A MONTE CARLO INVESTIGATION OF RADIATION DAMAGE TO CHROMATIN FIBERS AND PRODUCTION OF DNA DOUBLE STRAND BREAKS USING GEANT4-DNA CODE

A Thesis
Presented to
The Academic Faculty

by

Brian H Lee

In Partial Fulfillment
of the Requirements for the Degree
Masters of Science in Medical Physics in the
School of Mechanical Engineering

Georgia Institute of Technology
December 2014

A MONTE CARLO INVESTIGATION OF RADIATION DAMAGE TO CHROMATIN FIBERS AND PRODUCTION OF DNA DOUBLE STRAND BREAKS USING GEANT4-DNA CODE

Approved by:

Professor C-K Chris Wang, Advisor
School of Mechanical Engineering
Georgia Institute of Technology

Professor Eric Elder
School of Mechanical Engineering
Georgia Institute of Technology

Professor Nolan Hertel
School of Mechanical Engineering
Georgia Institute of Technology

Date Approved: November 25, 2014

ACKNOWLEDGEMENTS

I am immensely grateful to my advisor, Dr. Chris Wang, for guiding and mentoring me throughout the many challenges I have faced over the course of research. I would like to extend my gratitude to Dr. Eric Elder for his experience and for making time to provide insight. I would also like to thank Dr. Nolan Hertel for helping me to see broader opportunities in research. I am also grateful to Dr. Franklin Dubose for his timely support and advice.

I would also like to thank past and present members of my research group, Matthew Coghill, Spenser Lewis, and Corey Ginetz, as well as David Zabriskie, Gregory Szalkowski, Serdar Charyyev, and Jenna Fair.

I cannot even begin to express how grateful I am for my wonderful wife, Gloria Lee, whose love and support has been essential as a motivation and an inspiration for me. I would also like to thank my parents, Kyungsub and Bongsoo Lee, for their sacrifices and for instilling the work ethic that has made much of this effort possible.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Radiobiology is the study of the effect of ionizing radiation on living things. The field was introduced shortly following Wilhelm Röntgen's 1896 discovery of the X-ray, largely due to observations that X-rays were inducing biological effects[28]. Early studies were simple; X-ray exposure stopped the growth of a hairy mole, ionizing radiation was capable of inducing skin erythema, and radium was effective at treating cancerous tissue[28]. The understanding from these studies expanded rapidly over the past century, from the observation that different types and sources of ionizing radiation exist, to the different biological effects induced by ionizing radiation, and even the method by which ionizing radiation is capable of indirectly causing damage to the body through chemical action.

Two main areas where radiobiology is applied is in healthcare and radiation protection, for cancer treatment and for preventing radiation induced cancers[1]. In modern research, radiobiology involves collaboration between physics, chemistry, and biology, and its complexity necessitates mathematical, statistical and computational models[6][30]. Several historical formulae are prevalent in the field, largely due to their simplicity and effectiveness; however, analytical models based on older cell models require updating to provide a sufficient foundation for future research, as recent discoveries highlight the intricacies of cell biology in greater detail[36]. The currently accepted cell environment and DNA structure differs from what previous models were based upon, and there is a greater need for more detailed physics models to be applied that account for these[16][10].

The basic purpose of the presented research is to improve on the historically accepted radiobiological models through the application of a Monte Carlo simulation of radiation tracks passing through a cell nucleus modeled with the up-to-date subnuclear structures.

CHAPTER II

BACKGROUND

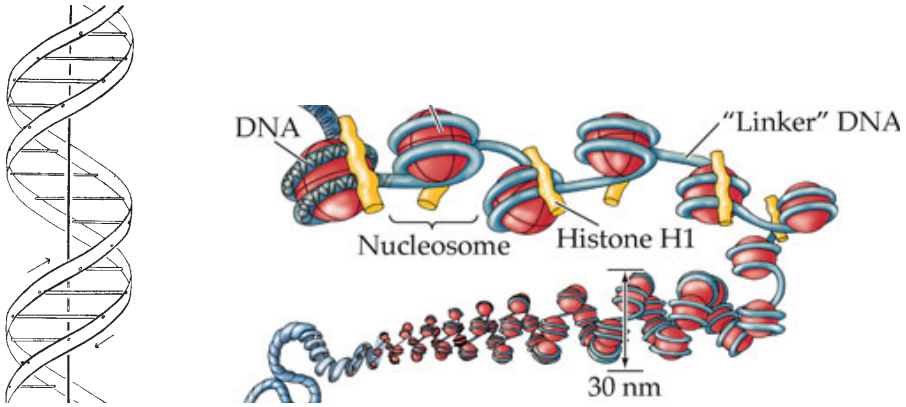
2.1 Microdosimetry and Nanodosimetry

Microdosimetry is a subset radiobiology that was introduced in the 1960's, which postulated the DNA-filled nucleus as the primary target of radiation damage[27]. In microdosimetry, the cell nucleus itself and the energy deposition within the cell nucleus were sufficient to be used as targets for radiation damage, with micro- representing the micrometer range as the target size. Several reasons factored into this, among which the biological effect of mitotic cell death involved chromosome aberrations in the nucleus, the analytical models were largely assumed to be in the 0.1 to 1 μm range, and the Rossi counter used to experimentally simulate small tissue volumes approximated the μm^3 range.

While it was understood that DNA itself is what ionizing radiation caused damage to, the actual organization of DNA was unknown at the time. The double helix had been imaged with X-ray crystallography, and chromosomes had been imaged around mitosis, but it would be several decades before more precise imaging techniques would reveal the extent to which DNA is organized in the cell nucleus. The current phase of research investigates nanodosimetry, wherein the nanometer range is used as the target size; this entails targeting key structures within the nucleus. Nanodosimetry follows the next logical progression from microdosimetry; instead of looking at the energy deposition in the DNA containing nucleus, the direct hits to DNA are modeled. This approach, owing to the alarming rate at which discoveries are being made and the scope of what is yet unknown, is complex and the potential for different models are nearly limitless[51][15][53].

2.2 DNA Double Helix, Nucleosomes, Chromatin Fibers, and Chromosome Domains

In 1953, James Watson and Francis Crick, with the help of Rosalind Franklin and her expertise in X-ray crystallography, published their seminal paper on the double-helix



(a) The DNA double-helix[52]

(b) The Nucleosome and Chromatin Fibers[45]

Figure 1: DNA as a double helix, compacted around nucleosomes and into chromatin fibers

structure of deoxyribose nucleic acid (DNA), shown in Fig. 1 [52]. At this point in history, radiotherapy had already been (primitively) applied in the treatment of cancers for several decades prior to the discovery of the actual target of radiotherapy. The DNA sequence, now known to encode the genetic information within the cell nucleus for every eukaryotic organism, is the consensus primary target in radiobiology. Radiation effects, desired or not, are mainly attributed to DNA damage followed by misrepair events, and this can manifest in the forms of mitotic cell death and carcinogenic mutations[37].

DNA has a double-helix structure in its most relaxed form, which has a diameter of approximately 2nm, with the sugar-phosphate backbone surrounding the purine-to-pyrimidine base pairs[52]. In a human cell nucleus, there are approximately 6 billion base pairs of DNA encoding 23 diploid chromosomes. With approximately 10 base pairs per 3.4 nm length of DNA, this equates to a total length of over 1 meter of DNA. In order for the DNA to fit into a cell nucleus, which is on the order of 11 μm diameter, several orders of compaction must take place.

The first step of DNA compaction involves nucleosomes comprised of histone proteins[39]. The DNA double helix wraps around the core histone protein structure approximately two times. This configuration resembles "beads on a string," with a bead diameter of 10 nm. Further modification of the histones allows for tighter compaction of the beads with one

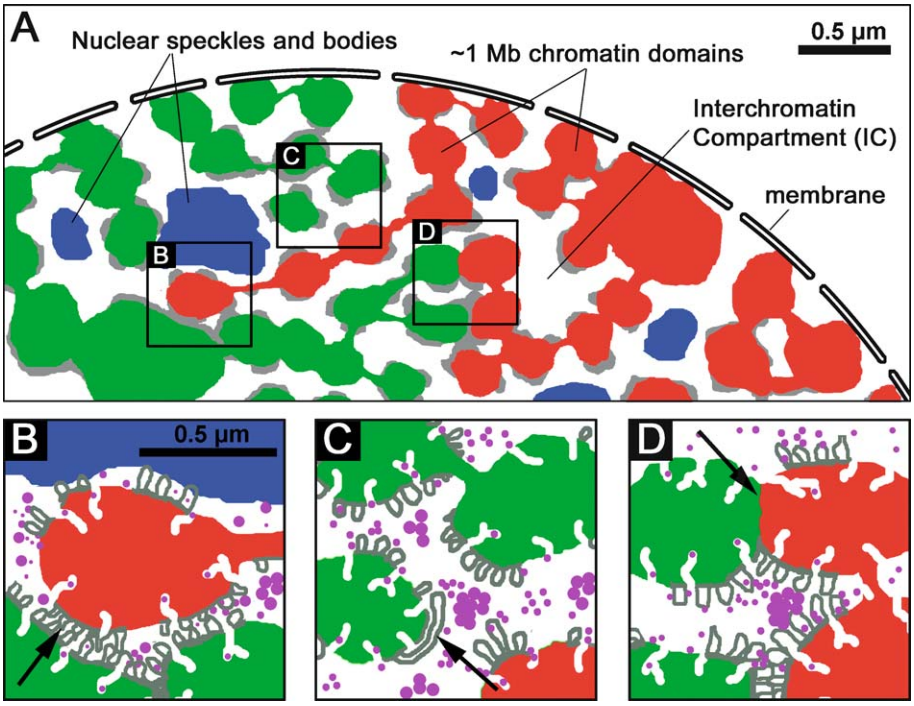


Figure 2: DNA compaction into chromatin domains[4]

another into a 30nm chromatin fiber structure[43] as shown in Fig. 1. This is the configuration that will be subsequently referred to in the proposed model as the chromatin fibers (CF). Current cancer biology studies are heavily invested in epigenetics, of which a large portion of research is dependent on histone proteins and their involvement in gene expression. While the proposed model does not directly address the intricacies of epigenetics, the spatial distribution of radiation damage to the chromatin fiber should inherently provide correlations relevant to epigenetics studies involving high-LET radiation.

The fully compacted chromosomes just before mitosis, the most recognizable form to most readers, are distinctly larger, with a metaphase width on the order of 1400nm[28]. Since cells spend the majority of their time in the G1/interphase state, which is a relaxed form that allows for enzymes to have access to the DNA for normal functions, the proposed cell model approximates the cell in the G1 phase. While the DNA is in the G1 phase is more relaxed than in mitosis, more recent discoveries using fluorescence imaging techniques demonstrate that there is a further level of compaction beyond the chromatin fiber during G1 phase[4]. Chromosomes occupy defined territories in the G1 phase[16]. Figure 2 shows

that within these chromosome regions, DNA is further organized into distinct Chromatin Domains (CDs), which each contain approximately 1 MBp of DNA and have an average diameter of approximately 400nm. Each of these CDs contain chromatin fibers, and interchromatin compartments (IC's) separating the CDs allow for enzymes to freely move around and access sections of DNA, allowing for gene expression and other major cellular functions to occur.

2.3 DNA Damage, Chromosome Aberrations, Repair and Misjoining Mechanisms

The field of radiobiology is based on the common knowledge that ionizing radiation causes damage to cells. Ionizing radiation causes damage to DNA through indirect and direct action to DNA as well as through the formation of reactive chemical species in close spatial proximity to DNA. Both direct as well as indirect damage to DNA are known to be capable of causing single strand breaks (SSBs) in DNA, wherein one side of the double helix's structure is lysed. Single strand breaks in isolation are rarely a major concern, as the overall structure of DNA is kept intact and these in turn can be repaired by various repair enzymes in the cell nucleus.

The primary interactions of interest in radiobiology are those that lead to DNA double strand breaks (DSBs), wherein the DNA is fractured into two fragmented ends. In general, two SSBs within 10 base pairs of each other on opposite strands, or 2nm, are capable of causing a DSB; SSBs that are too far apart are unlikely to weaken the DNA structure sufficiently to cleave the DNA[49]. It is also convenient to distinguish simple double strand breaks (sDSBs) from complex double strand breaks (cDSBs) defined as multiple sDSBs in close proximity or a DSB accompanied by an SSB or base deletion[50], as the distinction of DSB severity is an important factor for radiobiological consequence of DNA damage.

DNA damage is a normal occurrence in cells, and repair mechanisms for dealing with them are abundant. There are a very large number of enzymes that play a role in DNA repair, and these deal with SSBs as well as DSBs. The most important mechanisms identified for radiobiology are non-homologous end joining (NHEJ) and homologous recombination (HR)[50]. HR is conservative, slower, and more common in the DNA replication S phase

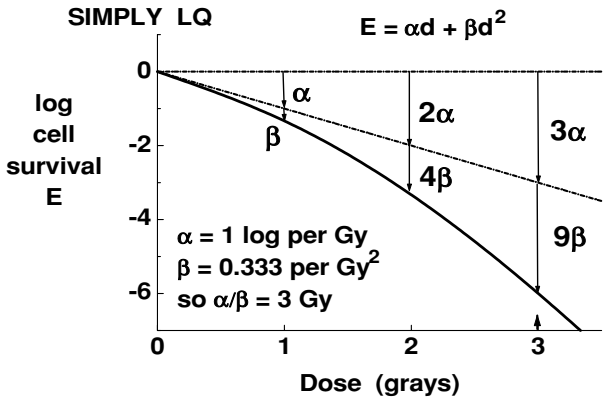


Figure 3: Linear Quadratic Units and Relationships (Fowler, 2006)

of the cell cycle and repairs the broken DNA ends by copying from the diploid sister chromosome in the cell nucleus; HR plays only a small role in DSB repair during the G0/G1 phase[25]. NHEJ is a much faster and much less accurate repair mechanism, wherein two broken ends are quickly spliced together and is the dominant repair mechanism in G0/G1 phases of the cell cycle[25]. As NHEJ is nonconservative, its repair response to DSBs can lead to the misjoining of DNA.

