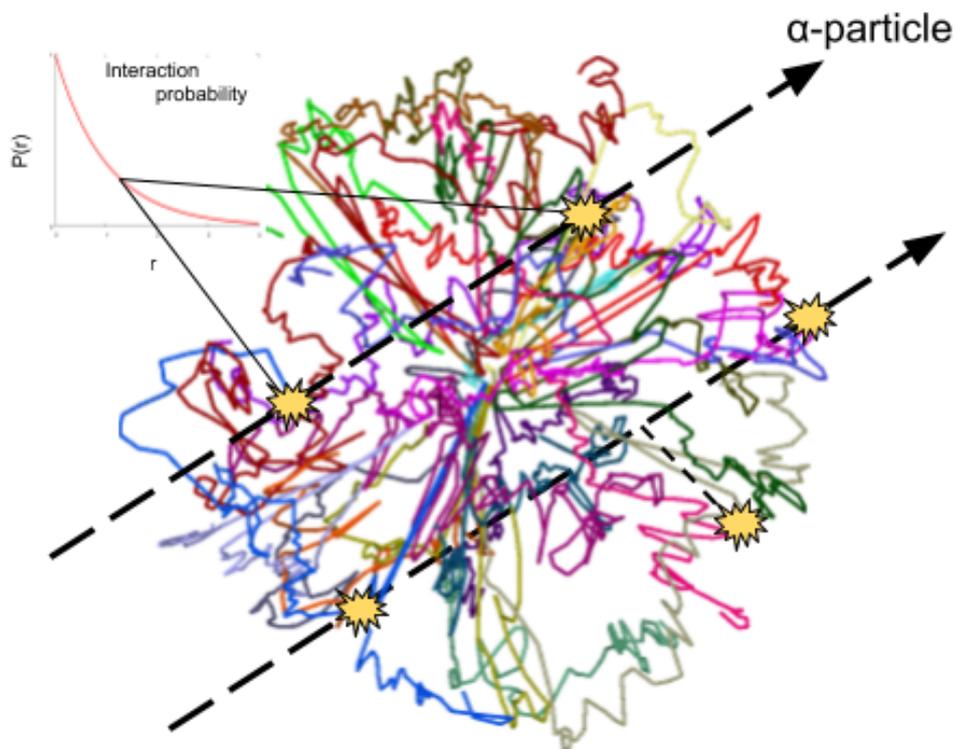




UNIVERSIDADE ESTADUAL DE CAMPINAS
UNIVERSITÀ DEGLI STUDI DI PAVIA
INFN - ISTITUTO NAZIONALE DI FISICA NUCLEARE
DOTTORATO DI RICERCA IN FISICA - XXXII CICLO

Biophysical modelling of radiation-induced chromosome aberrations

John James Tello Cajiao



Tesi per il conseguimento del titolo



UNIVERSIDADE ESTADUAL DE CAMPINAS
UNIVERSITÀ DEGLI STUDI DI PAVIA
INFN - ISTITUTO NAZIONALE DI FISICA NUCLEARE
DOTTORATO DI RICERCA IN FISICA - XXXII CICLO

Biophysical modelling of radiation-induced chromosome aberrations

John James Tello Cajiao

Dissertation presented to the Graduate School in
Physics in partial fulfillment of the requirements for
the degree of PhD in Physics.

Supervisor: Prof. Francesca Ballarini

Supervisor: Prof. Mario Antonio Bernal Rodriguez

THIS EXEMPLAR CORRESPONDS TO THE FINAL VERSION DEFENDED BY THE STUDENT JOHN JAMES TELLO CAJIAO, AND SUPERVISED BY PROF. FRANCESCA BALLARINI AND PROF. MARIO ANTONIO BERNAL RODRIGUEZ.

Campinas, Pavia

2019

Cover: Genome 3D structure reconstructed from a HiC map. A radiation track induces breaks that may interact with a given probability.

Biophysical modelling of radiation-induced chromosome aberrations

John James Tello Cajiao

PhD Thesis - University of Pavia & University of Campinas

Pavia, November 2019

To my beautiful daughter Isabella.

Acknowledgements

Firstly, I would like to thank to my supervisors Francesca and Mario for being more than just guides through this project. I learned so many things from them, they shared with me their knowledge and experience, and were always kind to me. Working with them along these years has made me a better person and scientist.

A big "thank you" to each of my friends and colleagues for their valuable academic and emotional support. Especially, to Mario C. (capo), Monica V., Daniel P., Sandra B., Brooke L., Yulle G., Pedro S., Alessandro C. and Alessia E., all of you contributed to this project in one way or another, and I always be grateful for that. Also to my old friends: Christian T., Andrés Q., Natalia Q., Bryan P., Christian M., Henry P., and the new ones: Andrea F. Francesco C., Rodi, Vania, Irene and many others that have come and gone.

Also, I want to thank to my family for all they have done for me. I have no words to describe my gratitude to my parents James and Carmenza, to whom I owe all what I am. To my sister Maye, for being unconditionally supportive and for always being there. To Isabella and Alejandro for being my inspiration.

Finally, this project was partially carried out within the INFN projects "ETHICS" and "MC-INFN". Additionally, the Brazilian National Council for Scientific and Technological Development (CNPq) also funded part of this PhD work. Many thanks to these institutions for the financial support.

*"El día que la mierda tenga algún valor,
los pobres nacerán sin culo."*

Gabo

Abstract

In this PhD work, the mechanisms of chromosome aberration formation by ionizing radiation were investigated by studying their dependence on proximity effects (i.e., dependence of the chromosome fragments rejoining probability on the fragment distance), radiation quality, cell line, and dose. To do so, several changes in the BIANCA (BIophysical ANalysis of Cell death and chromosome Aberrations) model, seeking to improve the performance and expand its capabilities, were carried out. As a result, the BIANCA model was upgraded to explicitly represent arm domains within chromosomes, two new rejoining probability functions were included and simulations of microbeam irradiations were also made available. Additionally, the overall performance of the code was considerably improved allowing more and better simulation setups. The upgraded version of BIANCA was then applied in a number of contexts testing: the new hypothesis regarding proximity effects, the model capability to provide theoretical explanations for microbeam experimental data, and the model predictive power in terms of cell survival.

First, BIANCA was applied to human lymphocytes and fibroblasts exposed to a wide range of radiation qualities and doses, investigating the behaviour of an exponential and a Gaussian probability function of the form $P(r) = e^{-r/\sigma}$ and $P(r) = e^{-r^2/2\sigma^2}$, respectively, to model proximity effects. Yields of the main chromosome aberration categories were calculated using both functions and compared against a variety of experimental datasets. The outcomes of the simulations showed that, for lymphocytes, an exponential function can describe proximity effects better than a Gaussian one, independent of radiation quality and for the considered dose range. On the other hand, for fibroblasts, while the exponential function still has the upper hand, the results at high LET values deserve further investigation. Moreover, not only the trends of the absolute yields, but also the trends of their relative ratios (F - and G -ratios), were reproduced with the exponential function as well.

