be monitored chromosome-specifically in Saccharomyces cerevisiae using pulse-field gel electrophoresis. *Mol Microbiol* **5**, 1615–1620 (1991). Cited on page 41.

67. Herschleb, J., Ananiev, G. & Schwartz, D. Pulsed-field gel electrophoresis. *Nat. Protoc* **2**, 677–684 (2007). Cited on page 41.

68. Olive, P., Banath, J. & Durand, R. Detection of etoposide resistance by measuring DNA damage in individual chinese hamster cells. *J Natl Cancer Inst* **82**, 779–783 (1990). Cited on page 42.

69. Lee, R. F. & Steinert, S. Use of the single cell gel electrophoresis/comet assay for detecting DNA damage in aquatic (marine and freshwater) animals. *Mutat. Res* **544**, 43–64 (2003). Cited on page 42.

70. Pilch, D. *et al.* Characteristics of gamma-H2AX foci at DNA double-strand breaks sites. *Biochem Cell Bio* **81**, 123–129 (2003). Cited on page 42.

71. Redon, C. E., Nakamura, A. J., Gouliaeva, K., Rahman, A., Blakely, W. F. & Bonner, W. M. The Use of Gamma-H2AX as a Biodosimeter for Total-Body Radiation Exposure in Non-Human Primates. *PLOS ONE* **5**, 1–8 (2010). Cited on page 42.

72. Nikjoo, H., O'Neill, P., Goodhead, D. & Terrissol, M. Computational modelling of low-energy electron-induced DNA damage by early physical and chemical events. *Int. J. Radiat. Biol* **71**, 467–483 (1997). Cited 8 times on pages 44, 45, 53, 84, 85, 104, 107, and 108.

73. Beaudré, A. Simulation spatio-temporelle des processus radiolytiques induits dans l'eua par des électrons. *Thése de l' Université Paul Sabatier, Toulouse, France* **371** (1988). Cited on page 45.

74. Terrissol, M. & Beaudré, A. Simulation of space and time evolution of radiolytic species induced by electrons in water. *Radiat. Prot. Dosim* **31**, 171–175 (1990). Cited 2 times on pages 45 and 107.

75. Buxton, G., Greenstock, C., Helman, W. & Ross, A. Critical review of rate constants for reactions of hydrated electrons, hydrogen atoms and hydroxyl radicals ($\cdot OH$-O^-) in aqueous solution. *J. Phys. Chem. Ref. Data* **17**, 513–886 (1988). Cited 5 times on pages 15, 46, 50, 81, and 82.

76. Kreipl, M. S., Friedland, W. & Paretzke, H. Time- and space-resolved monte carlo study of water radiolysis for photon, electron and ion irradiation. *Radiat Environ Biophys* **48**, 11–20 (2009). Cited 10 times on pages 12, 15, 46, 61, 62, 64, 65, 68, 69, and 97.

77. Ballarini, F. *et al.* Stochastic aspects and uncertainties in the prechemical and chemical stages of electron tracks in liquid water: a quantitative analysis based on Monte Carlo simulations. *Radiat Environ Biophys* **39**, 179–188 (2000). Cited 2 times on pages 46 and 49.

78. Chakraborty, T. (ed.) *Charge migration in DNA* (Springer-Verlag Berlin Heidelberg, 2007). Cited on page 48.

79. Friedland, W., Jacob, P., Paretzke, H. & Stork, T. Monte Carlo simulation of the production of short DNA fragments by low-linear energy transfer radiation using high order DNA model. *Radiat. Res.* **150**, 170–182 (1998). Cited on page 49.

80. Friedland, W., Jacob, P., Paretzke, H., Merzagora, M. & Ottolengui, A. Simulation of DNA fragments distribution after irradiation with photons. *Radiat. Environ. Biophys.* **38**, 39–47 (1999). Cited on page 49.

81. Incerti, S. *et al.* The Geant4-DNA project. *Int. J. Model. Simul. Sci. Comput.* **1**, 157–178 (2010). Cited 4 times on pages 15, 50, 55, and 58.

82. Francis, Z., Villagrasa, C. & Clairand, I. Simulation of DNA damage clustering after proton irradiation using an adapted DBSCAN algorithm. *Comput Methods Programs Biomed.* **101**, 265–270 (2011). Cited on page 51.