DNA misjoining is the major endpoint in radiobiology, both in radiation protection as well as in radiotherapy. Misjoining of DNA damage can lead to lethal chromosomal aberrations, which is the target of radiotherapy of cancerous tissue. By creating lethal chromosome aberrations, cancer cells upon attempting mitosis will be arrested by inseparable daughter chromatids and will eventually undergo apoptosis[28]. In radiation protection, DNA misrepair of healthy tissue exposed to ionizing radiation can lead to carcinogenesis. Healthy tissue undergoing apoptosis is of little concern for the radiation exposure deemed to be carcinogenic, and as such lethal chromosome aberrations are not nearly as large of a concern as the nonlethal types. Nonlethal chromosome aberrations may be as simple as chromosomal translocations, insertions or deletions that can lead to the loss of tumor suppressor or housekeeping gene functions or the over expression of oncogenic factors in normal cells, all of which can eventually lead to carcinogenesis[28].

2.4 Linear Quadratic Model and Biologically Equivalent Dose

The linear quadratic (LQ) equation is a deterministic approach used in radiobiology to plot cell survival versus dose delivered and is shown in Fig. 3. The cell survival S is:

$$S = e^{-(\alpha D + \beta D^2)} \tag{1}$$

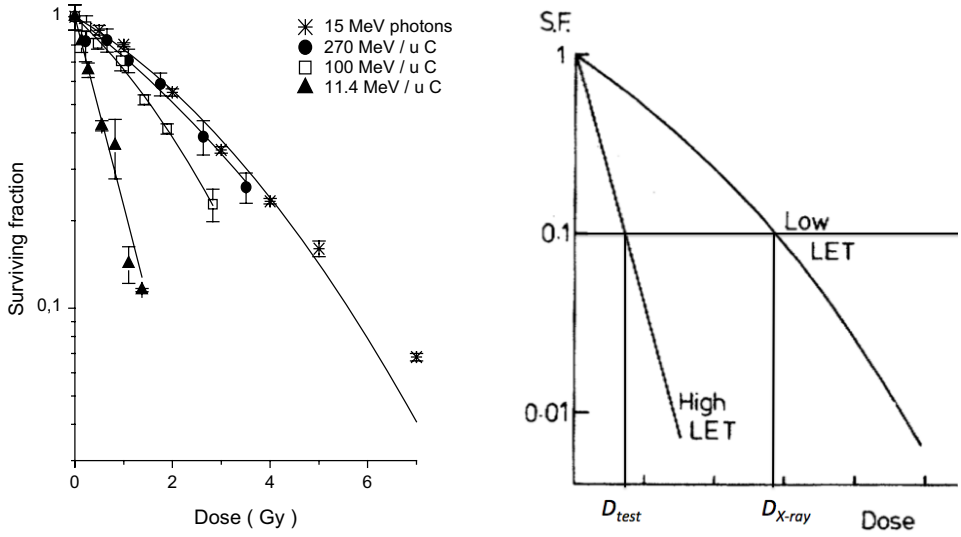
where D is the dose, α is the linear portion of the curve corresponding, while β is the quadratic portion of the curve related to D^2 corresponding to the interactions between tracks[28]. α is noted to be the initial slope of the curve and is relevant to the single track effect which will be discussed in more detail, while β is more involved in interactions between different ionizing radiation tracks in a cell nucleus.

The major significance of the LQ equation is that it is the primary correlation between delivered dose and cell survival applied in the field of radiation oncology[13]. A principle of the LQ relationship and its usefulness in the conventional photon radiation therapy is that most cancer cells have a higher value of α/ β than non-cancerous cells; hence, via a multi fraction dose delivery scheme, the same dose of radiation applied will have a more severe effect on a tumor while sparing a patient's critical organs[20].

A flaw in oversimplifying either cancer or radiation interactions with such a deterministic approach is that both of these fields are fundamentally stochastic. The LQ formalism is a very relevant, helpful, and readily applied approach to understanding radiation responses. However, there are a lot of fundamental issues in using it directly and accepting it outright; there are so many different α and β values for different cell lines, radiation types, and other factors that there is little consensus on how to tabulate it consistently. A major rationale for introducing a fully Monte Carlo approach is, simply, that it is how radiobiology is supposed to be handled. Stochastically simulating a cell model with the track structure should provide insight in how to more appropriately match the current LQ formalisms.

2.5 Relative Biological Effectiveness

Relative biological effectiveness (RBE) is a key term used to describe the effectiveness of a particular ionizing radiation relative to a standard type, which in effect is photon or



(a) Surviving fractions in CGL1 cells for heavy ions versus photons[11]

(b) High LET vs low LET survival curves[27]

Figure 4: Survival curves and RBE relationships

electron radiation. The rationale of defining the RBE of a particle is due to observations that certain types of radiation are more effective in inducing biological end points than are others[10]. General formulations have been developed by assigning specific RBE values to different particle types, which enable health physicists to simplify dosimetric calculations for the purposes of radiation shielding and personnel monitoring. ICRU and NCRP values are largely based on maintaining a safe working environment, and highlight the escalated hazards of certain types of radiation. The RBE is given by:

$$RBE = \frac{D_{X-ray}}{D_{test}} \tag{2}$$

where D_{test} is the dose of the test radiation in question required to illicit a specific biological response and $D_{X\text{-}ray}$ is the dose of X-ray radiation required to illicit the same biological response as the test radiation, which is visualized in Fig. 4 [11]. The assumption in the RBE equation is that the X-ray dose is in the numerator because it is less effective to induce a cellular response than the test radiation type, and in many cases it is given as the amount of radiation necessary to yield a 10% cell survival fraction.

High RBE treatment methods are desirable for dealing with radio-resistant cancer types,

as they have a poorer survival prognosis for patients as they tend to respond poorly to radiation within the constraints of critical organ dose limits. The increased concentration of damage from high RBE particles are more difficult to repair by radio-resistant cell lines, and the chance of inducing lethal damage to the tumor can be increased greatly[35]. While increased RBE is a highly desirable effect, in practice an increased RBE is difficult to achieve and effectively apply[5][23][24]. Figure 4 presents a plot that compares experimental survival curves for cells irradiated with 15-MeV photons as well as with carbon ions at various energies. Densely ionizing radiation types are considered to have a higher RBE than sparsely ionizing radiation types, as Fig. 4 shows with carbon ions; a major comparison to look at would be the curve of the 11.4-MeV/u carbon ion having an RBE of approximately 3.5 for the CGL1 cells using equation 2. It should also be noted that the curve for the 11.4-MeV/u carbon ions has a nearly linear shape.

In general, heavy ions are known to have a higher RBE than sparsely ionizing radiation and have a tendency to deposit high amounts of energy that is heavily clustered along the particle path length, while photons tend to deposit their energy infrequently and in a disconnected manner. While the exact effect is part of what the presented model aims to investigate, the general pattern accepted in the field is that a higher linear energy transfer (LET) leads to a higher RBE, as shown in Fig. 5. The linear energy transfer (LET) of a particle is a measure of the energy transferred by a particle over a unit distance and is analytically determined by the stopping power of the medium it traverses[8]. In this manner the LET is similar to dose; it provides a total energy deposition with some spatial information. While dose provides the net energy deposited to a unit volume, the LET provides the energy released along a unit length.

2.6 Track Structure

The track structure takes the analysis another step further from the basic assumption that higher LET is synonymous with higher RBE: the track structure preserves the energy deposition events to precise points in space. The purpose of using the track structure is to stochastically model the energy deposited by ionizing radiation to cellular structures, even

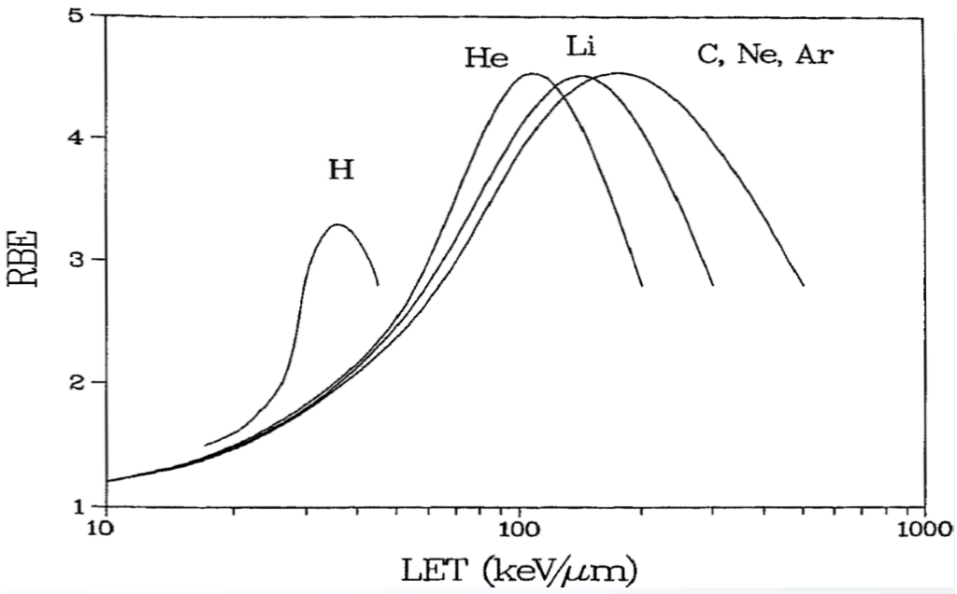
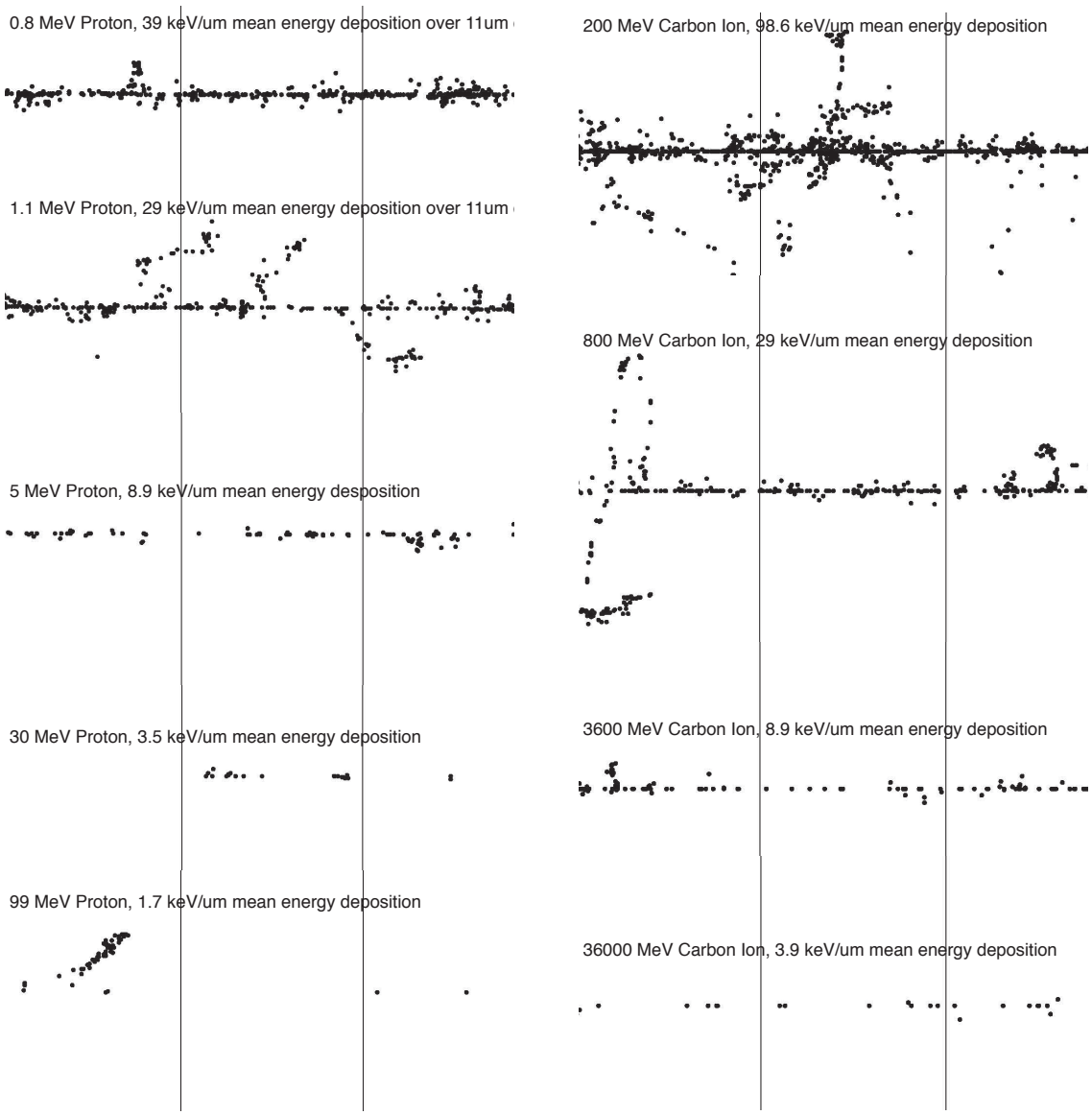


Figure 5: LET vs RBE of various particles based on approximations[46]

at the DNA level.

While particles behave stochastically in their interactions, specific particle types at specific energies can embody similar behaviors. The mean free path (MFP) is the mean distance between sites of interaction for a particle. For charged particles it is the distance between interaction sites and can be hard, soft, radiative, or nuclear collisions. The mean free path for particles of the same energy and type are identical, and while the Monte Carlo method will generate different paths for the two particles, their track structure will show similar features.