Later, in collaboration with the BioQuaRT (Biologically Weighted Quantities in RadioTherapy) project, the model was used as a means to interpret experimental data from a theoretical point of view in experiments on hamster ovary cells irradiated with α -particle microbeams. Higher yields of chromosome aberrations were observed at the highest LET radiation (85 *vs* 17 keV/ μ m). The irradiation conditions were reproduced with BIANCA, which allowed to interpret the results in terms of the particle effectiveness at inducing DNA cluster lesions. In another collaboration, the predictive power of BIANCA in terms of cell survival was tested by studying the behaviour of the model parameters in a number of datasets covering hamster and fibroblasts cells irradiated by protons, C- and He-ions, at energies and doses of interest for tumour hadrontherapy. Based on this study, an approach to predict cell survival curves for cell lines and LET values not available experimentally, was proposed.

Finally, in the pursue of further improvements to the model, part of this PhD work was invested in the development of a fortran code that aims at modelling the 3D structure of the genome starting from HiC data (i.e. data on *in vivo* contact frequency between all loci of the genome). The code is built on a restraint-based

method called LorDG. After benchmarking a first version of our code against the original implementation, several optimizations have been carried out in order to make it suitable for integration with BIANCA. So far, the construction of diploid cell models from haploid data and the implementation of libraries to manage the vast amount of data coming from HiC datasets, have been accomplished. Additionally, a coarse-grain/top-bottom method is under development. This method aims at producing higher resolution conformations (down to tens of kilo-basepair) of the chromatin, while keeping reasonable the demand of computational resources.

Key-words: Biophysical modelling. Chromosome aberrations. BIANCA. Computational radiobiology. HiC maps

List of Abbreviations

3C	Chromosome Conformation Capture
BIANCA	Biophysical ANalysis of Cell death and chromosome Aberrations
BioQuaRT	Biologically weighted Quantities in RadioTherapy
bp	base-pair
CA	Chromosome Aberration
CL	Cluster Lesion
CLy	Cluster Lesion yield
CT	Chromosome Territory
DNA	Deoxyribonucleic Acid
DNA-PK	DNA-Protein Kinase
DSB	Double-Strand Break
FISH	Fluorescent In Situ Hibridization
fLorDG	Lorentzian 3D Genome (fortran implementation)
FWHM	Full Width at Half Maximum
HR	Homologous Repair
HiC	High-throughput Chromatin Conformation Capture
IR	Ionizing Radiation
LEM	Local Effect Model
LET	Linear Energy Transfer
LorDG	Lorentzian 3D Genome
MC	Monte-Carlo
MKM	Microdosimetric Kinetic Model
MMEJ	Microhomology Mediated End Joining
NHEJ	Non-Homologous End Joining
PCC	Premature Chromosome Condensation
RBE	Relative Biological Effectiveness
RMF	Repair-Mis-repair Fixation
RMSE	Root Mean Squared Error
ROI	Region Of Interest
SSA	Single-Stranded Annealing
SSB	Single-Strand Break
SSD	Standard for DNA Damage
TAD	Topologically Associating Domain

List of Figures

Figure 1 – Direct and Indirect action of ionizing radiation on the genetic material. The very breakage of a bond that splits the molecule and provokes chemical reactions, constitutes a direct action. The creation of free radicals after the hit, which attack the DNA inducing changes, represents an indirect action of radiation. Source: Desouky <i>et al.</i> J. Rad: Res. (2015) [1].	25
Figure 2 – A scheme of the different organizational levels of the chromatin. At the nucleosomal scale, the DNA follows scale-invariant principles along the sequence and shows long range correlations in the primary structure that are related to nucleosome formation and positioning. Before the mega base pair (Mbp) regime, the chromatin self-organizes in globular states that are in agreement with the fractal globule. From chromatin loops to TADs and compartments, there appear correlations between chromatin states, accessibility and composition (e.g. gene content or presence of promoters/enhancers), and the spatial distribution and contact profile of the chromatin. At the chromosome scale (tens of Mbps and higher), the distinction between territories is prominent. Chromosomes are organized in well-defined domains, still, some regions intermingle with other domains. Source: Vuthy et al. Genes (2015) [3].	29
Figure 3 – Two-break types of chromosome and chromatid aberrations. exchanges can be Symmetric (S) or Asymmetric (A) depending on whether an acentric fragment is formed or not, provided that the exchange is complete (that is, all four ends rejoin). Additionally, the exchanges may occur between different chromosomes (<i>interchanges</i>), or between fragments of the same chromosome (<i>intra- and inter-arm intrachanges</i>). For chromatid-type aberrations, the intrachanges can be further classified as <i>intra-</i> or <i>inter-chromatid</i> depending on how the sister-chromatids mix. Image adapted from: Savage, J. Med. Genet. (1976) [4].	31
Figure 4 – Visualization of chromosome aberrations with cross-species color banding (Rx-FISH). (A) Translocation. (B) Symmetrical inter-arm exchange. (C) Asymmetrical inter-arm intrachange. Image adapted from: M. Durante <i>et al.</i> Mutat. Res. (2013) [5].	33

Figure 5 – Diagrammatic representation of the formation of some CAs by HR, SSA and NHEJ involving up to two DSBs and chromosomes. Red boxes and thunders represent DSBs. The green and blue horizontal lines represent chromosomes and their telomeres (vertical box). Pink arrows on top of a chromosome intend the orientation of the strand (needed by HR and SSA as they require sequence homology). (1-2) When HR occur between regions of sequence homology but located at different chromosomes (Ectopic HR), reciprocal translocations and dicentrics can arise along with acentric fragments. (3) In NHEJ, at least two DSB are needed to form an exchange. A dicentric is formed when two centromere-containing fragments are joined after being processed by the DNA-PK complex. One or two acentric fragments can also accompany the exchange, resulting in a complete or incomplete exchange respectively. (4) an example of an intra-chromosomal deletion as a consequence of SSA when the DSBs lie in an extensive region of sequence homology. Figure modified from G. Obe <i>et al.</i> <i>Mutat. res.</i> (2002) [6].	35
Figure 6 – Kinetics of the formation of exchanges in human cells after γ irradiation. Durante <i>et al.</i> [7] and Gotoh <i>et al.</i> [8] studied the formation of chromosome exchanges in G1 and G2 phase, respectively. They concluded that in both phases, each type of chromosome and/or chromatid exchange follow different kinematics so there is a contribution from fast and slow components to the total yield. Source: M. Durante <i>et al.</i> <i>Mutat. Res.</i> (2013) [5].	37
Figure 7 – Relationship between the average number of "lethal" aberrations (dicentrics + rings + large deletions) and the natural logarithm of the surviving fraction of cells. Source: Cornforth and Bedford, <i>J. Rad. Res.</i> (1987) [9].	38
Figure 8 – RBE values for monoenergetic ion beams of different qualities and in the dose range of 4 – 30 Gy. Color code: p = red; He = blue; C = gray; Ne = orange; heavier ions than Ne = green. The RBE values show a peak around 100 keV/ μ m and then a decrease even at higher LET and doses due to saturation effects. Source: Friedrich <i>et al.</i> <i>J. Radiat. Res.</i> (2013) [2].	41