83. Francis, Z. *et al.* Monte Carlo simulation of energy-deposit clustering for ions of the same LET in liquid water. *Phys. Med. Biol.* **57**, 209–224 (2012). Cited 2 times on pages 51 and 104.

84. Francis, Z. & Stypczynska, A. Simulation of DNA damage clustering after proton irradiation using an adapted DBSCAN algorithm. *Data mining: New Technologies, Benefits and Privacy Concerns* (2013). Cited on page 51.

85. Ester, M., Kriegel, H., Sander, J. & Xu, X. A density-based algorithm for discovering clusters in large spatial databases with noise. *Proc 2nd Intl Conf Knowledge Discovery and Data mining* (1996). Cited on page 51.

86. Incerti, S. *et al.* Energy deposition in small scale targets of liquid water using the very low energy electromagnetic physics processes of the Geant4 toolkit. *Nucl Instrum Methods Phys Res B* **306**, 158–164 (2013). Cited on page 52.

87. Bernal, M. *et al.* Performance of a new atomistic geometrical model of the B-DNA configuration for DNA-radiation interaction simulations. *J. Phys. Conf. Ser* **490** (2014). Cited on page 52.

88. Delage, E. *et al.* PDB4DNA: Implementation of DNA geometry from the Protein Data Bank (PDB) description for Geant4-DNA Monte Carlo simulations. *Comput. Phys. Commun* **192**, 282–288 (2015). Cited on page 52.

89. Lampe, N. *et al.* Mechanistic DNA damage simulations in Geant4-DNA part 1: A parameter study in a simplified geometry. *Phys. Med.* **48**, 135–145 (2018). Cited 4 times on pages 53, 85, 107, and 108.

90. Ivanchenko, V. N. *et al. GEANT4 Physics reference manual v10.2* (2015). Cited on page 54.

91. Emfietzoglou, D. Inelastic cross-sections for electron transport in liquid water: a comparison of dielectric models. *Radiat. Phys. Chem.* **66**, 373–385 (2003). Cited on page 57.

92. Kyriakou, I., Incerti, S. & Francis, Z. Technical note: Improvements in Geant4 energy-loss model and the effect on low-energy electron transport in liquid water. *Med. Phys.* **42**, 3870–3876 (2015). Cited on page 57.

93. Kyriakou, I., Šefl, M., Nourry, V. & Incerti, S. The impact of new Geant4-DNA cross section models on electron track structure simulations in liquid water. *J. Appl. Phys* **119** (2016). Cited on page 57.

94. Grosswendt, B. & Waibel, E. Transport of low energy electrons in nitrogen and air. *Nucl. Instr. Meth. Phys. Res* **155**, 145–156 (1978). Cited on page 57.

95. Kim, Y. K. & Rudd, M. E. Binary-encounter-dipole model for electron-impact ionization. *Phys. Rev. A* **50**, 3954 (1994). Cited on page 57.

96. Mott, N. & Massey, H. The theory of atomic collisions. *Oxford: Clarendon Press* (1965). Cited on page 57.

97. Michaud, M. & Sanche, L. Total cross sections for slow-electron (1–20 eV) scattering in solid H_2O. *Phys. Rev. A* **36**, 4672 (1987). Cited on page 57.

98. Kernbaum, M. Décomposition de l'eau par l'aigrette. *Le Radium* **7**, 275–278 (1910). Cited on page 59.

99. Karamitros, M. *et al.* Modeling radiation chemistry in the Geant4 toolkit. *Nucl. Sci. Tech* **2**, 503–508 (2011). Cited on page 60.

100. Karamitros, M. *Extension de l'outil Monte Carlo généraliste Geant4 pour la simulation de la radiolyse de l'eau dans le cadre du projet Geant4-DNA*. Ph.D. thesis, École Doctorale des Sciences Physiques et Ingénieries, L'Université Bordeaux 1 (2012). Cited 2 times on pages 44 and 60.

101. Ermak, D. L. & McCammon, J. A. Brownian dynamics with hydrodynamic interactions. *J. Chem. Phys* **69**, 1352–1360 (1978). Cited on page 63.