Generally speaking, for identical heavy particle types, the hard collisions of higher energy particles will more frequently generate higher energy delta rays than for identical particles of lower energy. In addition to higher energy particles having more energetic delta rays (i.e., the knocked out secondary electrons from ionization interactions) that tend to deposit their energy farther away from the path of the heavy particle, their mean free path is longer than their lower energy counterparts. This makes heavy particles to have a more disconnected track structure at higher energies than lower energies[46]. Figure 6 demonstrates this property, and also how particles decrease in LET as their energy increases. As the LET increases and the particle energy decreases, the mean free path between interactions



(a) Protons (b) Carbons

Figure 6: Particle track energy deposition events, 100nm grid spacing

decreases, leading to a particle track with smaller distances between energy deposition sites.

The general perception of LET's relationship to RBE is that a higher LET corresponds to a higher RBE. However, if a certain threshold exists wherein a particle enters a cell nucleus with high enough LET and sufficient range to guarantee cell killing, a higher LET will not correspond to a higher RBE. Any excess LET from such a threshold will decrease its RBE; it would deposit more energy with an identical radiological consequence.

2.7 Monte Carlo Methods

The Monte Carlo method, named after the casino and dating back to the Manhattan Project, is noted for being the gold standard for simulations of radiation interactions[41][18]. Its general usage covers a large breadth of statistical applications, and in the most simple terms it is a stochastic approach to solving complex mathematical problems. The Monte Carlo method makes use of known or assumed probability distribution functions and pseudo-random number generation to model the behavior of stochastic processes, and is advantageous for complex problems. Rather than dealing with a mathematical problem as an enormous integro-differential equation, a statistical approach is taken wherein a very large number of random samples of single events with a known probability are run. A large enough accumulated history of such events subsequently yields a result that effectively models the complicated integro-differential equation[41].

The application of the Monte Carlo method requires a very large sample history and a powerful pseudo-random number generator to work; essentially, a large amount of computing power is required but the mathematical complexity of the problem is made very simple through this approach. As computers have grown more powerful, the practical applications of Monte Carlo have increased dramatically. In the case of radiation interactions with the human body, the Monte Carlo method has shown itself to be highly effective in estimating the distribution of dose on a macroscopic scale[18]. In radiobiology and nanodosimetry, there are many publications and models that have been proposed that extensively make use of the Monte Carlo method at some stage[9][26][42]. The new model being presented is unique in that it makes use of Monte Carlo methods for every step of the procedure,

from the generation of particle tracks all the way through the placement and seeding of the biological structures in the cell nucleus.

2.7.1 Geant4 and Geant4-DNA

Geant4 is a freely available Monte Carlo particle simulation toolkit developed and used by CERN and includes a large number of physics models that can be appropriately selected depending on the end users' needs[3]. Geant4 stands for geometry and tracking, 4th generation, is C++ based and can reference a multitude of outside software implementations for visualization and data processing. Within Geant4, several examples are included that present either an educational tutorial on the operation of Geant4, or a more extensive simulation of a real-world scenario. These vary in complexity from a basic calorimetry detector to a full simulation of a heavy ion medical therapy unit. Extremely complex examples are also available by contacting more experienced end users. Each example problem consists of a primary .cc file, several included .h and source .c files, and macro and library files: most of these will call to other files available throughout Geant4's libraries or provide data for running the simulations required. Geant4 is a Monte Carlo program, and as such it calls to the respective files in accordance with a fairly standard coding flow. The order by which different files used in a simulation are called can be traced to determine their place within the overall example.

Geant4-DNA is an extension of Geant4 that handles discrete physics processes at the nanoscale level required for DNA studies[33][32]. Geant4-DNA allows for simulation of particles at very low energies and can track particles' step-by-step interactions in liquid water, and this added nano-scale level of detail enables modeling particle interactions at the scale level of DNA as well as the diffusion interactions that are prevalent in radiochemistry[22]. The selection of Geant4-DNA for this study was due to its power, availability, accuracy, and customizability[21]. The presented model requires a database of particle tracks that were obtained from Geant4 in the form of spatial coordinates and energy deposition for individual ionization and excitation events in liquid water approximating the cell nucleus.

Several publications compare Geant4-DNA to other available Monte Carlo codes and highlight the statistical similarities in dose deposition and radial dose distribution, and as such for the current model the Geant4-DNA physics accuracy is assumed to hold true[14][7][34]. Future investigations will necessitate a more comprehensive validation of the Geant4-DNA package.

CHAPTER III

METHODS

3.1 Geant4 Microdosimetry Example

The microdosimetry example was the foundation for generating the track structure database. In order to obtain results pertinent to our study, several parameters of the microdosimetry example were modified. Among the most important to verify within Geant4 is randomization of the start position of the random number generator (RNG). Geant4 contains several powerful RNG's for the user to choose from, but in the example problems, including Microdosimetry, the RNG's are often set to a fixed starting seed. Every time the software is restarted the starting RNG seed is reset to the original value, which is excellent for educational purposes, but requires for all simulations to be run during a single session to get a truly randomized output. In order to get a truly random sampling from the example throughout multiple sessions, some form of enforced randomization is needed, and in our case we chose to link the start seed of the RNG to the system time. The location for doing so is in the .cc file, with examples in the appendices.

The original microdosimetry example is primarily a tutorial, showing how to set up the physics lists within Geant4-DNA alongside the Geant4 standard physics lists within different volumes of a simulation. The example problem tracks each individual discrete step in the Geant4-DNA physics processes and its output file structure is very similar to what we were interested in, so very little modification was required to output the .csv files for our particle database. The detector volume parameters were modified from the original 1mm x 1mm x 1mm cubic detector volume to a basic cylindrical volume with a 15μm diameter and a 30μm length. The particles were fired using the General Particle Source (GPS) in Geant4 from the center of the detector volume with an initial direction along the axis toward one end of the cylinder.

Instead of using the hybrid Geant4/Geant4-DNA physics processes, only the Geant4-DNA physics lists were used. The discrete history was filtered to only record processes of high enough energy for electron excitation, as events below 1 eV dominated the original recorded data. A lower threshold of 7eV was used and vibrational excitation events were excluded. Data were recorded using the AIDA analysis manager, with file outputs also referencing the C standard library to get .csv file outputs. For each track, two files were output: the positively charged energy deposition data were recorded in one file and the negatively charged energy deposition data were recorded in another. This was intentional to preserve the path of the primary positively charged particle for the cell model intersection for the majority of simulation data. For all energy deposition events, the x, y, and z cartesian coordinates were recorded, with a 0,0,0 start position, along with the energy deposition and type of event. For the primary positively charged particle, the particle's instantaneous energy was also recorded to verify the energy deposition to compare the final energy with the software-recorded linear energy transfer.

3.2 Track Structure Database

The database was constructed by acquiring 30 tracks for each particle type and energy under investigation. Alpha particles at energies of 2.5-MeV, 3.5-MeV, 4.8-MeV, 7-MeV, 10-MeV, 15-MeV, and 20-MeV were recorded, as were carbon-12 ions at 50-MeV, 125-MeV, 200-MeV, 500-MeV, and 800-MeV. Protons at initial energies of 0.8-MeV, 1.1-MeV, 2-MeV, and 3-MeV were also used. Each track was numbered sequentially, as primary.#.csv and secondary.#.csv, where # represents its track number from 1 through 30. The track numbers selected in the model were then called out by a uniform RNG, scaled from its 0∼1 to a 1∼30.9999 value instead. The reason for separating the particle data from the primary positively charged particle and secondary negatively charged particles was to simplify the later data analysis for aligning the cell nucleus model with the track based on the primary particle only. The LET and number of tracks per Gy of radiation are shown in table 1.

Table 1: LET and number of tracks per 1 Gy for simulated particle tracks[a]

	LET (keV/μm)	Tracks per Gy		LET (keV/μm)	Tracks per Gy
Alpha 20-MeV	29	13.63	Carbon 800-MeV	29	13.63
Alpha 15-MeV	37	10.67	Carbon 500-MeV	46	8.58
Alpha 10-MeV	50	7.91	Carbon 275-MeV	73	5.45
Alpha 7-MeV	71	5.57	Carbon 200-MeV	100	3.95
Alpha 4.8-MeV	100	3.95	Carbon 125-MeV	150	2.63
Alpha 3.5-MeV	131	3.02	Carbon 50-MeV	299	1.32
Alpha 2.5-MeV	187	2.11			
Proton 3-MeV	12.7	31.14	Electron 1-MeV	0.2	1977
Proton 2-MeV	18.0	21.97			
Proton 1.1-MeV	29	13.64			
Proton 0.8-MeV	39	10.14			

[a]LET is reported as track-averaged over the 11μm diameter of the cell nucleus

Table 2: Modeled Cell Nucleus Structures and Characteristic Dimensions

Structure	Characteristics
Cell Nucleus	Sphere (11 μm-dia)
Chromatin Domain	Sphere (400 nm-dia)
Chromatin Fiber	Cylinder (30 nm-dia, 150 nm-length)
Ave no. of CDs in a nucleus	6000
Ave no. of CFs in a CD	56

3.3 Cell Nucleus Model

The first step in establishing the cell nucleus model was to conceptually develop how to best approach the simulation of damage by incident radiation while keeping a fully stochastic Monte Carlo approach viable. Once the model parameters were determined, scripts were written for input into Geant4; this was done primarily to make use of the Geant4 random number generation engines. A single track approach was used wherein a single primary particle and its associated delta rays were observed for each individual track. The single track effect refers to a single particle as having the ability to cause lethal cell damage on its own. As such, the interactions of DNA damage sites from one particle track were assumed to have no interactions with the energy depositions of other particles. While this approach is found lacking in the example of sparsely ionizing radiation, the results section will demonstrate that for high LET radiation, the quantity of particle tracks required to deposit a relatively high dose is fairly low and as such the effect within the energy range of the presented research is small, albeit non-negligible.

The presented cell model approximates the aforementioned chromatin domains (CDs) and chromatin fibers (CFs) within the cell nucleus (see section 2.2). The single track effect was noted for its simplicity of programming, as it allows for the generation of the critical targets within the cell nucleus by way of a mean free path (MFP) approach in establishing the probability of the particle track crossing intersecting the CDs and CFs while ignoring the majority of structures. This is performed by using their geometric cross sections to

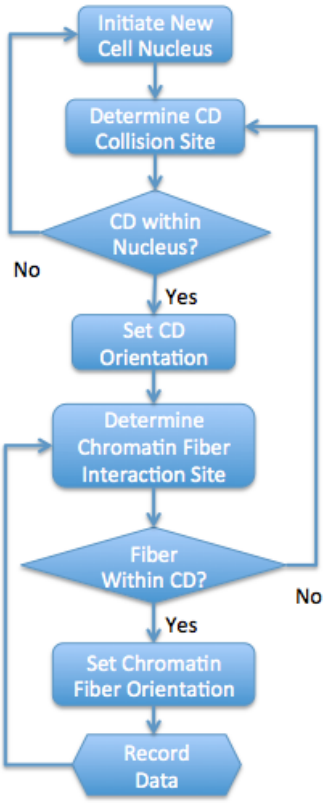


Figure 7: Flowchart for cell model geometry seeding

calculate the MFP along the primary particle track to determine the locations of CDs, then subsequently using a mean free path within the CDs to determine the placement of fibers. In this way, CDs and fibers are placed stochastically while simplifying the cell model, only approximating the structures within the nucleus that are the most likely to undergo DNA lesions.

The particle track and the nucleus model were intersected using a combination of in house and publicly available Matlab codes. Detector volumes were generated within Matlab to approximate the cylindrical fibers, and the particle track structure data that were contained within these volumes were recorded to output files. This data show the spatial deposition of energy to CFs, and further processing was used to determine the likelihood of cell lesions developing.

3.4 Monte Carlo procedure for implementing particle tracks intersecting a cell nucleus

A cell nucleus was modeled as a 11μm diameter sphere interspersed with 6000 chromatin domains (CD's). Each CD has a 400nm diameter and represents 1 million base pairs (MBp) of DNA, and the resulting model approximates a typical human cell nucleus. In the single track approach used, the particle was sent through the nucleus across the greatest chord length possible through the center of the nucleus, from one point on the sphere in the direction of the opposing point. The probability of crossing a chromatin domain was determined by calculating the MFP of the CD based on its cross section. A 400nm diameter spherical CD occupies 0.03351 μm^3, while 6000 CDs occupy 201.1 μm^3 of a 11 μm diameter spherical nucleus's total volume of 696.9 μm^3, or 28.85% of the nuclear volume of a cell.

The Monte Carlo cell nucleus model development followed a basic coding sequence along the projected path of the primary particle and is shown in Fig. 7. The most simplistic flow of the program followed the particle track from an x-position of 0nm until the end of the cell nucleus at a position of 11,000nm. Once the particle passed a position of 11,000nm, it was assumed to have fully exit the cell nucleus and was no longer tracked; as such, only structures within the cell nucleus were modeled.

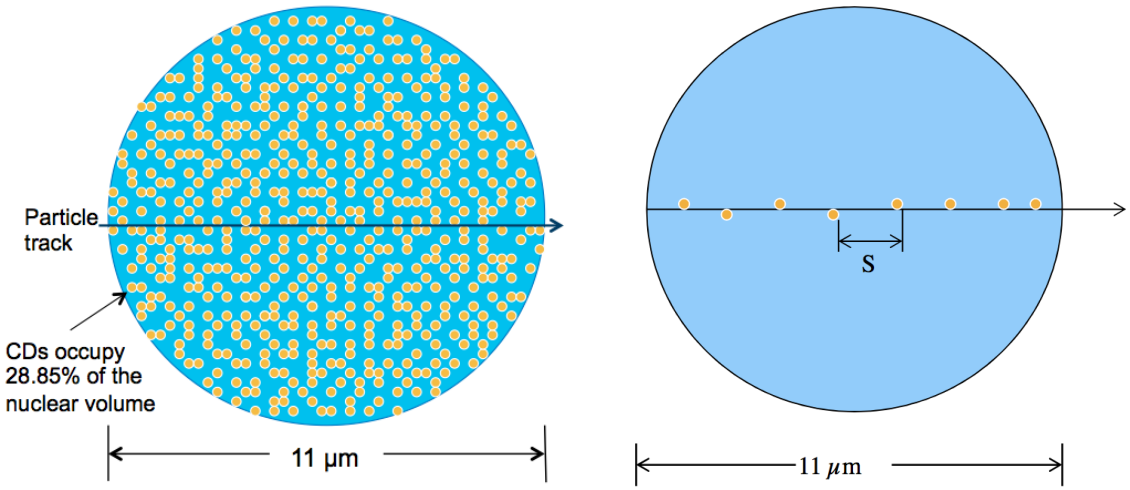
The first action the program executes is to allocate a cell number and particle track number. The cell number is allocated sequentially from 1 until a sufficiently high number of cells are accumulated. The cell is then allocated a track number to pull from the track database in section 3.2. The calculation for this is simple:

$$track\ number = \mathcal{R} * 29.99999 + 1 \tag{3}$$

where R is a uniform random number from $0 \sim 1$. The rationale behind the programming is to get a uniform probability of getting an integer output from $1 \sim 30$.