Figure 9 – Illustration of the track structure of two different radiation qualities. Blue (top): 10 MeV proton ($\sim 5 \text{ keV}/\mu\text{m}$). Red (Bottom): 200 MeV Carbon ion ($\sim 100 \text{ keV}/\mu\text{m}$). In both cases, the primary particles deposit a small fraction of their energy along their track (solid lines). Most of the energy is deposited by secondary electrons (points) near the track core, with notorious differences in their corresponding energy distribution patterns. A DNA structure is superimposed for scale. This highlights the importance of nanodosimetric calculations as densely ionizing radiation (e.g. Carbon ions in the figure) has the potential to induce more breaks in the genome than sparsely ionizing radiation (e.g. protons in the figure) at such scale. Source: S. McMahon and K. Prise. Cancers. (2019) [10].	46
Figure 10 – Scheme of the NHEJ. The process, as implemented in Friedland <i>et al.</i> (2010), is divided in three phases: The <i>pre-synaptic</i> phase, in which DNA ends diffuse freely and, eventually, can be attached to Ku70/80 heterodimer and DNA-PKcs. The <i>synaptic</i> phase, where two DNA ends in sufficient close proximity, move together while are prepared for the next phase. In the <i>post-synaptic</i> phase, repair enzymes attempt to clean up dirty DSBs and finalize the rejoining process. Source: W. Friedland <i>et al.</i> . Radiat. Res. (2010) [11].	49
Figure 11 – Survival curves for V79 cells exposed to X-rays and to four carbon-ion beams of differing energy. Simulations were carried out using the first version of BIANCA. Source: Ballarini <i>et al.</i> Rad. Res. (2013) [12].	57
Figure 12 – From top to bottom: survival of AG1522 cells exposed to γ -rays (two curves: the upper one compared with data from Hamada et al. 2006 [13], and the lower one compared with data from Neti et al. 2004 [14]), $16.2 \text{ keV}/\mu\text{m}$ He ions, $437 \text{ keV}/\mu\text{m}$ Ne ions, $76.3 \text{ keV}/\mu\text{m}$ C ions, $321 \text{ keV}/\mu\text{m}$ Ne ions, $108 \text{ keV}/\mu\text{m}$ C ions, and $132 \text{ keV}/\mu\text{m}$ α -particles. The lines are simulation results, whereas the points are experimental data taken from Hamada et al. 2006 (upper gamma curve, He ions, C ions, and Ne ions) or Neti et al. 2004 (lower gamma curve and α -particles). Source: Ballarini <i>et al.</i> Rad. Res. (2014) [15].	58
Figure 13 – Yields (mean number per cell) of different aberration types (dicentrics, rings and deletions, as well as total aberrations) in AG1522 primary normal human fibroblasts exposed to different doses of X-rays. The lines are simulation outcomes, the points are experimental data taken from [9].). Source: Carante <i>et al.</i> Rad. Res. (2016) [16].	59

Figure 14	–Input file of BIANCA. The admitted values of each key are shown on the right as comments. The target information is controlled by the <i>Cell_Dish</i> keyword. <i>Beam</i> , <i>Particle</i> and <i>doses</i> determine the irradiation setup. The adjustable parameters, namely, the CLy and σ and the selection of the model are passed through the <i>Cluster_Yield</i> , <i>my-dist</i> and <i>Model</i> keywords, respectively. The order of the lines does not matter but for a given line, the key values order does.	61
Figure 15	–An example of the cell models available in BIANCA. On top, an AG fibroblast grown as monolayer modelled as an elliptical cylinder. At the bottom, on the left, a lymphocyte grown in suspension simulated as a sphere. V79 cells are shaped as circular cylinders. Dimensions of each geometrical model are those shown in Table 3 . Each color represents a chromosome territory. Changes of tonality within a chromosome are intended to highlight its arm domains.	63
Figure 16	–Scoring scheme adopted in BIANCA. For simple exchanges (panels (a),(c) and (e)), each aberration must be accompanied by at least one linear fragment. (b) multicentric linear fragments must be associated to at least $n - 1$ linear fragments (where n is the number of centromeres) in order to be scored. (b) In the case of multicentric circular fragments, at least n linear fragments are expected. Complex exchanges as in (b) and (d) can be distinguished and scored separately or translated to simple exchanges. The number of "equivalent simple exchanges" is equal to the minimum number of linear fragments associated to the multicentric fragment. Acentric linear fragments smaller than 3 <i>Mbps</i> are not considered during the assessment of chromosome aberrations. When one or more exchanges are not complete (illustrated in the upper case in the brackets), one or more Excess Acentric Fragments (EAFs) may be counted as well. The order of the panels reflects the priority of the aberration categories, when there is more than one type of aberration presented.	73
Figure 17	–Simplification of the HiC map concept. The entries of the matrix represent the number of times that a pair of chromosome loci were in close proximity to each other. In the figure, the construction of an inter-chromosome HiC map is illustrated.	74
Figure 18	–Iterative and integrative process for model building. The iterative process consists of data acquisition, model representation and scoring, model optimization and model analysis. Source: Dekker <i>et al.</i> . Nat. Rev. Gen. (2013) [17].	75