102. Pimblott, S. M. & LaVerne, J. A. Scavenger concentration dependences of yields in radiation chemistry. *J. Phys. Chem* **96**, 746–752 (1992). Cited on page 69.

103. GLACTONE. The glactone project. (2002). URL http://chemistry.gsu.edu/Glactone/PDB/DNA_RNA/dna.html. Cited on page 79.

104. Tartier, L. *et al.* Radiolytic signature of Z-DNA. *Nucleic Acids Res* **22**, 5565–5570 (1994). Cited on page 81.

105. Aydogan, B., Bolch, W., Swarts, S., Turner, J. & Marshall, D. Monte Carlo simulations of site-specific radical attack to DNA bases. *Radiat. Res.* **169**, 223–231 (2008). Cited on page 81.

106. Bernal, M. A. Evaluation of the mean energy deposit during the impact of charged particles on liquid water. *Phys. Med. Biol* **57**, 1745–1757 (2012). Cited on page 104.

107. Belli, M. *et al.* DNA DSB induction and rejoining in V79 cells irradiated with light ions: a constant field gel electrophoresis study. *Int. J. Radiat. Biol.* **76**, 1095–1104 (2000). Cited 3 times on pages 14, 106, and 114.

108. Belli, M. *et al.* DNA fragmentation in mammalian cells exposed to various light ions. *Adv. Space Res.* **27**, 393–399 (2001). Cited 6 times on pages 14, 42, 105, 106, 112, and 114.

109. Campa, A. *et al.* DNA DSB induced in human cells by charged particles and gamma rays: Experimental results and theoretical approaches. *Int. J. Radiat. Biol.* **81**, 841–854 (2005). Cited 4 times on pages 14, 42, 106, and 114.

110. Nikjoo, H., O'Neil, P., Wilson, W. E. & Goodhead, D. T. Computational approach for determining the spectrum of DNA damage induced by ionizing radiation. *Radiat Res.* **156**, 577–83 (2001). Cited 5 times on pages 14, 105, 106, 107, and 114.

111. Rydberg, B. *et al.* Spatial distribution and yield of DNA double-strand breaks induced by 3-7 MeV helium ions in human fibroblasts. *Radiat. Res.* **158**, 32–42 (2002). Cited 3 times on pages 14, 106, and 108.

112. Bordage, M. *et al.* Implementation of new physics models for low energy electrons in liquid water in Geant4-DNA. *Phys. Med.* **32**, 1833–1840 (2016). Cited 2 times on pages 107 and 108.

113. Zein, S. *et al.* Microdosimetry in 3D realistic mitochondria phantoms: Geant4 Monte Carlo tracking of 250keV photons in phantoms reconstructed from microscopic images. *Phys. Med.* **42**, 7–12 (2017). Cited on page 107.

114. Boudaïffa, B., Cloutier, P., Hunting, D., Huels, M. A. & Sanche, L. Cross sections for low-energy (10–50 eV) electron damage to DNA. *Radiat Res* **157**, 227–234 (2001). Cited on page 109.

115. Panajotovic, R., Martin, F., Cloutier, P., Hunting, D. & Sanche, L. Effective cross sections for production of single-strand breaks in plasmid DNA by 0.1 to 4.7 eV electrons. *Radiat Res* **165**, 452–459 (2006). Cited on page 109.

116. Boudaïffa, 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). Cited on page 109.

117. Bernhardt, P., Friedland, W., Jacob, P. & Paretzke, H. Modeling of ultrasoft X-ray induced DNA damage using structured higher order DNA targets. *Int. J. Mass Spectrom* **223-224**, 579 – 597 (2003). Cited on page 109.

118. Leloup, C. *et al.* Evaluation of lesion clustering in irradiated plasmid DNA. *Int J Radiat Biol.* **81**, 41–54 (2005). Cited 2 times on pages 14 and 112.

119. Vyšín, L. *et al.* Proton-induced direct and indirect damage of plasmid DNA. *Radiat Environ Biophys* **54**, 343–352 (2015). Cited 2 times on pages 14 and 112.

120. Souici, M. *et al.* Single- and double-strand breaks of dry DNA exposed to protons at bragg-peak energies. *J. Phys. Chem. B* **121**, 497–507 (2017). Cited 2 times on pages 14 and 112.