The method for handling the cross section is based on the 6000 CD's within the cell nucleus and the cross section for interaction per unit volume of the nucleus:

$$\sigma = \pi(0.2\mu m)^2 = 0.1257\mu m^2 bead^{-1} \tag{4}$$



(a) Cell nucleus with CDs

(b) Cell nucleus MFP approach

Figure 8: Cell nucleus model and the single track approach applied

While the density of beads within the cell nucleus is calculated by:

$$N = \frac{6000 beads}{\frac{4}{3}\pi(5.5\mu m)^3} = \frac{6000 beads}{696.9(\mu m)^3} = 8.609 \frac{beads}{\mu m^3} \ . \tag{5}$$

The overall interaction probability is then determined to be:

$$\Sigma = N\sigma = 1.082 \mu m^{-1} \tag{6}$$

The mean free path for interaction with a CD is then found to be:

$$\lambda = \frac{1}{\Sigma} = \frac{\mu m}{1.082} = 0.924 \mu m = 924 nm \tag{7}$$

To determine the pathlength to the chromatin domain intersection using the MFP approach is then performed by:

$$S = \lambda \ln(\frac{1}{\mathcal{R}}) = 924 nm \ln(\frac{1}{\mathcal{R}}) \tag{8}$$

Following the placement of a CD entry point by its mean free path, the direction vector was then calculated for the CD and the CF intersections within the CD were then calculated. The pathlength for the next CD entry point, calculated from the prior CD entry point, was then calculated by means of the prior equation. Each CD is allocated a sequential number beginning with 1; if a CD is traversed without a single CF hit, it is ignored an the next sequential CD is determined until the cell nucleus periphery is reached.

3.4.1 Chromatin Domain Direction Vector

Once the mean free path was determined for the intersection of a particle track with the cell model, each CD was assigned a direction vector to the center of the CD. It was assumed that the intersection of the particle track determined by the MFP would be at the boundary of the CD, and as such the CD's center would be 200nm from the intersection point. Since the intersection point was calculated to be at the proximal surface of the CD relative to the particle source origin, the direction vector of the CD center was forward directed. All forward directions were assumed to be equally likely and the direction vector was also isotropic and called by a random number generator for both direction angles.

To isotropically obtain the direction for the CD in the forward direction, the polar angle φ and the azimuthal angle ϑ must be determined stochastically. The polar angle's probability distribution function is given by:

$$f(\phi)d\phi = \frac{dA}{A} = \frac{r^2 d\Omega}{A} = \frac{2\pi r^2 \sin\phi d\phi}{2\pi r^2} = \sin\phi d\phi \tag{9}$$

where dΩ is the differential solid angle:

$$d\Omega = 2\pi \sin\phi d\phi \tag{10}$$

Integrating the probability distribution function of the polar angle φ yields:

$$F(\phi) = \int_0^{\phi} f(\phi)d\phi = \int_0^{\phi} \sin\phi d\phi = 1 - \cos\phi = \mathcal{R} \tag{11}$$

Where R is a uniform random number with value from 0 ∼1. This yields:

$$\phi = \arccos(1 - \mathcal{R}) = \arccos(\mathcal{R}) \tag{12}$$

The azimuthal angle ϑ has a probability distribution function of:

$$f(\theta)d\theta = \frac{d\theta}{2\pi} \tag{13}$$

Integrating the probability distribution function of the azimuthal angle yields:

$$F(\theta) = \int_0^{\theta} f(\theta)d\theta = \int_0^{\theta} \frac{d\theta}{2\pi} = \mathcal{R} \tag{14}$$

Simplifying this yields:

$$\theta = 2\pi\mathcal{R} \tag{15}$$

where R is a uniform random number with value from 0 to 1. From this information, the unit vector to the CD's center can be calculated in cartesian coordinates as:

$$\hat{\Omega} = u\hat{i} + v\hat{j} + w\hat{k} \tag{16}$$

where the unit vectors u, v, and w are calculated by:

$$u = \cos\phi \tag{17}$$

$$v = \cos\theta \sin\phi \tag{18}$$

$$w = \sin\theta \sin\phi \tag{19}$$

The center of the CFs were then calculated by multiplying the unit vectors by 200nm, the radius of the chromatin domains:

$$x = r \cdot u = r\cos\phi \tag{20}$$

$$y = r \cdot v = r\cos\theta \sin\phi \tag{21}$$

$$z = r \cdot w = r\sin\theta \sin\phi \tag{22}$$

where r is the radius of the CD.

The placement of CD's was a very important element of the proposed model, but only insofar as the primary particle intersected with a CF. While every CD was placed within the calculation, its position was recorded in the context of the CFs within it; if no CFs were crossed by the primary particle within the chromatin domain, it was not recorded. The data recorded for the CD included the sequential number of the CD, its x, y, and z coordinates, and the chord length of the particle track crossing the CD. Once the CD center was determined, a local coordinate system was implemented with the origin set to the center of the CD to assist in the placement of CFs.

3.4.2 Cell Model Parameter Basic Setup

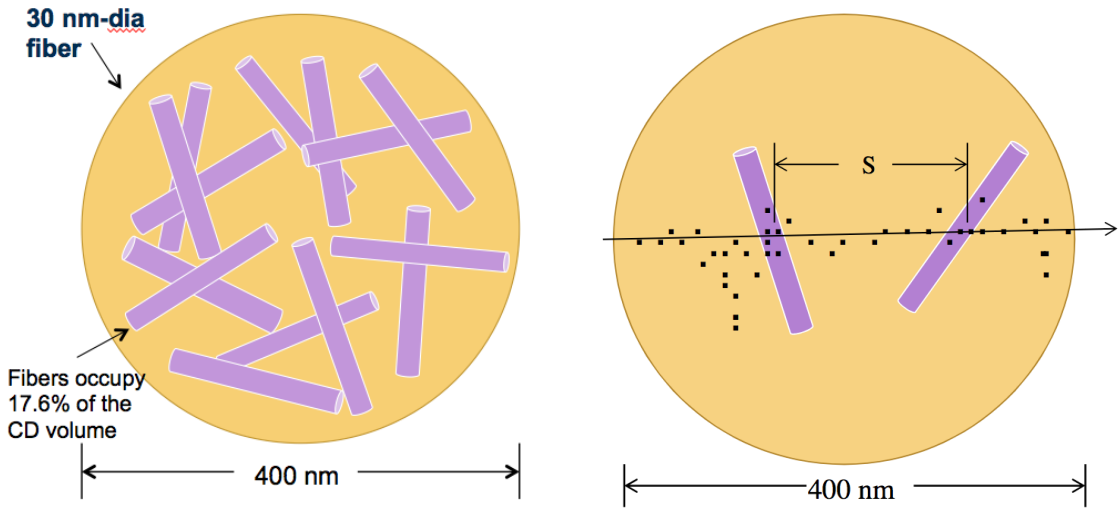
From an x-position of 0nm, the path length to the first CD intersection was calculated based on its MFP. Once the CD was entered by the particle, the direction vector of the CD was calculated to determine the boundaries of the CD as well as the chord length along the primary particle track. The path length to a CF was determined based on its MFP within the CD. If the MFP to the CF was determined to be greater than the chord length of the CD traversed by the primary particle, it was assumed that the primary particle missed any major targets and the CD and its fibers were ignored. The details in how this was performed, as well as the cross sections and other numerical calculations are described in sections 3.4 and 3.4.1.

If the MFP to the CF was less than the chord length, the direction vector to its central axis and the direction vector of its central axis was determined. The CF was assumed to extend to the boundaries of the CD in both axial directions, provided that its axial length was less than 150nm. If its axial length was greater than 150nm, its length was halved until it was less than 150nm[12][47]. Once a CF was placed, the path length from the CF initial intersection point to the next CF intersection point was determined from its MFP and the placement steps were repeated until the primary particle exit the CD. A thorough explanation of parameters used to determine the CF locations is included in section 3.4.3.

The path length from the CD's intersection point to the next CD intersection point was then calculated, and the same procedure followed in placing the CFs within the next CD. This was repeated until the primary particle was at an x-position of 10,600nm, 400nm from the end of the cell nucleus to prevent placement of a CD outside of the region of interest. The entire cell nucleus model seeding was repeated extensively for each particle type and energy of interest so that enough cell histories could be accumulated for better statistics.

3.4.3 Chromatin Fibers within Chromatin Domains

The CFs were assumed to reside entirely within the CDs in this simulation, and their positions, similar to the CDs, were determined using a MFP approach. Each fiber was modeled with a diameter of 30nm, with an overall length of up to 150nm but not exceeding



(a) Chromatin fibers within CDs

(b) Chromatin fiber MFP approach

Figure 9: Chromatin fibers within chromatin domains

the periphery of the CD. The length of 150nm was chosen to be similar to existing CF loop structures in Geant4-DNA, and represents approximately 18 kbp of DNA[47]. Unlike the modeled CD that simply has a central point determining the location and orientation of the CD, the fiber has a central axis location in addition to an axial orientation due to its cylindrical profile. The mean free path was calculated by the ratio of CFs occupying the chromatin domains versus the chromatin free portions.

As a whole, approximately 5.07% of the cell nucleus is occupied by CFs. Given that the CD's occupy 28.85% of the cell nucleus, the volume of each CD in our model occupied by CFs would be 17.6%. The mean-free-path method is then applied to determine the position of 30nm diameter CFs within the CD, and the position of the central axis of the CFs is stochastically determined by an isotropically forward-directed calculation with a distance of 15nm. A MFP of 147nm used for the CF.

If the pathlength for the CF is smaller than the chord length traversed by the primary particle through the CD, the CF data are then fully calculated. If the pathlength to the CF is larger than the chord length of the CD, the particle is assumed to traverse the CD without interacting with any CFs. The polar and azimuthal angles, as well as the unit vector for calculating the direction to the center of the CF, are calculated using equations

(9) - (22), with a radius of 15nm for the CF used instead of the 200nm for the CD's. The direction vector of the CF axis was calculated with equations (9) - (19). In keeping with the published findings of the CF rosette structure, the CF length was halved until it was less than 150nm in total length[47].

The central axis point of the CF was also converted to the local coordinate system of the CD. This point was used to determine the intersection of the CF with the CD boundary; the distance from each end of the CF to the CD center would be equal to the radius of the CD.

$$r = \sqrt{(x - x')^2 + (y - y')^2 + (z - z')^2} \tag{23}$$

$$x = x_1 + u \cdot L \tag{24}$$

$$y = y_1 + v \cdot L \tag{25}$$

$$z = z_1 + w \cdot L \tag{26}$$

$$r^2 = x^2 - 2x \cdot x' + x'^2 + y^2 - 2y \cdot y' + y'^2 + z^2 - 2z \cdot z' + z'^2 \tag{27}$$

With the center of the CD was set to a local origin, the equation simplifies to:

$$r^2 = x^2 + y^2 + z^2 =$$

$$x_1^2 + 2u \cdot L + u^2 \cdot L^2 + y_1^2 + 2v \cdot L + v^2 \cdot L^2 + z_1^2 + 2w \cdot L + w^2 \cdot L^2 \tag{28}$$

where x_1, y_1, and z_1 are in the local coordinate system of the CD.

$$r^2 = x_1^2 + 2u \cdot L + y_1^2 + 2v \cdot L + z_1^2 + 2w \cdot L + L^2 \tag{29}$$

$$L^2 + 2(u + v + w) \cdot L + (x'^2 + y'^2 + z'^2 - r^2) \tag{30}$$

The quadratic formula simplifies this to:

$$L = \frac{-2(u + v + w) \pm \sqrt{(2(u + v + w))^2 - 4 \cdot 1 \cdot (x_1^2 + y_1^2 + z_1^2 - r^2)}}{2 \cdot 1} \tag{31}$$

Since all values except for the length L from the CF axis point location to the boundary of the CD is known and the positive value of length is desirable, the length values chosen are:

$$L_1 = \frac{-2(u+v+w) + \sqrt{(2(u+v+w))^2 - 4(x_1^2 + y_1^2 + z_1^2 - r^2)}}{2}$$
$$L_2 = \frac{2(u+v+w) + \sqrt{(-2(u+v+w))^2 - 4(x_1^2 + y_1^2 + z_1^2 - r^2)}}{2} \quad (32)$$

Where L_1 is the length in the positive unit direction, L_2 is the length in the negative unit direction, and r is the 200nm radius of the CD. The global coordinates of both CF ends were then calculated using equations 24 - 26. The next CF path length is calculated the CF placement process is recorded until the path length yields a value beyond the CD. The process is then repeated for each CD to seed CFs, which are numbered sequentially.

3.4.4 Cell Model File Output

The placement of a CF signals Geant4 to record the data into a .csv output file. The data recorded for the CF include: the cell number, particle track number, CD number, CD cartesian coordinates, CD chord length traversed by particle track, CD intersection x-coordinate, CF intersection x-coordinate, CF end point 1 cartesian coordinates, and CF end point 2 cartesian coordinates. In total, the output file has 14 columns. Considerably more data are accumulated in the simulation, but only these values are relevant for the next programs used for verifying damage sites in the model. If a CF is not hit within a CD, the CD data are not recorded. In the rare cases when a primary particle track crosses a cell nucleus without hitting any CFs, no data from that cell are recorded and the next cell is modeled; in such cases the cell number is still increased to identify a total miss. A typical cell model output file includes enough data for $500 \sim 1000$ cells.