Figure 19	–Aberrations in lymphocytes calculated by the exponential model. Full symbols are experimental values taken from [18]; the error bars represent one standard deviation from the mean, calculated from the raw numbers (both ^{137}Cs γ –ray data sets pooled) assuming Poisson statistics. Empty symbols, connected by lines to guide the eye (from top to bottom: total aberrations, dicentrics, excess acentric fragments and centric rings), are results of simulations performed by the exponential model, with $\sigma = 0.8 \mu\text{m}$ and $\text{CL} = 3.1 (\text{GyCell})^{-1}$. Source: Tello <i>et al.</i> DNArepair (2017) [19].	85
Figure 20	–Aberrations in lymphocytes calculated by the Gaussian model. Like in Figure 19, full symbols are experimental values taken from [18], with standard deviations calculated assuming Poisson statistics. Empty symbols, connected by lines to guide the eye (from top to bottom: total aberrations, dicentrics, excess acentric fragments and centric rings), are results of simulations performed by the Gaussian model, with $\sigma = 1.1 \mu\text{m}$ and $\text{CL} = 2.3 (\text{GyCell})^{-1}$. Source: Tello <i>et al.</i> DNArepair (2017) [19].	86
Figure 21	–F- and G-ratio for lymphocytes irradiated by $3 \text{Gy } ^{137}\text{Cs}$ γ –rays. Points represent values of F (open symbols) and G (closed symbols) calculated with the exponential model for different values of σ , whereas the upper and lower solid lines represent the (mean) experimental value of F and G, respectively, according to [18]. Source: Tello <i>et al.</i> DNArepair (2017) [19].	88
Figure 22	–Aberrations in AG1522 fibroblasts calculated by the exponential model. Full symbols represent experimental values taken from [20] ("Day 0" data set, with error bars representing one standard deviation calculated from the raw numbers assuming Poisson statistics), whereas the empty symbols, connected by lines to guide the eye (from top to bottom: total aberrations, dicentrics, excess acentric fragments and centric rings), are results of simulations performed by the exponential model (with $\sigma = 0.7 \mu\text{m}$ and $\text{CL} = 4.0 (\text{GyCell})^{-1}$). Source: Tello <i>et al.</i> DNArepair (2017) [19].	89
Figure 23	–Aberrations in AG01522 fibroblasts calculated by the Gaussian model. Like in Figure 22, full symbols represent experimental values taken from [20]; empty symbols, connected by lines to guide the eye (from top to bottom: total aberrations, dicentrics, excess acentric fragments and centric rings), are results of simulations performed by the Gaussian model with $\sigma = 1.3 \mu\text{m}$ and $\text{CL} = 2.0 (\text{GyCell})^{-1}$. Source: Tello <i>et al.</i> DNArepair (2017) [19].	91

- Figure 24 – Aberrations in human lymphocytes exposed to $3.5 \text{ keV}/\mu\text{m}$ protons. Empty symbols, connected by lines to guide the eye (solid line for dicentric, dashed line for excess acentric and dotted line for centric rings), represent simulation outcomes obtained with the exponential model, taking $CL = 0.02 \mu\text{m}^{-1}$. Full symbols are experimental values taken from Bauchinger and Schmid [18]; the error bars represent one standard deviation from the mean, calculated from the raw numbers assuming Poisson statistics. Source: Tello *et al.* DNArepair (2018) [21]. 94
- Figure 25 – Aberrations in human lymphocytes exposed to $5.3 \text{ keV}/\mu\text{m}$ protons. Empty symbols, connected by lines to guide the eye (solid line for dicentric, dashed line for excess acentric and dotted line for centric rings), represent simulation outcomes obtained with the exponential model, taking $CL = 0.03 \mu\text{m}^{-1}$. Full symbols are experimental values taken from Bauchinger and Schmid [18]; the error bars represent one standard deviation from the mean, calculated from the raw numbers assuming Poisson statistics. Source: Tello *et al.* DNArepair (2018) [21]. 95
- Figure 26 – Aberrations in human lymphocytes exposed to $19.0 \text{ keV}/\mu\text{m}$ protons. Empty symbols, connected by lines to guide the eye (solid line for dicentric, dashed line for excess acentric and dotted line for centric rings), represent simulation outcomes obtained with the exponential model, taking $CL = 0.21 \mu\text{m}^{-1}$. Full symbols are experimental values taken from Bauchinger and Schmid [18]; the error bars represent one standard deviation from the mean, calculated from the raw numbers assuming Poisson statistics. Source: Tello *et al.* DNArepair (2018) [21]. 95
- Figure 27 – Aberrations in human lymphocytes exposed to $150 \text{ keV}/\mu\text{m}$ α -particles. Empty symbols, connected by lines to guide the eye (solid line for dicentric, dashed line for excess acentric and dotted line for centric rings), represent simulation outcomes obtained with the exponential model, taking $CL = 1.06 \mu\text{m}^{-1}$. Full symbols are experimental values taken from Bauchinger and Schmid [18]; the error bars represent one standard deviation from the mean, calculated from the raw numbers assuming Poisson statistics. Source: Tello *et al.* DNArepair (2018) [21]. 96

Figure 28	– Aberrations in human fibroblasts exposed to $116 \text{ keV}/\mu\text{m}$ α –particles. Empty symbols, connected by lines to guide the eye (solid line for di-centrics, dashed line for excess acentrics and dotted line for centric rings), represent simulation outcomes obtained with the exponential model, taking $CL = 0.47 \mu\text{m}^{-1}$. Full symbols are experimental values taken from Cornforth et al. [20]; the error bars represent one standard deviation from the mean, calculated from the raw numbers assuming Poisson statistics. Source: Tello <i>et al.</i> DNArepair (2018) [21].	100
Figure 29	– (a) Survival of V79 cells irradiated with protons of different LET (7.7, 11.0, 20.0 and $30.5 \text{ keV}/\mu\text{m}$), as well as X-rays. The lines are simulation outcomes, the points are experimental data taken from Belli <i>et al.</i> (1998) [22]. (b) Survival of AG01522 cells irradiated with protons of different LET (from top to bottom: 0.63, 1.68, 2.45 and $7.50 \text{ keV}/\mu\text{m}$). The lines are simulation outcomes, the points are experimental data taken from Marshall <i>et al.</i> (2016) [23]. Source: Carante <i>et al.</i> PMB. (2018) [24]. . .	107
Figure 30	– CL yields used to simulate the survival of proton-irradiated V79 cells (lower line and points) and AG01522 cells (upper line and points). Each point represents the mean number of CLs per micrometer used as a code input, whereas the lines are linear-quadratic fits. A 5% relative error was assigned to each CL yield. Source: Carante <i>et al.</i> PMB. (2018) [24].	109
Figure 31	– Survival of V79 cells irradiated by Carbon beams of different (dose-averaged) LET, as well as X-rays as a reference. Panel (a): 22.5, 360.0, 102.0 and $137.0 \text{ keV}/\mu\text{m}$. Panel (b): 31.0, 502.0, 78.5 and $206.0 \text{ keV}/\mu\text{m}$. The points are simulation outcomes, whereas the lines are experimental data fits of the form $S(D) = \exp(-\alpha D - \beta D^2)$ taken from Furusawa <i>et al.</i> (2000) [25]. Source: Carante <i>et al.</i> PMB. (2018) [24].	110
Figure 32	– Survival of AG01522 cells irradiated by Carbon beams of different LET (from top to bottom: 48.8, 76.3, 108.0 and $147.6 \text{ keV}/\mu\text{m}$). The lines are simulation outcomes, whereas the points are experimental data taken from Hamada <i>et al.</i> (2006) (76.3 and $108.8 \text{ keV}/\mu\text{m}$) [13] and Kavanagh <i>et al.</i> (48.8 and $147.6 \text{ keV}/\mu\text{m}$) (2013) [26]. Source: Carante <i>et al.</i> PMB. (2018) [24].	110