APPENDIX A – Publications and Events

Publications

1. Falzone N., Ackerman N., de la Fuente L., Bernal M., Liu X., Peeters S., Sarmiento M., Corroyer-Dulmont A., Grimoin E., Bernaudin M., Touzani O., Sibson N., Vallis K., "Dosimetric evaluation of radionuclides for VCAM-1-targeted radionuclide therapy of early brain metastases", Theranostics 2018; 8(1): 292-303.

   This article was the result of an international collaboration and was part of the work performed during this PhD research. It was published at a high impact journal "Theranostics" (IF 8.766) and was also selected as back cover of that issue (Fig. 49). This study evaluated the efficacy of α-emitting and β-emitting radionuclides, specifically ^{212}Pb and ^{177}Lu for treating early brain metastasis. The geometry was conceived from the results obtained by photon microscopy of mouse brain parenchyma and Monte Carlo simulations were carried out using the GEANT4-DNA toolkit. Our contribution to this work was during the simulations performed in our group cluster and in the further post-processing with the radiobiological code in order to estimate DNA damage yields. Also, the Relative Biological Effectiveness (RBE) was determined from the DSBY obtained due to physical interactions, using the ^{60}Co as the reference beam. Although in this work only the direct action was considered since the indirect module was in the testing stage, it greatly influenced the first stage of the current thesis. This research showed that the ^{212}Pb, which has the advantage that can be used for SPECT imaging, has a favorable dose profile and RBE.

2. de la Fuente Rosales L., Bernal M. A., "Accounting for radiation-induced indirect damage on DNA with the GEANT4 code", Physica Medica, Available online 19 June 2018, ISSN 1120-1797, https://doi.org/10.1016/j.ejmp.2018.06.006. (https://www.sciencedirect.com/science/article/pii/S1120179718304873) Invited full paper in the Special Edition of Physica Medica journal (MCMA2017)- European