3.5 Matlab Data Analysis

The distribution of radiation damage was analyzed in Matlab by intersecting the particle track structures with the cell models. Since we desired to call to a track database for our cell model, Geant4 was deemed to be impractical for running the entirety of the simulations. In addition, the dynamic seeding of CF detector volumes within Geant4 along a primary

particle track as it traverses a cell nucleus is not practical to set up. Matlab offered ease and flexibility for reading input files in a simple and intuitive manner. Several publicly available Matlab codes were used to convert the cell nucleus model data into a volumetric detector setup for use with the particle track structure database that was accumulated.

3.5.1 Particle Track Interactions with the Chromatin Fibers

The first step involved in intersecting the particle track with the cell model data is to input the data for the cell models used as described in section 3.4.4. The quantity of cells to be run is determined, and a FOR loop is used to partition all of the CFs for a single cell to be analyzed. Since the particle track number from the database as described in sections 3.2 and 3.4.2 will be the same for all CFs, this value is taken for only one of the CFs in the partition. The primary particle data and secondary electron data are then read from the database based on the number allocated by the cell model called.

3.5.1.1 Chromatin Fiber Position Offset

Heavy charged particles typically move in a straight path, but even a 0.16^{o} deflection over the 11μm length of a cell nucleus can alter the particle trajectory by over 30nm at the exit point. This would be sufficient to cause a primary particle to entirely miss CFs modeled to intersect its trajectory. Since our original CF placements are assumed to be directly along the primary particle path and some of the heavy charged particles in the track structure database have slight axial deflections, it was important to offset the CF positions in the cell model by the particle's axial deflection. This was performed at the CF intersection point from the cell model by adding the y- and z-coordinate of the primary particle position nearest the intersection point to the y- and z- coordinate values of the CF end points, respectively. This process was repeated for all of the CFs for the cell being modeled. Following the offset of the CFs, the primary and secondary energy deposition data were merged to simplify the intersection algorithm.

Figure 10: Flowchart for Intersection Algorithm

3.5.1.2 Intersection Algorithms

The intersection algorithms required the placement of detector volume regions within Matlab based on the endpoint locations of the CFs, and the general process is demonstrated in Fig. 10. Once the detector volume regions were developed, the intersection of the spatial data of the particle track energy depositions within the detector volumes were processed. The method chosen for performing this was to seed a cylinder surface based on the cylinder2P Matlab code[48], translation of the cylinder surface code to a delauney triangulated surface boundary using the alphavol code[40], and then finally to run a volume intersection based on the inpolyhedron code[31]. The cylinder2P Matlab code took the initial CF end point positions and output a cylinder geometry comprised of external surface points. For

the purpose of developing the intersection algorithm, a constant diameter was used in its source code. A sampling of 10 points along the axis and 10 points along the circumference of the cylinder defined a total number of 100 points per cylinder. The output data of cylinder2P were a 100x3 matrix of the cartesian coordinates of the cylinder points.

The alphavol program was used to develop a Delauney triangulated surface boundary of the cylinder output from the cylinder2P code. Alphavol optimized the uniform point layout from the cylinder2P outputs to efficiently produce a triangulated surface boundary. In layman's terms, alphavol allowed for the inside of the DNA containing CF cylinders to be separated from the outside DNA absent chromatin domain regions. The output file of alphavol that was specifically used in the simulation was an Mx3 matrix comprised of the cylinder point identifiers that made up triangular boundary surfaces of the cylinder volume.

The inpolyhedron program ran the actual intersection algorithm for each CF in the cell nucleus model. Its inputs were the triangular boundary surface output identifiers of the alphavol program, the point coordinates of the cylinder2P program, and the Nx3 cartesian coordinates of all primary and secondary particle energy deposition events, where N is the total number of all primary and secondary events. The output of the inpolyhedron program is an Nx1 vector that corresponds to all primary and secondary energy deposition events, where the value is 0 if there is no intersection and the value is 1 if the events are physically located within the volume. The output of the inpolyhedron program is then processed to record CF hits, allocating the CF number and CD number to each primary and secondary energy deposition event that falls within the CFs. A filtering code is also implemented to prevent energy deposition events from being double-counted in the event that CFs physically intersect with each other.

Once all of the CFs in the cell nucleus in question have been processed, its data were output as a .csv file of a naming scheme *particlename.energy.cellnumber.csv*. After recording, the next cell was processed following the same procedure until all cells were processed and their data recorded. The file output from the hit intersection algorithm includes the cartesian coordinates of the energy deposition event, the total energy deposited in the interaction, the type of interaction, the CF number and the CD number, with the cell number,

particle type and initial particle energy preserved in the name of the file.

3.5.2 Clustering of Radiation Damage in the Chromatin Fibers

3.5.2.1 DBSCAN Cluster Processing

The hit intersection data was subsequently processed through a scanning code whose framework was based on DBSCAN[19]. DBSCAN, or Density Based Spatial Clustering of Applications with Noise, would take the hit intersections and output cluster information containing the number of interactions and energy deposited within each cluster. It efficiently analyzes data that bear similarities, and in the proposed cell nucleus model it is used to determine proximity between coordinates in 3D space, specifically the deposition of energy to CFs by particle track excitation and ionization events. DBSCAN's input as used in our model required a radial distance parameter, a minimum cluster size parameter, and a three dimensional spatial coordinate matrix of hits. The radial distance parameter sets the maximum distance from which events spaced apart will be included in a cluster, while the minimum cluster size parameter sets the minimum number of events clustered together to register the cluster. DBSCAN is applied in the cell nucleus model for identifying the energy depositions to CFs that are the most likely to lead to DNA double strand breaks (DSBs), which are most relevant to the observable biological consequences.

Setting the parameters represents the first step of calibration in our model; variations in the values of the parameters can greatly alter the outputs and results of post-processing. Since the target of the new radiobiological model presented is to correlate a Monte Carlo method to experimentally validated data, the parameters were varied significantly in multiple ways in order to gain insights. Eventually it was settled on to use a very small cluster size parameter of 2 in the DBSCAN algorithm and to apply the minimum cluster size filter in post-processing. This prevented the need to re-run DBSCAN multiple times. The maximum distance was varied from as small as 0.1nm to as large as 3nm. Eventually it was hypothesized that a distance of 2nm agreed with the literature best, since this corresponds to the most likely scenario for a double strand break. The output from the algorithm was an output matrix listing all individual clusters, with the output data being: cell number,

CD number, CF number, cluster number as some CFs had multiple clusters, quantity of ionizations in the cluster, quantity of excitations in the cluster, quantity of total events in a cluster, and energy deposited in the cluster.

3.5.3 Clustered Damage Post-Processing Steps

At this point sufficient data were collected to be analyzed in post-processing. The benefit provided by this method is that the accumulated data took into account more recent cell nucleus chromatin structure into effect while also accounting for nanodosimetric data from the particle track structure. Comparisons that were undertaken include the number of DSBs that were generated per Gy of radiation, as well as the severity of damage based on the cluster data.

3.5.4 Electron Calibration Algorithm

One final calculation that was applied was for the purpose of calibrating the hit size energy that would correspond to a DSB. 0.9999-MeV Electrons were used to determine the hit size required for a DSB under the generally accepted premise that 1 Gy of sparsely ionizing radiation leads to approximately 40 observable DSB's in a cell nucleus[28]. There was an expectation that high-LET particles would introduce some variability in DSB's per Gy at different energy ranges, and as such it was sparsely ionizing radiation was used to calibrate this value. Geant4-DNA has the capability of running electrons below 1-MeV, but no higher. Since higher energy electrons typically follow a straighter path than lower energy electrons, the highest available energy was used.

3000 electron tracks and 3000 cell models were generated for 3 data sets. A random number generator pairing output file was developed to match up different electron tracks with the different cell models. This was performed by using a the Geant4 uniform random number generator in equation 3, multiplied by 2999.99999 instead of 29.99999, for a two column matrix, where all values were calculated the same for a random matching between electron track number in the first column and cell model number in the second column. The matrix was then processed in Matlab to remove identical rows to insure no cells and electron tracks would be repeated.

A total of 1977 electron tracks at approximately 0.2 keV/μm were calculated to be equivalent to 1 Gy of radiation by Geant4-DNA. The electron radiation was processed in an identical method as for the hadron tracks, with the cell model being offset by the y- and z-coordinate positions of the primary electron path. The key to this approach in Geant4 is that the electron output file moves linearly from the initialization point of the electron until the exit point of the electron from the geometry of the cell, and then processes the delta rays from the exit point of the cell in the direction of the initialization point of the cell; as such, the primary particle data are preserved, even though electron interactions are typically indistinguishable from each other. In this manner, all CFs in the cell model were aligned with the primary particle tracks.

One of the drawbacks for this approach occurs when an electron undergoes a severe hard collision or loses energy via bremsstrahlung, altering the electron path considerably. In these cases, the cell model beyond said interaction would effectively be ignored. With the single-track approach this is unavoidable; however, the relative effect on the electron simulation, while not negligible, is not considered to be so high as to compromise the calibration model by more than 5 to 10 % . Since the major emphasis on the model is for heavier ion interactions, this was deemed to be acceptable for the current phase of research.

CHAPTER IV

RESULTS AND DISCUSSION

4.1 Energy Deposition to the Whole Chromatin Fiber

Several approaches were taken in processing the acquired data to correlate the CF hits to cell death. In the most simplistic comparisons, the total hit size to entire CFs irrespective of clustering was calculated and displayed on a histogram to show a frequency comparison of hits. This showed the expectation values of energy deposition to CFs, and analysis of the various histograms for different particle types and energies yielded insightful comparisons.

Figure 11 is a histogram of the total energy deposited to individual CFs per Gy of radiation deposited to the cell nucleus by various particles at very high LET. The shape of the profile is nearly poisson distributed and demonstrates the general tendency of energy deposition. Owing to the stochastic nature of the orientation of the CF and the intersection path through the fiber by the primary particle, there is a fairly large spread of energies observed, including in the region of very high energy depositions.

The lower energy, higher LET alpha particles can clearly be seen to deposit considerably more energy on average than higher energy, lower LET alpha particles, which is expected due to the higher connectivity of particle tracks. While the total energy deposited to a CF by a particle track has correlations to the expected RBE escalation seen in Fig. 5, it is not analogous as it does not equate to actual double stranded DNA breaks directly. Regardless, the general property observed does back up some experimental observations, and the clustered data to be presented will bring this approach into better agreement.

Figure 11 shows a comparison of the hit sizes to CFs for alpha particles and carbon ions at high-LET. From left-to-right, the curves are the 200-MeV carbon ion at 100keV/μm energy deposition, the 5.25-MeV alpha particle at 87keV/μm, the 2.5-MeV alpha particle at 187 keV/μm, and then the 50-MeV carbon ion at 300keVμm. The key property of the plot in Fig. 11 is that the entire curve tends to shift to higher hit sizes as the LET of the particle

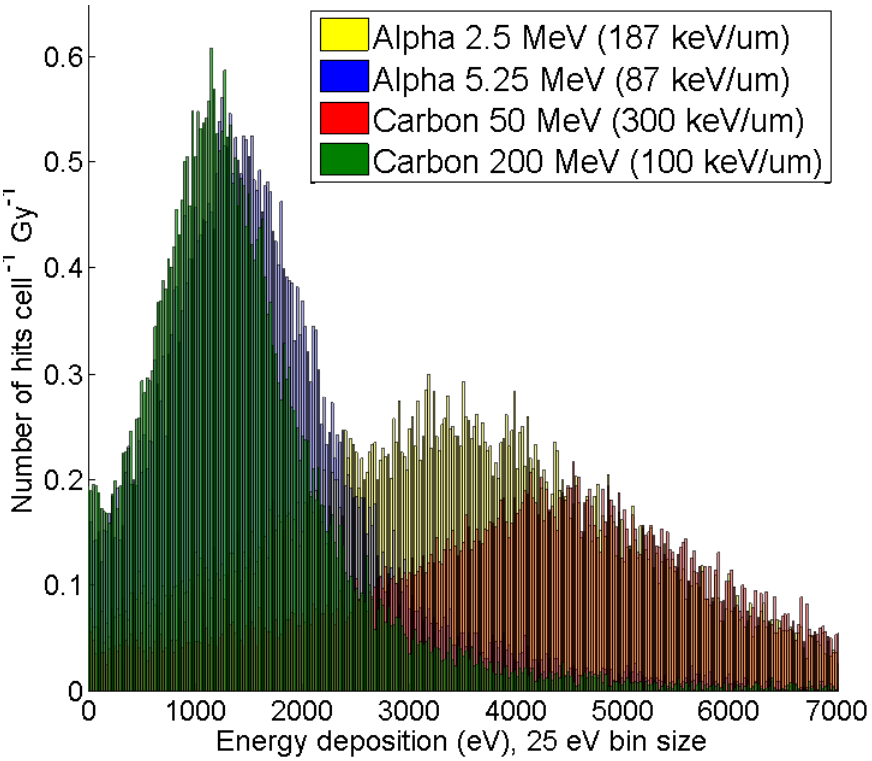


Figure 11: Hits per Fiber for High-LET Particles

increases, which is a basic take on expected observations. The most interesting observation here is in the 5.25-MeV alpha particle compared to the 200-MeV carbon ion, as conventional approaches would expect the higher LET carbon ion to have a larger average hit size than the lower LET alpha particle, but this is not observed; rather, the alpha particle has a notably larger mean hit size than the carbon ion. There is a tendency to equate LET to RBE directly, but the difference in hit sizes show that the full picture is not so deterministic. It is generally understood that lighter hadrons have a smaller MFP at the same LET as heavier hadrons, and this behavior is seen even between alpha particles of lower LET than carbon ions in Fig. 11. The increased average hit size of alpha particles relative to carbon ions in the CFs is on account of the increased frequency of primary particle events as well as the less energetic delta rays from ionization events depositing energy in the CF.