- Figure 33 –(a) CL yields used to simulate the survival of carbon-irradiated AG01522 cells (upper line and points) and V79 cells (lower line and points) up to $\sim 150 \text{ keV}/\mu\text{m}$. Each point represents the mean number of CLs per micrometer used as a code input, whereas the lines are linear-quadratic fitting functions. A 5% relative error was assigned to each point. (b) CL yields used to simulate the survival of Carbon-irradiated V79 cells over the whole LET range ($22.5 - 502 \text{ keV}/\mu\text{m}$). Each point represents the mean number of CLs per micrometre used as a code input, whereas the line is a fit of the form $c \cdot \arctan(aL + bL^2)$. A 5% relative error was assigned to each point. Source: Carante *et al.* PMB. (2018) [24]. 111
- Figure 34 –(a) Survival of V79 cells irradiated by He-ion beams with different (dose-averaged) LET (from top to bottom: 18.6, 29.9, 50.0, 73.9 and $90.8 \text{ keV}/\mu\text{m}$). The points are simulation outcomes, whereas the lines are experimental data fits of the form $S(D) = \exp(-\alpha D - \beta D^2)$ taken from Furusawa *et al.* (2000) [25]. (b) Survival of AG01522 cells irradiated by He-ion beams of $16.2 \text{ keV}/\mu\text{m}$ (upper line and points) and $132 \text{ keV}/\mu\text{m}$ (lower line and points). The lines are simulation outcomes, whereas the points are experimental data taken from Hamada *et al.* (2006) ($16.2 \text{ keV}/\mu\text{m}$) [13] and Neti *et al.* (2004) ($132 \text{ keV}/\mu\text{m}$) [14]. Source: Carante *et al.* PMB. (2018) [24]. 113
- Figure 35 –CL yields used to simulate the survival of He-ion-irradiated V79 cells (lower line and points) and AG01522 cells (upper line and points). Each point represents the mean number of CLs per micrometre used as a code input, whereas the lines are linear-quadratic fits. A 5% relative error was assigned to each CL yield. Source: Carante *et al.* PMB. (2018) [24]. 113
- Figure 36 –(a) Survival of AG01522 cells irradiated with protons of $0.63 \text{ keV}/\mu\text{m}$ (upper line and points) and $2.45 \text{ keV}/\mu\text{m}$ (lower line and points). The lines are model full predictions, the points are experimental data taken from Marshall *et al.* (2016) [23]. (b) Survival of AG01522 cells irradiated with proton beams of $1.68 \text{ keV}/\mu\text{m}$ (upper line and points) and $7.50 \text{ keV}/\mu\text{m}$ (lower line and points). The lines are model full predictions (see the text for the details), the points are experimental data taken from Marshall *et al.* (2016) [23]. Source: Carante *et al.* PMB. (2018) [24]. 115

Figure 37	<p>–(a) Survival of U87 cells irradiated by protons of different LET (from top to bottom: 1.1, 7.0 and 18.0 $keV/\mu m$). The lines are full predictions, the points are experimental data taken from Chaudhary <i>et al.</i> (2014) [27]. (b) Survival of U87 cells irradiated by protons of different LET (from top to bottom: 4.0, 11.9 and 22.6 $keV/\mu m$). The lines are full predictions, the points are experimental data taken from Chaudhary <i>et al.</i> (2014) [27]. Source: Carante <i>et al.</i> PMB. (2018) [24].</p>	116
Figure 38	<p>–Reconstructed models of the GM06990 entire genome at 1 <i>Mbp</i> using (a) LorDG and (b) fLorDG with $\alpha = 1.0$. The fLorDG implementation produces structures similar to that produced by LorDG. Features like well-defined chromosome territories and the tendency of telomeres and/or elongated regions of large chromosomes to intrude into the center of the center of the genome, are also presented in the fLorDG models (c.f. figure 12 in [28]). Figure of panel (a) was taken from Trieu <i>et al.</i> Nuc. Acid Res. (2017) [28].</p>	117
Figure 39	<p>–(a) Base model of 1 <i>Mbp</i>. (b) 500 <i>kbp</i> Refined structured. The resolution of model (a) was doubled by using 500 <i>kbp</i> data in fLorDG-v3. The effect was an over-segregation of the chromosome territories (the topology was highly preseved) expressed as a radial shift of the chromosome with respect to the center of the nucleus. (c) 500 <i>kbp</i> Reference structure. A model at 500 <i>kbp</i> resolution was constructed directly from the corresponding HiC maps instead of applying the refinement to a 1 <i>Mbp</i> model. The structure is more compact but still, territories can be distinguished.</p>	119

List of Tables

Table 1	– Comparison of the main features of the A-, B- and Z-DNA conformations.	27
Table 2	– Summary of radiation doses and units.	40
Table 3	– Shape, size and number of chromosomes of the cell lines implemented in BIANCA.	61
Table 4	– Experimental ("Exper.") F- and G-ratios for human lymphocytes exposed to different doses of ^{137}Cs γ -rays compared with values calculated by the exponential model ("Exp.") or the Gaussian model ("Gaus."). The errors associated to the experimental values represent standard deviations calculated applying Poisson statistics to the raw numbers reported in the experimental paper ([18], both γ -ray data sets pooled).	87
Table 5	– Experimental ("Exper.") F- and G-ratios for AG1522 human fibroblasts exposed to different doses of ^{137}Cs γ -rays compared with values calculated by the Gaussian model ("Gaus.") or the exponential model ("Exp."). The errors associated to the experimental values represent standard deviations calculated applying Poisson statistics to the raw numbers reported in the experimental work ([20], "Day 0" data set).	92
Table 6	– F-ratio (total dicentrics/total centric rings in the whole dose range) and G-ratio (total interstitial deletions/total centric rings in the whole dose range) in human lymphocytes, compared with data taken from Bauchinger and Schmid [18].	98
Table 7	– F-ratio (total dicentrics/total centric rings in the whole dose range) and G-ratio (total interstitial deletions/total centric rings in the whole dose range) in AG1522 human fibroblasts, compared with data taken from Cornforth et al. [20].	101
Table 8	– Calculated aberration yields. The corresponding experimental values for comparison are reported in parenthesis.	105
Table 9	– Ensemble-averaged Spearman coefficients per chromosome of models as reconstructed by the two implementations. LorDG values are taken from Table 1 in [28], which represent the ensemble average over 50 structures (no standard errors provided). fLorDG values correspond to 50 structures as well (all standard deviations were smaller than 2%).	118
Table 10	– Agreement with data and dependence on LET and dose for the F-ratio and the G-ratio.	121