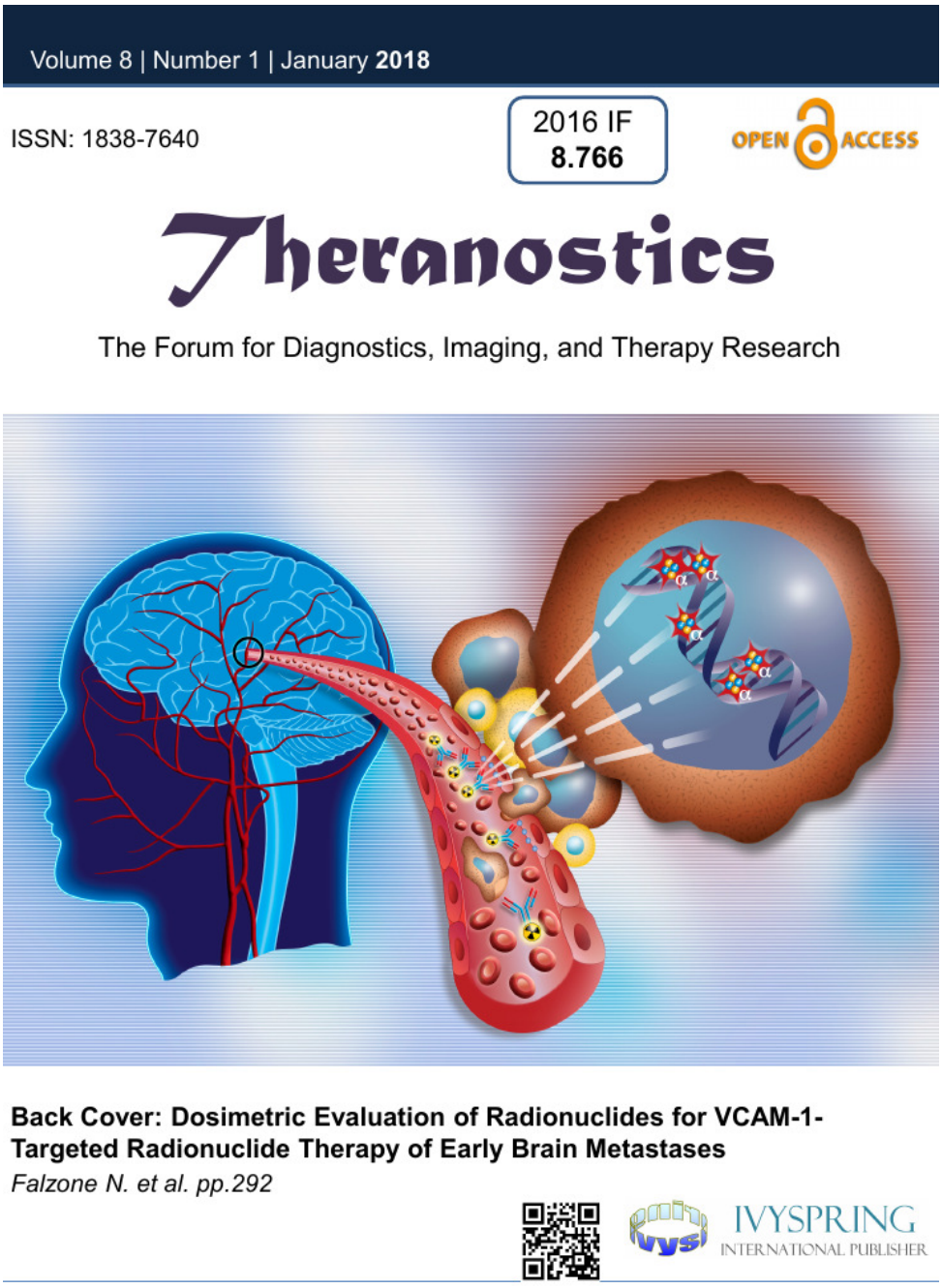

Figure 49 – Back cover *Theranostics* journal (Theranostics 2018; 8(1): 292-303).

Journal of Medical Physics.

This work was presented at the International Conference on Monte Carlo Techniques for Medical Applications (MCMA) and invited as a full article for the special issue of this event in the Physica Medica journal. The scope of the PhD thesis and the results obtained were included in this article.

3. Ackerman N., de la Fuente Rosales L., Falzone N., Vallis K. A., Bernal Rodriguez M., "Change of RBE of ^{225}Ac and ^{212}Pb as a Consequence of Recoiling Daughters", Physica Medica, 2018,ISSN 1120-1797, https://doi.org/10.1016/j.ejmp.2018.05.020. Invited full paper in the Special Edition of Physica Medica journal (MCMA2017)- European Journal of Medical Physics.

   Also presented at the International Conference on Monte Carlo Techniques for Medical Applications (MCMA) and invited as a full article for the special issue of this event in the Physica Medica journal. This work is the second one developed as part of the international collaboration already mentioned. Again we had an active role in this work, performing the simulation runs, the damage yield determination and supporting the manuscript writing. This research was focused on the comparison of two α-emitting radionuclides for targeted therapy, in this particular case ^{225}Ac and ^{212}Pb were studied. The physical dose and DNA damage was calculated using the GEANT4-DNA and the radiobiological code for both radionuclides. The results showed that the ^{225}Ac has a higher dose and double strand break yield per decay than ^{212}Pb. The major concern in this research was related to the efficacy of both nuclides on retaining the daughter nuclei at the target location in the brain vasculature. When the entire decay chain was considered the RBE of ^{225}Ac and ^{212}Pb were similar. If the initial daughter is lost, the RBE of ^{212}Pb is completely reduced to 1 or lower and the RBE of ^{225}Ac is approximately 2 only for the first 40 μm.

Events

1. de la Fuente Rosales L., Bernal Rodriguez M., Abstract ID: 90 "Accounting for radiation-induced indirect damage on DNA with the GEANT4 code", Vol. 42, p19, Physica Medica, International Conference on Monte Carlo Techniques for Medical Applications MCMA, Napoli, Italy, 2017. (Poster)

2. de la Fuente Rosales L., Bernal Rodriguez M., "Upgrading a biophysical model to compute radiation induced damage on a DNA molecule with atomic resolution", III School on Medical Physics, Joint Conference LASNPA & WONPNURT, Havana, Cuba, 2017. (Oral presentation)