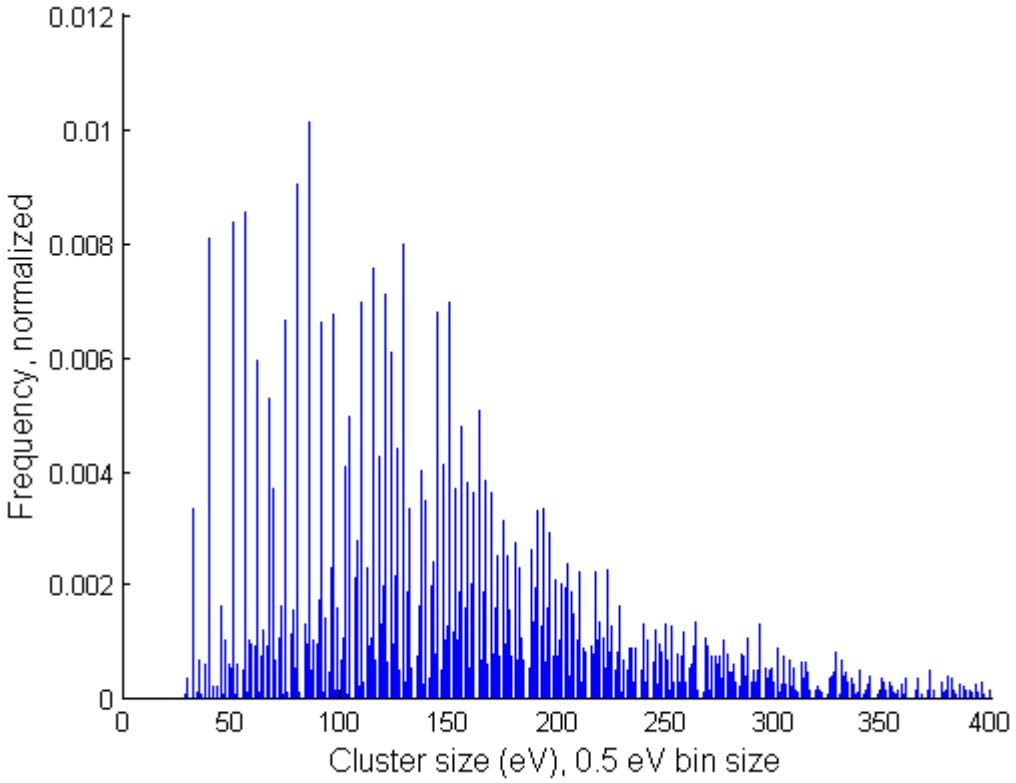


Figure 12: Cluster size distribution for 800-MeV carbon ions

4.2 Clustering of Energy Deposition Events

The total energy deposited in the CF is analogous to the total dose of radiation given to the DNA-containing structures, and lacks the main spatial data that are more relevant to assess the probability of damage leading directly to double strand breaks. The DBSCAN clustering algorithm provided considerably more insights into the probability of a DSB to be produced in a CF. A small histogram bin size of 0.5 eV demonstrates the discretized nature of energy deposition events in liquid water, as seen in Figure 12. This property decreases considerably for larger histogram bin sizes, but is present for the majority of energy depositions likely to account for DSB's.

Once DBSCAN was applied to the CF energy deposition data, additional complexities were added to the model. For sparsely ionizing radiation tracks, many of the energy deposition events were filtered out by the minimum radial distance parameter as they were too

far from other energy depositions to be registered as a cluster, and some CFs receiving hits did not meet the criteria for registering even a single cluster. For more densely ionizing radiation tracks, the opposite happened; many of the individual CFs registered multiple clusters from a single track. However, the most densely ionizing radiation tracks had a tendency of yielding a smaller number of clusters per fiber, most likely due to the distance parameter being large enough that the interactions in the primary particle track were close enough to link clusters together.

4.3 Cluster Count Simplification and Rationale

Using the DSB yield from a direct summation of clusters for RBE evaluation proved to vary significantly to the published literature. Initially the individual clusters were counted as independent DSB's, but this heavily skewed the data based on the DBSCAN setup parameters. In order to get the 100 keV/μm alpha particles having the highest RBE, the distance parameter and cluster size requirements had to be set to unintuitive values. As in any model, values could be substituted so as to match the literature, but this approach is highly suspect and thus rejected. The distance and cluster parameters required to match the literature were determined to be 0.08nm and a cluster size of 8, which did not make sense with respect to the size of the DNA double helix. To keep consistency, the distance parameter was fixed at 2nm, the diameter of the DNA double helix. The minimum cluster size for the DBSCAN input parameter was set to 2, so that all clusters greater than 2 events would be recorded: this allowed for a minimum cluster size to be determined by the user while further processing the data.

A very important observation was utilized to restrict the counting of clusters for DSBs. Since the maximum optical resolution attainable in typical fluorescence imaging is much larger than the 2nm diameter DNA double helix[38][17]. Hence, the counting of DNA damage repair mechanisms should only be visible around a CF being repaired, and not for individual DSB's in this regard. In a cell irradiation experiment wherein a CF has several highly clustered energy depositions, distinct from one another but in close proximity, fluorescence-labelled repair enzymes recruited to repair DSBs would appear as a single

lesion, and not multiple distinct lesions [29]. As such, the counted DSB's from such a cell study should represent repair to a chromatin fiber or domain in a general region within the spatial resolution of the lesions. More modern imaging techniques are improving on the attainable resolution but still fall short of distinguishing independent lesions in proximity on a 30nm chromatin fiber[38].

In addition to the visual observation of DNA damage repair mechanisms through FISH, the counting of DNA fragments by gel electrophoresis is another method by which DNA DSB's are counted that also should limit the counting of clusters[44][2]. Gel electrophoresis counting of DNA fragments typically target the mega-base-pair (mbp) size range or high kilo-base-pair range[2]. The specific clustered DNA damage observed in the proposed model using DBSCAN has a distance parameter set to 2nm, which is the diameter of the double helix and also a distance that accounts for approximately 7 base pairs along the length of the DNA double helix. As such, the clusters output by DBSCAN will have a very high likelihood of accounting for sub-kbp DNA fragments. Due to being overlooked, these sub-kbp fragments would not likely contribute to total counts of DSB's reported in the literature.

For modern bioinformatics discussions and gene sequencing, small DNA fragments are assayed as they may code for important data. For most DSB counting methods in cell irradiation studies, though, such small fragments are entirely neglected. As such it was considered that for the purposes of counting DSB's, it was a safe approach to count a single chromatin fiber as having a maximum of one DSB when multiple clusters are registered within it.

The restricted counting of clusters to individual s was executed by filtering through the clusters to select only the largest cluster to each . Figure 13 shows the result of this approach compared to the unfiltered data, and the shape of the filtered histogram more closely resembles the total energy deposited to a whole from Fig. 11. This simplified the counting of clusters to a , and also allowed for a more intuitive scoring method. This also removed a very complex calibration protocol with the DBSCAN distance and cluster size parameters in order to match the RBE of different radiation types and energies to the

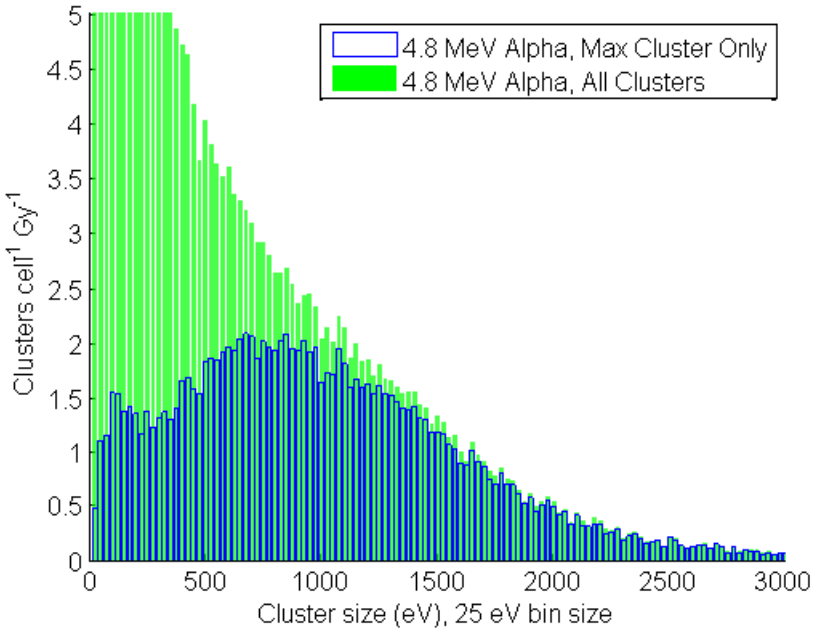


Figure 13: Clusters per Gy Histogram for 4.8-MeV Alpha Particles, All Clusters Versus Max Clusters to Fiber Only

literature, which was highly suspect and ultimately discarded. While the appearance of these figures are very similar, the analysis using the filtered data was both simpler and more similar to the experimental findings.

Since chromosome aberrations are considered to be the most prominent cause of cell death in irradiation studies, a more basic approach was used in assessing DSB yields and the probability of misjoining between two nearby DSBs to occur based on the radiation type and energy.

4.4 Energy Deposition Cluster Threshold Setting

Correlation of the track structure to DSB's was performed based on the hit size to individual s. This was calibrated using the electron track clustering algorithm in section 3.5.4, and Fig. 14 shows the 1-MeV electron clusters relative to carbon-12 ions and protons in the 29 keV/μm energy deposition ranges. 100eV was selected as the threshold for registering a DSB, which comprises of an average cluster size of 8 events of ionizations or excitations, as this value yields a close value to the 40 DSB's per Gy yield as generally accepted in the

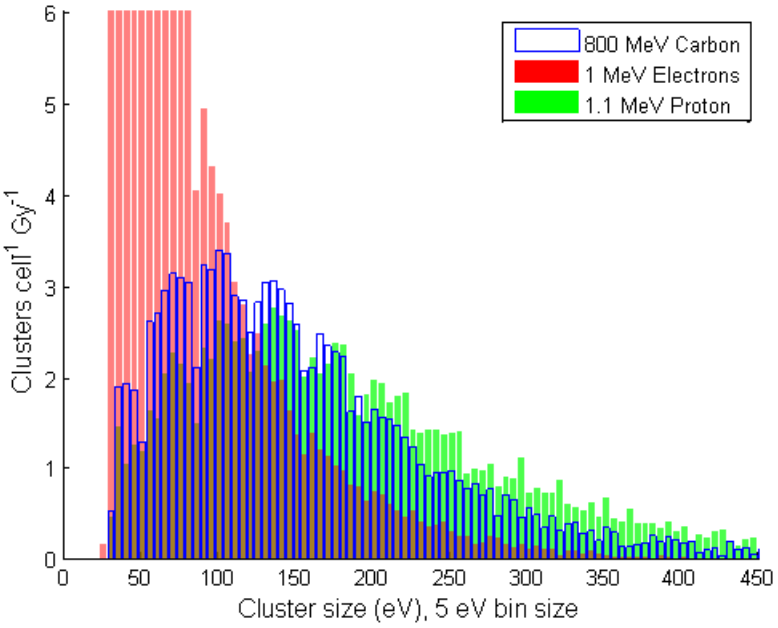


Figure 14: Cluster Sizes for Various Particles

literature for X-ray and electron radiation[28]. This was chosen under the assumption that the threshold represents simple DSBs (sDSBs).

Table 3 shows the number of chromatin fibers receiving clustered energy depositions at or above an energy threshold. A quick assessment shows that the number of s with clusters at or above threshold per 1 Gy of radiation varies based on the particle type and energy, and shows some of the expected variations based on track structure. In lower-energy higher-LET particles, such as the 2.5-MeV alpha particle or the 50-MeV carbon ions, the number of s per Gy does not vary with increasing energy thresholds, even though the threshold varies tremendously. This implies that the typical highest energy cluster to each respective is large. In the higher-energy lower-LET particles, increasing the energy threshold decreases the number of s meeting the threshold criteria. Comparing the different particles along a fixed threshold also yields some insight in how the track structure influences energy depositions.