Contents

Introduction	21
1 Action of radiation on living systems	22
1.1 Ionizing radiation and its effects	22
1.2 DNA and chromatin structure	26
1.3 Chromosome aberrations	30
1.3.1 Observation and classification	31
1.3.2 DSB repair and aberration formation	34
1.3.3 Cell death and carcinogenesis	37
1.4 Radiobiological quantities of interest	39
2 Models of DNA/chromosome damage and cell death: analytical and Monte Carlo approaches	42
2.1 Analytical approaches	42
2.2 Monte Carlo approaches	45
2.2.1 DNA damage and repair	45
2.2.2 Chromosome and cell level response	50
3 BIANCA model: Former versions and version developed in this work	53
3.1 Overview	53
3.1.1 Model assumptions	53
3.1.2 Evolution of BIANCA	54
3.1.2.1 Early versions	54
3.1.2.2 The first BIANCA	56
3.1.2.3 BIANCA II	58
3.2 The version developed in this work	60
3.2.1 Geometry	60
3.2.2 Irradiation	64
3.2.2.1 Initial damage	65
3.2.2.2 Chromatin fragments	67
3.2.3 Interaction models	67
3.2.3.1 Probability functions	68
3.2.3.2 Formation of Aberrations	70
3.2.4 Scoring	71
3.2.4.1 Classification of the fragments	72
3.2.4.2 Chromosome aberration assessment	72

3.3	Chromatin modelling using HiC data	73
3.3.1	Restraint-based modelling with HiC data	74
3.3.2	The LorDG method	76
3.3.2.1	Algorithm	76
3.3.2.2	Input data	78
3.3.2.3	Testing the structures	79
4	Results	81
4.1	BIANCA upgrade	81
4.1.1	Technical aspects	82
4.1.2	Model improvements	83
4.2	Proximity effects in chromosome aberration induction by low-LET ionizing radiation	83
4.2.1	Human lymphocytes	84
4.2.1.1	Dose-response curves	84
4.2.1.2	Aberration ratios	86
4.2.2	Human fibroblasts	88
4.2.2.1	Dose-response curves	88
4.2.2.2	Aberration ratios	90
4.3	Dependence of proximity effects on radiation quality, cell type and dose . .	92
4.3.1	Human lymphocytes	93
4.3.1.1	Proton aberration yields	93
4.3.1.2	Alpha-particle aberration yields	94
4.3.1.3	Aberration ratios	97
4.3.2	Human fibroblasts	99
4.3.2.1	Alpha-particle aberration yields	99
4.3.2.2	Aberration ratios	101
4.4	Applications for hadrontherapy	102
4.4.1	Analysis of radiation-induced chromosome aberrations after alpha particle microbeam irradiation	103
4.4.1.1	Absorbed dose estimation	103
4.4.1.2	Calculation of aberration yields.	104
4.4.2	Application of BIANCA to cell survival data by protons, C-ions and He-ions at energies and doses used in hadrontherapy	106
4.4.2.1	Protons	106
4.4.2.2	Carbon ions	108
4.4.2.3	Helium ions	112
4.4.2.4	Full predictions of ion-survival based on a reference cell line	112
4.5	Including HiC-maps data in BIANCA	116
4.5.1	FORTTRAN implementation of LorDG	116

4.5.1.1	Contributions in fLorDG	118
4.5.2	Future prospects	120
	Conclusions	121
	List of Publications	124
	Bibliography	126

Introduction

When ionizing radiation impinges on living cells during the *G0/G1* phase of the cell cycle, it can produce a variety of chromosome aberrations, which mainly consist of large-scale incorrect rearrangement of chromosome fragments. This is an important end-point in radiobiology because it is related to both cell death and cell conversion to malignancy. Since the very beginning, it was clear that the involved fragments have to be created close in space, and investigating the ratios of specific aberration types can help characterizing such "proximity effects". In this context, biophysical modelling can contribute clarifying the features of proximity effects. Throughout the years, different models have been proposed to describe the formation of chromosome aberrations and their relation to cell death. To account for proximity effects, a number of functions describing the probability for two chromosome fragments (mis-)rejoin have been proposed. Typically, such functions depend on the initial distance between the two fragment free-ends. However, evaluating whether a certain function can model the rejoining process better than another one is not trivial, because in general different models have been implemented in different simulation codes and tested under different conditions in terms of considered cell types, dose values, radiation quality, etc.

The present thesis work tackles the topic of chromosome aberration (CA) formation from a biophysical perspective. During this PhD project, the BIANCA (BIophysical ANalysis of Cell death and chromosome Aberrations) model, which is implemented as a Monte-Carlo code, was upgraded in order to expand its capabilities and performance. The model is based on three main assumptions, linking the initial DNA cluster damage, dependent on both radiation quality and the cell line features, to the production of chromosome aberrations, which in turn can lead to clonogenic cell death. Previous model versions have successfully reproduced experimental results with regard to inter-chromosome exchanges. Nevertheless, a finer distinction of chromosome aberration products at the intra/inter-arm intra-chromosome level was necessary. Furthermore, going towards more refined computational models of CA formation implies not only to revisit the mechanisms that govern proximity effects, but also to aim at including sophisticated models of the DNA capable to describe the process of CA formation down to the molecular level. In this matter, a new scenario for chromatin 3D modelling has been opened with the advent of the so called HiC maps.

In chapter one, the action of ionizing radiation on living systems is discussed. It includes a description of the main physical mechanisms of interaction of radiation with

matter, basic concepts in radiobiology and chromatin organization, and a review on the chromosome aberrations topic. Chapter two deals with the biophysical modelling of chromosome aberrations. The basic components of the modelling problem are discussed and some of the different approaches that have been developed in recent years are summarized.

Chapter three is divided in three sections in order to present, first, an overview of the BIANCA model foundations, theoretical assumptions, and evolution of the model/code throughout the years. Next, a section describing the current implementation of the code follows. This section constitutes the core of the work carried out during this PhD thesis, as new functionalities were included and the whole architecture of the code was revisited in this project. Finally, a third section dealing with chromatin modelling based on HiC maps is introduced. An ongoing complementary project to this thesis aimed at improving the geometrical modelling in BIANCA is discussed, and a restraint-based optimization method called LorDG is presented.

Chapter four comprises five sections presenting the results of the activities carried out during this PhD thesis. The technical and model improvements of BIANCA are presented first. Then, the next two sections display the results of a systematic study on the formation of chromosome aberrations in human lymphocytes and fibroblasts. Two different proximity functions, added during this project, were tested against experimental data on a wide range of LET and dose values for both cell lines. A fourth section, summarizes the conclusions of two collaborations in which BIANCA was applied in the context of hadrontherapy. In collaboration with the BioQuaRT (Biologically Weighted Quantities in RadioTherapy) project, BIANCA served as a means to interpret experimental data on microbeam irradiation with α -particles, in terms of critical DNA damage. In the second collaboration, experimental survival data on V79, human fibroblast and U87 cells exposed to protons, C-ions and He-ions, was analyzed in terms of cluster lesion yields per micron, and a method to predict dose-response curves for ions was proposed. The last section of chapter four is about the current state of a project that started during this PhD, which aims at including HiC data sets in the chromatin modelling of BIANCA by means of the LorDG method mentioned above. Thus, the current fortran implementation and tests results of such method are presented.

1 Action of radiation on living systems

Radiobiology consists of the study of the effects of ionizing radiation (IR) in biological systems such as cells, biological tissues and living organisms. The smallest unit of all living organisms, capable of independent existence, is the cell. Cells contain both organic and inorganic materials and their structure is divided into two main components: the cytoplasm, in which metabolic functions are supported, and the nucleus, which contains all the genetic information (DNA). It is commonly accepted that the cell nucleus is the sensitive target when IR interacts with a living system [29]. Nevertheless, there exist other known effects that cannot be explained within this target theory based on the cell nucleus, for instance, genomic instability¹ [30] and bystander effects² [31]. The study of these phenomena is also very important for the understanding of radiation-induced damages and the development of corresponding study models [32, 33]. Nonetheless, the effects of IR on the DNA continue to play a major role in the radiobiological problem. Thus, it is still necessary to analyze the action of radiation down to the molecular (DNA) level.

1.1 Ionizing radiation and its effects

The radiation that is able to leave charged a given atom or molecule is considered *Ionizing Radiation*. Living organisms are continuously exposed to IRs from different sources as IR is, in fact, present everywhere coming from the outer space. There are several other mechanisms to produce it such as X-ray machines, cyclotrons, and nuclear reactors for instance. This radiation can be emitted by the decay of an unstable nucleus or by de-excitation of atoms and/or nuclei. According to its origin, it can be regarded as extraterrestrial (e.g. cosmic rays), NORM (Naturally Occurring Radiation Material such as Uranium, Thorium, and Radium) or artificial (e.g. Medical facilities and Nuclear Reactors, among others). In 2008 it was published a summary of public and worker exposures to different types of radiations by the UNSCEAR [34]. The fact that IRs can ionize the media that they traverse implies that they may cause injuries to cells. If the amount of energy deposited by a given radiation is enough to directly disrupt the structure of an atom or molecule, it is considered *directly ionizing radiation*. It may also happen that the solely ionizing particle is not able to induce a direct damage, but produces a number of secondary particles (e.g. secondary electrons) that induce chemical and/or biological damage. When this occurs, such IR is called *indirectly ionizing radiation*.

¹ High frequency of mutations within the genome due to defects in the checkpoint and/or repair machinery of a cell lineage, during mitosis or DNA damage events.

² A phenomenon in which unirradiated cells display biological damage due to signaling from nearby irradiated cells.

From a physical point of view, ionizing radiation can be classified into electromagnetic (EM) and particulate. When radiation traverses a material it may or may not interact with the atoms of the media. Biological effects can occur when some energy is deposited within the cell or tissue [35].

Photons constitute the EM radiation and they are characterized by their wavelength or frequency spectrum. They interact with matter through photoelectric effect, Compton scattering, and pair production processes mainly. The chance for one of these processes to happen, is determined by the cross section, which in turn depends on the characteristics of the target and the photon incident energy E . In photoelectric absorption, an X -ray or γ photon transfers all its energy to an inner-shell electron leading to a vacancy after ejection of the latter. Then, a fluorescence or Auger electron emission follows when the vacancy is occupied by other outer electron of the atom. The difference between the photon incident energy and the binding energy of the core electron corresponds to the kinetic energy of the ejected electron. The cross section of photoelectric absorption strongly depends on E and the Z number of the material [36]. For instance, in human tissues where $Z \simeq 7.4$ [37], this is the main mechanism of interaction up to $E \sim 30 - 50 keV$.

Compton scattering refers to the collision of a photon with an electron of the atom. During the collision, only a fraction of the photon energy is transferred to the electron leading to its ejection from the orbital [36]. The photon goes on with the remaining energy leaving behind the atom ionized. The emitted electron will, typically, interact with other atoms producing a cascade of ionizations that will result in the emission of secondary electrons, also known as δ -rays. In biological tissues, Compton scattering is the dominant mechanism at photon energies from $100 keV$ to $10 MeV$. This energy range is in the therapeutic radiation range.

Photons with energy above $1.022 MeV$ can also interact with the Coulomb field of an atomic nucleus, converting themselves into an electron-positron pair [36]. This process is called pair production. Positrons have a very short lifetime and they annihilate with an electron at the end of its range producing two new γ -rays of about $0.511 MeV$. Pair production can also happen in the presence of the Coulomb field of an atomic electron yielding to a triplet-production. In this process, the atomic electron is also ejected from its shell along with the electron-positron pair. The threshold energy for triplet-production is twice the threshold energy for pair-production.

In all the aforementioned cases, the number of ionizations produced by secondary electrons are much greater than those directly produced by the primary photons. Thus, EM radiation is considered as indirectly ionizing.

Particulate radiation refers to a beam of massive particles that is able to excite and ionize the atoms and molecules of a target. Such particles may or may not be charged. Additionally, charged particles can be classified as light and heavy particles [35]. For instance, electrons and positrons are regarded as light charged particles, while ions belong to the heavy charged particle category. Protons and α -particles are generally considered as light ions as well. The behavior of light particles is different than that of heavy particles of the same energy, in terms of the amount of energy that they deposit per unit path length. Such quantity is the so called *Linear Energy Transfer* (LET) or *electronic stopping power*. A semi-classical expression for this quantity was obtained by Bohr back in 1915 [38]. However, Bethe and Bloch derived a more sophisticated formulation accounting for quantum mechanical effects years later [39, 40]. The Bethe-Bloch formula depends on several parameters and quantities describing the absorbing material, the projectile and their interaction. In Eq. 1.1, ρ , Z , and A represent the absorbing material's density, atomic number, and atomic weight, respectively. The projectile velocity is v , its charge is z and W_{max} , the maximum energy transfer in a single collision. The Bethe-Bloch formula reads,

$$-\frac{dE}{dx} = 4\pi N_A r_e^2 m_e c^2 \rho \frac{Z}{A} \frac{z^2}{\beta^2} \left[\ln \left(\frac{W_{max}}{I} \right) - \beta^2 - \delta/2 - \frac{C}{Z} \right] \quad (1.1)$$

where N_A is the Avogadro's number, r_e is the classical electron radius, m_e is the mass of the electron, c is the speed of light in vacuum, and β represents the ratio v/c . I is the mean excitation potential of the absorbing material, δ is the density effect correction and C is the shell correction. An exhaustive description of the various terms and their meaning can be found for example in [41].

It is worth noticing that the energy loss depends on the inverse square of the velocity of the incident particle and on the square of the ion charge. On the contrary, there is no dependence on the projectile mass. At non-relativistic energies the stopping power is dominated by the $1/\beta^2$ term and decreases with increasing velocity, until a minimum is reached around $v = 0.96c$. As the energy further increases, dE/dx increases again due to the logarithmic term. The dependence of the energy loss on the energy of the projectile is of great importance for hadrontherapy, since charged particles deposit most of their energy just before the end of their path. The penetration of charged particles into a given medium will depend on their LET and, in general, is much higher for low LET radiation. Energetic ions show a well defined range, just after a very sharp peak in the energy deposition distribution as a function of depth (Bragg peak). Thus, ions offer advantages for cancer treatment over photon and electron beams, both from the dosimetric and radiobiological point of views.