In observing the proton energy depositions for a 100eV cluster threshold in table 3, there is an increase followed by a decrease in the number of s per 1 Gy registering above the

Table 3: Chromatin Fibers per Gy at or above different energy thresholds for various particles and initial energies with a 2nm DBSCAN radial distance parameter

Threshold(eV)	60	70	80	90	100	110	120	150	200
Alpha 20-MeV	105.1	99.9	94.6	88.4	81.9	76.7	70.6	54.1	33.6
Alpha 15-MeV	87.0	84.7	82.5	79.6	76.5	73.7	70.3	60.1	43.3
Alpha 10-MeV	67.2	66.5	65.9	65.0	64.0	63.1	61.9	58.1	50.4
Alpha 7-MeV	49.0	48.7	48.4	48.1	47.8	47.5	47.1	46.1	43.8
Alpha 4.8-MeV	35.0	34.9	34.8	34.6	34.5	34.3	34.1	33.6	32.8
Alpha 3.5-MeV	27.1	27.1	27.0	26.9	26.8	26.7	26.6	26.3	25.8
Alpha 2.5-MeV	19.0	18.9	18.9	18.8	18.8	18.7	18.7	18.6	18.3
Proton 3-MeV	142.9	123.3	106.5	90.4	77.9	68.3	58.6	36.9	17.1
Proton 2-MeV	133.6	120.8	109.4	98.3	87.6	78.2	68.9	47.5	24.2
Proton 1.1-MeV	108.7	105.1	101.1	97.1	92.3	87.9	83.0	68.0	46.6
Proton 0.8-MeV	86.5	85.4	84.1	82.6	81.0	79.4	77.5	71.0	58.8
Carbon 800-MeV	102.0	96.4	92.3	84.9	78.3	72.7	66.9	49.8	28.6
Carbon 500-MeV	70.6	69.3	67.8	65.9	63.8	61.8	59.5	51.7	38.5
Carbon 275-MeV	46.7	46.4	46.1	45.7	45.3	45.0	44.6	43.4	40.5
Carbon 200-MeV	35.2	35.1	34.9	34.7	34.6	34.4	34.2	33.6	32.6
Carbon 125-MeV	23.5	23.4	23.3	23.2	23.1	23.0	22.9	22.7	22.2
Carbon 50-MeV	11.9	11.9	11.9	11.9	11.8	11.8	11.8	11.7	11.5
Electron 1-MeV	99.4	78.0	64.2	52.9	43.9	37.6	31.7	19.4	8.8

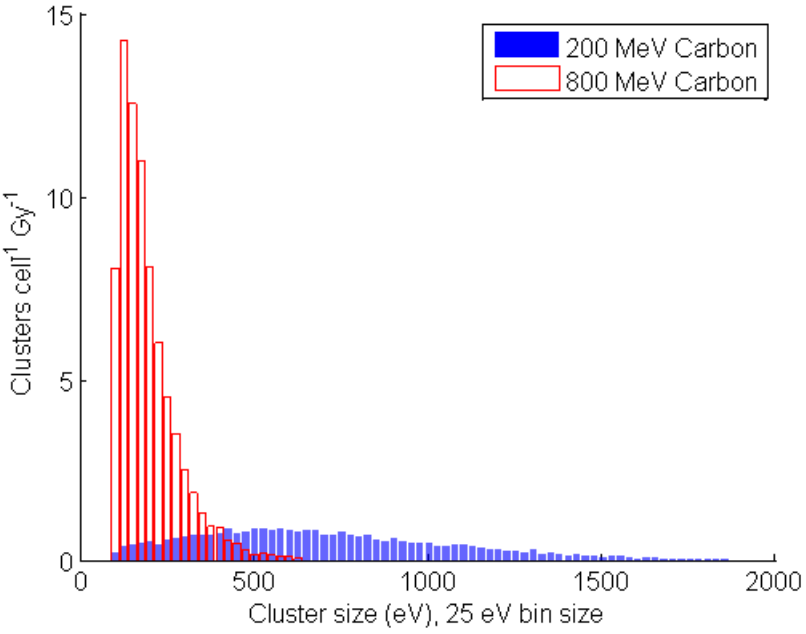


Figure 15: Hits per Gy for carbon ions of different LET with a 100 eV threshold

threshold as the LET increases. The increase is attributed to the connectedness of the track structure increasing with increasing LET of the proton track, wherein the excitations and ionizations from the proton, as well as its delta rays, have a shorter MFP between events. This yields clusters of larger quantity and higher energy when filtered through DBSCAN that are more capable of meeting the threshold requirements. The decrease in the 0.8-MeV proton from the 1.1-MeV proton is due to an overdose effect in the s and cell nucleus, wherein both the 0.8 and 1.1-MeV protons' energy deposition clusters to s are sufficient to meet the energy threshold, but the number of tracks per 1 Gy are higher in 1.1-MeV protons than in the 0.8-MeV protons. It should be noted that this pattern is visible in the 200 eV thresholds for the alpha particles and carbon ions, as well, and would also be visible if higher energy, lower-LET alpha particles and carbon ions were included in the 100 eV threshold from table 3.

4.5 Comparison of Clustered Energy Depositions Between Particles

A few important notes have to be made regarding observations from this section. The cluster threshold of 100 eV at a DBSCAN radial distance parameter of 2nm is used for

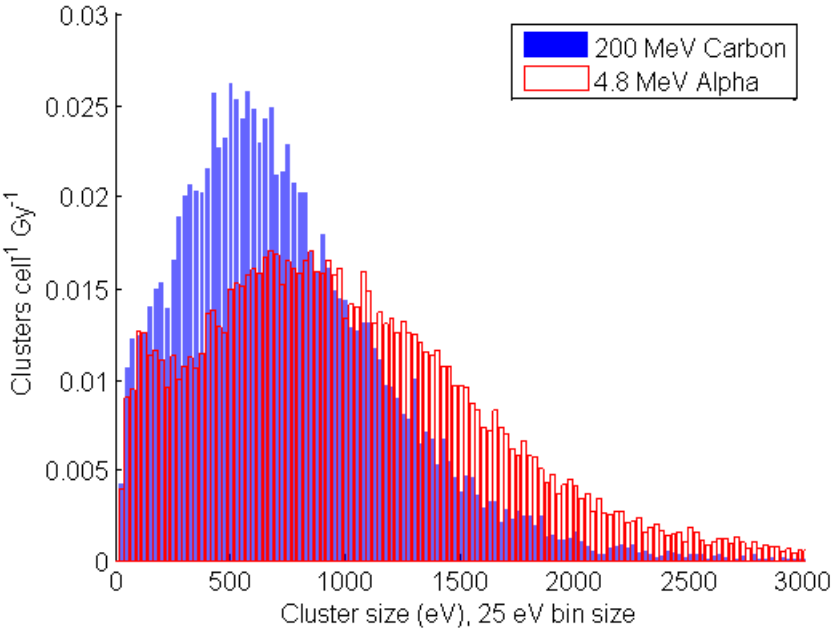


Figure 16: Hits per Gy for 100 keV/μm particles with a 100 eV threshold

counting DSB's, but not radiobiological consequence. The purpose of this is to correlate the proposed cell model to a quantifiable step in calibrating the model to experimentally observed responses to radiation damage, in this case the DSB's per 1 Gy. Table 3 simply quantifies the number of s with clusters above the threshold, but it does not assess the amount by which hits may exceed the threshold nor does it assess any probability of misjoining of two nearby DSBs. To stress this point, Fig. 15 shows the difference in energy depositions for 200-MeV and 800-MeV carbon ions of 100 and 29 keV/μm LET, respectively. While the total quantity of hits is higher in the 800-MeV carbon ions, the severity of the tallied clusters in the 200-MeV carbon ions are clearly higher on average. While the 800-MeV carbon ions deposit over double the quantity of DSBs relative to the 200-MeV carbon ions, their energy deposition is considerably smaller.

Figure 16 shows a more visual comparison between 4.8-MeV alpha particles and 200-MeV carbon ions, of which both have the same LET of 100 keV/μm. Table 3 shows that for a 100eV threshold, the DSB yield between 200-MeV carbon ions and 4.8-MeV alpha particles are nearly identical. The higher threshold energy values show that there are more

hits of higher severity for the 4.8-MeV alpha particles, and Fig. 16 shows this pattern clearly, as the mean energy of the clusters is significantly higher in the alpha particles than in carbon ions. This implies that, at this LET, the alpha particles are more effective in delivering more severe damage to s, and should have a higher radiological consequence since a larger and more energetic cluster will lead to more complex DSBs, which are more difficult to repair.

CHAPTER V

CONCLUSIONS & REMARKS

The presented research is a starting point in the development of a viable Monte Carlo based cell nucleus model for assessing damage caused by various types of ionizing radiation. The current approach demonstrates the expected behavior of the RBE-LET relationship in that more densely ionizing radiation tracks yield more heavily clustered energy depositions. Comparisons of particle tracks show that lighter ions produce larger hit sizes in chromatin fibers than do heavier ions of identical LET, which is consistent with the fact that the RBE of light ions is higher than that of heavy ions of identical LET (see Fig. 6). In other words, the behavior of charged particles and their energy depositions to the presented cell nucleus model matched what is understood about the physics processes and biology involved.

Data from the model have been output with quantitative detail on energy deposition and spatial geometry, which will allow for calculations related to specific biological endpoints of cell death or carcinogenesis. Current development of the presented model involves the interactions between chromatin fibers and domains to yield chromosomal aberrations, but is in the early stages. References have also been made regarding the radiobiological importance of DSB severity to chromatin fibers as quantified by clustered hit sizes, which will require comparisons to published experimental results.

The presented model is novel in that it adopts a step-wise Monte Carlo approach to a complex radiobiological problem. The end goal is to develop a fully Monte Carlo-based cell model that incorporates the particle track structure into the calculation of radiation induced damage. While the overall scope is challenging, a major advantage of the presented model is that it is inherently flexible owing to the stochastic approach it is predicated on. Future developments will involve the continued development of the cell model to incorporate chromatin fiber interaction probabilities, as well as a translation of the model outputs to allow comparisons to the wealth of epigenetic data being investigated.

APPENDIX A

GEANT4 INSTALLATION PARAMETERS

Geant4 is a powerful but complex tool for particle physics simulation; as such, its proper setup is crucial for extracting accurate and pertinent data. The Geant4 collaboration provides excellent documentation and user support for operating Geant4, provided the end user has some familiarity with C++ programming and a general understanding of the Monte Carlo method. As such, the setup details will largely ignore areas better covered by more qualified experts and focus on the topics that specifically address the modifications made to the Geant4-DNA microdosimetry example problem to output the particle track database required by the presented model. For an experienced user, such a basic explanation is a waste of time, but for a beginner, it is important to note that there is a sequence that Geant4 utilizes in calling to these different files.

The version of Geant4 used was 4.9.5 p02 on a Mac OSX 10.8 platform. Prior to any Geant4 installation, several other software packages should be installed, as this will allow for Geant4 to reference outside resources of which many Geant4 features are dependent. A functional version of CMake should be used; the newest revisions are not always compatible with some software packages, and for these simulations CMake 2.8.8 was chosen. Visualization managers should be assessed for the purpose of their use, and if specific file outputs are required an appropriate analysis manager would also become necessary. By default, many of the features specific users would necessitate are not activated in Geant4.

The default Geant4 visualilzation manager is a basic OpenGL command-driven window; however, most of the alternative visualization managers are an improvement on the default, and it is highly advisable to change from the default during installation. The SoQT OpenGL visuallization package was selected in our Geant4 installation commands due to the simplicity of its GUI for navigation. SoQT is nearly identical to the default OpenGL

visualization manager, but the viewing window can be controlled using a mouse and keyboard more intuitively, and its GUI accessible help tree provides a very good tutorial on the command structure of the Geant4 user interface, which provided the commands used in our macro files to drive our Geant4 simulations. Use of SoQT requires for it to be installed prior to activation in Geant4, and the sequence of installing SoQT is: installation of Coin3D, installation of QT4, installation of SoQT, and activation of SoQT in the Geant4 installation commands.

Since our use of Geant4 required specific file outputs, two analysis managers were installed. The original file manager explored was the Root manager, which refers to the Root analysis package made available by CERN. Root, like Geant4, is based on a C++ programming structure with slightly different commands recommended for use, but also differs from Geant4's modified C++ structure. Hence, for the beginner, utilization of Root may involve for learning the Geant4 C++ structure, a Root C++ structure, a Geant4 command line structure, and the operating system command line structure, which is a daunting task. For more complex analyses it may be beneficial to learn and utilize Root, as several of the Geant4-DNA examples were developed with Root output files designed into them. However, the Root software and programming metrics are complex and lacking in compatibility with external software packages, and so an AIDA based analysis manager was used to output particle data in a comma separated value, or .csv, file format, which allowed for post processing with any basic programming language, in our case Matlab. This dramatically reduced the Geant4 learning curve.

APPENDIX B

CELL MODEL CODE

The cell model code was run in Matlab using files output from Geant4. Several of the output data were processed from the Matlab user interface using basic sorting and filtering commands. However, cluster.m and scanner.m are the more intensive algorithms used for the cell nucleus model. Output files from scanner.m are processed using simple sorting and filtering commands from the Matlab user interface.

B.1 cluster.m: Cell Nucleus and Track Intersection Algorithm

```
% Note: cluster.m file reads primary.#.csv, secondary.#.csv, and HitToChromatin.#.csv.
% Output name has particle type & cell number, file has every event with CD & CF data
% cluster.m outputs: [x, y, z, interaction, energy, CF #, CD #, CD x, CD y, CD z]
for runs=1:5
    repository=strcat('HitToChromatin.',int2str(runs),'.csv');
    chromatins=csvread(repository,1,0); tracksinfile= max(chromatins(:,1));
    highestrow=max(chromatins(:,1)==tracksinfile);
    ra=cell(1,1,highestrow); rb=cell(1,1,highestrow);
    X=cell(1,1,highestrow); Y=cell(1,1,highestrow); Z=cell(1,1,highestrow);
    Points=cell(1,1,highestrow); S=cell(1,1,highestrow);
for trial=1:5000 r1=find(chromatins(:,1)==trial); topchrom=max(r1); botchrom=min(r1);
if isempty(botchrom) eventfinal=[0 0 0 0 0 0 0 0 0 0];
else
    beadfibers=topchrom-botchrom+1; primfile=chromatins(botchrom,3);
    primname = strcat('primary.', int2str(primfile), '.csv');
    secname = strcat('secondary.',int2str(primfile), '.csv');
    primar=csvread(primname,1,0); secon=csvread(secname,1,0);
```

```
    prima=[primar(:,1:3) primar(:,[4 6])]; seco=[secon(:,1:3) secon(:,[4 5])];
    primmy=[prima;seco]; inPri=zeros(size(primmy,1),1); inPr=[inPri,inPri];
for chromvalue=(botchrom:topchrom)
    chx=chromatins(chromvalue,15); primcomp=abs(primar(:,1)-chx);
    [unused,primind]=min(primcomp); yoff=primmy(primind,2); zoff=primmy(primind,3);
    chromposition=chromvalue-botchrom+1;
rachromvalue=[chromatins(chromvalue,4) (chromatins(chromvalue,5)+yoff)
    (chromatins(chromvalue,6)+zoff)];
rbchromvalue=[chromatins(chromvalue,7) (chromatins(chromvalue,8)+yoff)
    (chromatins(chromvalue,9)+zoff)];
[Xchromvalue,Ychromvalue,Zchromvalue,Pointschromvalue] =
    cylinder2Pb(15,10,rachromvalue,rbchromvalue);
    [V,Schromvalue]=alphavol(Pointschromvalue);
    inP=inpolyhedron(Schromvalue.bnd,Pointschromvalue,primmy(:,1:3));
    eventqty=sum(inP);
if eventqty==0
else
    try
    inPri=inPri+inP; inZ = inPr(:,1)==0; inY = inP.*inZ;
    inPr(inZ,1)=inY(inZ)*(chromvalue);
    inPr(inZ,2)=inY(inZ)*(chromatins(chromvalue,2));
    inPr(inZ,3)=inY(inZ)*(chromatins(chromvalue,11));
    inPr(inZ,4)=inY(inZ)*(chromatins(chromvalue,12));
    inPr(inZ,5)=inY(inZ)*(chromatins(chromvalue,13));
    catch err
    end
end end
if (sum(inPri))==0
    event=[0 0 0 0 0 0 0 0 0 0]; eventfinal=[0 0 0 0 0 0 0 0 0 0];
```

```
else event=[primmy(:,1:5) inPr]; eventhits=find(inPr(:,1)~=0); eventfinal=event(eventhits,:);
end
    eventname = strcat('proton1.1.', int2str(trial), '.csv'); any2csv(eventfinal,',',0,eventname);
    inPri=0; inPr=0;
end end clearvars -except runs; end
```

B.2 scanner.m: Cluster Processing Algorithm

```
% scanner.m reads the particle track intersections from the cluster.m output files
% algorithm runs intersections through DBSCAN and outputs all clusters and their info
% Output: [cell #, CD #, CF #, cluster #, ionizations, excitations, ion & exci,
%              energy in cluster, cluster x position, cluster y position, cluster z position]
    clustermatrix=zeros(300000,14);
    groggy=1;
for runs=1:5000
    repository=strcat('proton1.1.',int2str(runs),'.csv');
try chromatins=csvread(repository,1,0);
catch err
    chromatins=[ ];
end
if isempty(chromatins)
else
    beadmax=unique(chromatins(:,7)); fibmax=unique(chromatins(:,6));
for fibid=fibmax'
    chromatina=chromatins(chromatins(:,6)==fibid,:); clustersegment=chromatina(:,1:3);
try
    [clustid,idtype]=dbscan(clustersegment,clustsize,radius);
catch err
    clustid=[ ]; idtype=[ ];
end
```

```
if isempty(clustid)
else
    idtype=idtype'; clustid=clustid';
    clusterdata=[chromatina((idtype(:)~=(-1)),[1:5,7:10]),clustid(idtype(:)~=(-1))];
for beadid=beadmax'
    beadfib=clusterdata(clusterdata(:,6)==beadid,:);
if isempty(beadfib)
else clustermax=unique(beadfib(:,10));
for clusterbead=clustermax'
    beadfibclust=beadfib(beadfib(:,10)==clusterbead,1:9);
    elionidx=beadfibclust(beadfibclust(:,5)==13,2);
    prionidx=beadfibclust(beadfibclust(:,5)==18,2);
    ionqty=size(elionidx,1)+size(prionidx,1);
    elexciidx=beadfibclust(beadfibclust(:,5)==12,2);
    prexciidx=beadfibclust(beadfibclust(:,5)==17,2);
    exciqty=size(elexciidx,1)+size(prexciidx,1); ioexqty=ionqty+exciqty;
    ioexnrg=sum(beadfibclust(:,4)); ioexsize=size(beadfibclust,1);
clusterposition=[(sum(beadfibclust(:,1))/ioexsize),(sum(beadfibclust(:,2))/ioexsize),
    (sum(beadfibclust(:,3))/ioexsize),beadfibclust(1,7),beadfibclust(1,8),beadfibclust(1,9)];
    clustersumdata=[runs,beadid,fibid,clusterbead,ionqty,exciqty,ioexqty,ioexnrg,clusterposition];
try clustermatrix(groggy,:)=clustersumdata; groggy=groggy+1;
catch err
    clustermatrix=[clustermatrix;zeros(300000,14)]; clustermatrix(groggy,:)=clustersumdata;
    groggy=groggy+1;
end end end end end end end
    clearvars -except runs clustsize clustermatrix radius groggy;
end
    scanname=strcat('proton1.1clusterS',int2str(clustsize),'R',num2str(radius),'.csv');
    clustermatrix(~any(clustermatrix,2),:)=[ ]; dlmwrite(scanname,clustermatrix);
```

Bibliography

[1] Relative biological effectiveness in ion beam therapy. Technical Report 461, IAEA and ICRU, Vienna, 2008.

[2] D Ager, W Dewey, and et al. Measurement of radiation-induced dna double-strand breaks by pulsed-field gel electrophoresis. *Radiat Res*, 122:181–187, 1990.

[3] S Agostinelli and et al. Geant4- a simulation toolkit. *Nucl Instr & Methods in Phys Res A*, A(506):250–303, 2003.

[4] H Albiez, M Cremer, and et al. Chromatin domains and the interchromatin compartment form structurally defined and functionally interacting nuclear networks. *Chromosome Research*, 14:707–733, 2006.

[5] K Ando. Carbon-ion radiotherapy - basic and clinical studies. Gunma University, Maebashi Japan.

[6] K Ando and Y Kase. Biological characteristics of carbon-ion therapy. *Int J Radiat Biol*, 85(9):715–728, 2009.

[7] T. André and et al. Comparison of geant4-dna simulation of s-values with other monte carlo codes. *Nucl Instr & Methods in Phys Res B*, 319:87–94, 2014.

[8] Frank Herbert Attix. *Introduction to Radiological Physics and Radiation Dosimetry*. Wiley-VCH, Weinheim, 2004.

[9] M Bellamy. *A Deterministic Model of the Relative Biological Effectiveness of Photons and Electrons*. PhD thesis, Georgia Institute of Technology, Atlanta, GA, 2013.

[10] M Belli, F Cera, and et al. Rbe-let relationship for cell inactivation and mutation induced by low energy protons in v79 cells: further results at the lnl facility. *Int J Radiat Biol*, 74(4):501–509, 1998.

[11] D Bettega, P Calzolari, and et al. Neoplastic transformation induced by carbon ions. *Int J Radiat Oncol Biol Phys*, 73(3):861–868, 2009.

[12] M Branco and A Pombo. Chromosome organization: new facts, new models. *TRENDS in Cell Biology*, 17(3):127–134, 2007.

[13] D Brenner. The linear-quadratic model is an appropriate methodology for determining isoeffective doses at large doses per fraction. *Seminars in Radiat Oncol*, 18:234–239, 2008.

[14] C. Champion, S. Incerti, Y. Perrot, R. Delorme, M.C. Bordage, M. Bardiès, B. Mascialino, H.N. Tran, V. Ivanchenko, M. Bernal, Z. Francis, J.-E. Groetz, M. Fromm, and L. Campos. Dose point kernels in liquid water: An intra-comparison between geant4-dna and a variety of monte carlo codes. *Applied Radiation and Isotopes*, 83, Part B:137 – 141, 2014. Quantum scattering codes and Monte Carlo simulations to model dynamical processes in biosystems.

[15] D Charlton, D Goodhead, W Wilson, and H Paretzke. Energy deposition in cylindrical volumes. Monograph 85/1, M.R.C. Radiobiology Unit, Chilton, UK, 1985.

[16] T Cremer and C Cremer. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nature*, 2:292–301, April 2001.

[17] N Desai, E Davis, and et al. Immunofluorescence detection of clustered γ-h2ax foci induced by hze-particle radiation. *Radiat Res*, 164(4):518–522, Oct 2005.

[18] I El Naqua, P Pater, and J Seuntjens. Monte carlo role in radiobiological modelling of radiotherapy outcomes. *Phys Med Biol*, 57:R75–R97, May 2012.

[19] M Ester, H Kriegel, J Sander, and X Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. pages 226–231. AAAI Press, 1996.

[20] J Fowler. Development of radiobiology for oncology - a personal view. *Phys Med Biol*, 61:R263–R286, June 2006.

[21] Z Francis, S Incerti, V Ivanchenko, C Champion, M Karamitros, M A Bernal, and Z El Bitar. Monte carlo simulation of energy-deposit clustering for ions of the same let in liquid water. *Physics in Medicine and Biology*, 57:209–224, 2012.

[22] Z. Francis, S. Incerti, M. Karamitros, H.N. Tran, and C. Villagrasa. Stopping power and ranges of electrons, protons and alpha particles in liquid water using the geant4-dna package. *Nucl Instr & Methods in Phys Res B*, 269:2307 – 2311, 2011. 12th International Conference on Nuclear Microprobe Technology and Applications.

[23] N. Franken, R. Cate, and et al. Comparison of rbe values of high-let α-particles for the induction of dna-dsb's, chromosome aberrations and cell reproductive death. *Radiat Oncol*, 6(64):1–8, 2011.

[24] N. Franken, S. Hovingh, and et al. Relative biological effectiveness of high linear energy transfer α-particles for the induction of dna-double-strand breaks, chromosome aberrations and reproductive cell death in sw-1573 lung tumour cells. *Oncol Reports*, 27:769–774, Nov 2012.

[25] M Frankenberg-Schwager, A Gebauer, and et al. Single-strand annealing, conservative homologous recombination, nonhomologous dna end joining, and the cell cycle-dependent repair of dna double-strand breaks induced by sparsely or densely ionizing radiation. *Radiat Res*, 171:265–273, 2009.

[26] W Friedland, P Jacob, and P Kundrát. Stochastic simulation of dna double-strand break repair by non-homologous end joining based on track structure calculations. *Radiation Research*, 173(5):677–678, 2010.

[27] D Goodhead. *An Assessment of the Role of Microdosimetry in Radiobiology*, volume 91 of *45-76*. Radiation Research Society, July 1982.

[28] E Hall and A Giaccia. *Radiobiology for the Radiobiologist*. Lippincott Williams & Wilkins, Philadelphia, sixth edition, 2006.

[29] L Hernandez and et al. Highly sensitive automated method for dna damage assessment: gamma-h2ax foci counting and cell cycle sorting. *Int J Mol Sci*, 14:15810–15826, 2013.

[30] K Hofer, X Lin, and M Schneiderman. Paradoxical effects of iodine-125 decays in parent and daughter dna: A new target model for radiation damage. *Radiation Research*, 153:428–435, 2000.

[31] S Holcombe. Inpolyhedron - are points inside a triangulated volume (http://www.mathworks.com/matlabcentral/fileexchange/37856-inpolyhedron-are-points-inside-a-triangulated-volume-) matlab central file exchange, Feb 2014.

[32] S Incerti and et al. The geant4-dna project. *Int J Model Simul Sci Comput*, 01:157–178, 2010.

[33] S Incerti, A Ivanchenko, and et al. Comparison of geant4 very low energy cross section models with experimental data in water. *Med Phys*, 37(9):4692–4708, 2010.

[34] S Incerti, M Psaltaki, and et al. Simulating radial dose of ion tracks in liquid water simulated with geant4-dna: A comparative study. *Nucl Instr & Methods in Phys Res B*, 333:92–98, 2014.

[35] K Kanai, M Endo, and et al. Biophysical characteristics of himac clinical irradiation system for heavy-ion radiation therapy. *Int J Radiat Oncol Biol Phys*, 44(1):201–210, 1999.

[36] C Lanctôt, T Cheutin, M Cremer, G Cavalli, and T Cremer. Dynamic genome architecture in the nuclear space: regulation of gene expression in three dimensions. *Nature*, 8:104–115, Feb 2007.

[37] Y Li and F Cucinotta. Modeling non-homologous end joining. *Journal of Theoretical Biology*, 283:122–135, 2011.

[38] E Lubeck and L Cai. Single-cell systems biology by super-resolution imaging and combinatorial labeling. *Nature Methods*, 9(7):743–750, July 2012.

[39] K. Luger and T. Richmond. Dna binding within the nucleosome core. *Current Opinion in Structural Biology*, 8(1):33–40, 1998.

[40] J Lundgren. Alpha shapes (http://www.mathworks.com/matlabcentral/fileexchange/28851-alpha-shapes/content/alphavol.m) matlab central file exchange, March 2012.

[41] N. Metropolis and S. Ulam. The monte carlo method. *Journal of the American Statistical Association*, 44(247):335–341, 1949.

[42] H. Nikjoo and P. Girard. A model of the cell nucleus for dna damage calculations. *Int J Radiat Biol*, 88:87–97, 2012.

[43] D. Olins and A. Olins. Chromatin history: our view from the bridge. *Nat Rev Mol Cell Biol*, 4(10):809–814, Oct 2003.

[44] P Olive and J Banath. The comet assay: a method to measure dna damage in individual cells. *Nature Protocols*, 1(1):23–29, 2006.

[45] W Purves and et al. *Life: the science of biology.* Sinauer Associates and W.H. Freeman, seventh edition, 2004.

[46] M R Raju. Review: proton radiobiology, radiosurgery, and radiotherapy. *Int J Radiat Biol*, 67(3):217–259, 1995.

[47] M Dos Santos, C Villagrasa, I Clairand, and S Incerti. Influence of the dna density on the number of clustered damages created by protons of different energies. *Nucl Instr & Methods in Phys Res B*, 298:47–54, 2013.

[48] P Sundqvist. Cylinder between 2 points, July 2004.

[49] R Taleei, P Girard, S Krishnaswami, and H Nikjoo. The non-homologous end-joining (nhej) pathway for the repair of dna double-strand breaks: Ii. application to damage induced by ultrasoft x rays and low-energy electrons. *Radiation Research*, 179(5):540–548, 2013.

[50] R Taleei and H Nikjoo. The non-homologous end-joining (nhej) pathway for the repair of dna double-strand breaks: I. a mathematical model. *Radiation Research*, 179(5):530–539, 2013.

[51] C K Wang. The progress of radiobiological models in modern radiotherapy with emphasis on the uncertainty issue. *Mutation Research*, 704:175–181, Feb 2010.

[52] J Watson and F Crick. Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. *Nature*, 171:737–738, April 1953.

[53] Z.Francis, S. Incerti, and et al. Monte carlo simulation of energy-deposit clustering for ions of the same let in liquid water. *Phys Med Biol*, 57:209–224, Dec 2012.