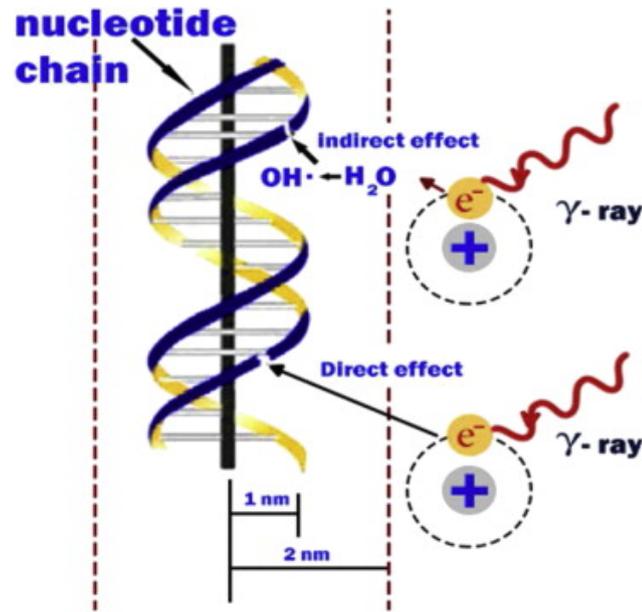


Figure 1 – Direct and Indirect action of ionizing radiation on the genetic material. The very breakage of a bond that splits the molecule and provokes chemical reactions, constitutes a direct action. The creation of free radicals after the hit, which attack the DNA inducing changes, represents an indirect action of radiation. Source: Desouky *et al.* J. Rad: Res. (2015) [1].

Neutrons are within the uncharged particle category and they deposit their energy into the medium through different processes that depend on their initial energy and the target mass. Such processes are, elastic scattering, inelastic scattering, neutron capture and fission [42]. The microscopic cross section is a quantity defined to represent the measure of the probabilities of each such reaction. As neutrons are not able for themselves to interact with the electronic structure of an atom, they are regarded as indirectly ionizing radiation.

When an atom or molecule of a biological sample is ionized or excited, it follows a series of chemical reactions between such an atom or molecule and the medium that surrounds it. The typical time scale for an ionized molecule to become a free radical is only about 10^{-15} to 10^{-10} s. In turn, free radicals may lead to cell damage by starting chain reactions on the neighbouring molecules. In lipids, for instance, these chain-reactions are related to damage in the cell membrane. In the case of direct damage, a biomolecule (BioM) hit by IR produces radicals that subsequently may undergo cross linking with other radicals. The chemical reaction path can be represented as,



where R^* and H^* stands for radicals and $R^* - R^*$ is the product of the cross linking.

This mechanism is the dominant effect when high LET radiation (see below) impacts on biological media.

On the other hand, the interaction of IR with water produces H_2O^+ and electronically excited water molecules. After radiolysis, such products decompose into $\cdot OH$ and $H\cdot$ radicals. Also, electrons can be produced from IR-water interactions. These electrons rapidly thermalize and are solvated by dielectric interactions becoming e_{aq}^- radicals. Finally, all these radicals, namely, $\cdot OH$, $H\cdot$, and e_{aq}^- , as being highly reactive with cells, DNA, and lipids, induce damages in the cell. This is the mechanism for indirect damage and is more frequent in the case of low LET irradiation. Figure 1 shows the mechanisms of direct and indirect damage schematically.

1.2 DNA and chromatin structure

DNA stands for *Deoxyribose Nucleic Acid* and consists of a long strand of nucleotides that form a helix. Nucleotides are made up of a pair of chemical bases attached to sugar-phosphate groups at the ends. The chemical bases are formed by Guanine-Cytosine (GC) or Adenine-Thymine (AT) pairs connected by hydrogen bonds. This pairing results in base pairs with similar shapes and sizes. The backbone of the helix is bounded by covalent bonds between the sugar-phosphate groups at the extremes of the nucleotides. The instructions that control cell function are coded in the DNA. Genes are long genetic material sequences that contain specific information for the organism functioning, such as those used for protein synthesis. Genes vary in extent and content and are the fundamental objects of study of *genetics* [46].

Back in the early 40's it was hypothesized that the DNA was the carrier of the genetic information. This hypothesis was confirmed years later when Watson and Crick showed how the genetic information can be stored in and copied from the DNA [47]. They discovered that Guanine pairs with Cytosine and Adenine with Thymine in such a way that the helical structure can store arbitrarily long sequences of different nucleotides despite the shape differences between those purines and pyrimidines. The Watson and Crick work was inspired by previous works about the composition of the DNA, mainly those of Visher, E. Chargaff [48, 49], Maurice Wilkins [50] and Rosalind Franklin and R. G. Gosling [51]³. They also discovered that the backbone of the helix runs in opposite directions and that this mechanism makes it possible the successful replication of DNA.

Despite having the same composition, not all the DNA molecules have the same chemical structure. In human cells, there are three main configurations of DNA, namely,

³ "However, the data which really helped us to obtain the structure was mainly obtained by Rosalind Franklin, who died a few years ago." Crick's letter to Monod. December 31th 1961 [52]

the B-, A- and Z-DNA [53]. Each of them have different biological functions. For instance, B-DNA is the primary (canonical) form of DNA that encodes the genetic information due to its flexibility. This was the DNA configuration described by Watson and Crick in their famous paper. In B-DNA, the base pairs (bp) are nearly perpendicular to the helix axis, the distance between two adjacent nucleotides is about 0.34 nm and the twisting angle is 36° , which results in 10 bps per turn. The helix diameter is 2.4 nm .

A-DNA can be observed when dehydrated DNA is put under X-rays, also, in some RNA-DNA hybrids and doubled stranded RNA. In both A- and B-DNA the twisting of the helix is right handed and the bases follow the Watson-Crick pattern. However, in the A conformation, the base pairs are no longer perpendicular to the helix axis but are tilted about 19° . Thus, A-DNA is wider and shorter than B-DNA. A-DNA seems to be related with DNA-drug as well as DNA-protein interactions. In addition, it has been looked up to its role in genome structure and function [54].

A third form of DNA was described by Alexander Rich and coworkers when they were studying the $d(GC)_3$ hexamer. They concluded that this hexamer also forms a helical structure but, unlike the other conformations, it is left-handed and the sugar-phosphate backbone shows a zigzag pattern. Hence the name, Z-DNA for this configuration [55, 56]. Despite it is observed in certain physiological cellular processes before decaying to the B form [57], the biological function of the Z-DNA is still unknown. Nevertheless, it has been suggested to play a role in the pathology of poxviruses [58]. Table 1 displays a summary of some features of the aforementioned DNA conformations and how they compare to each other.

Table 1 – Comparison of the main features of the A-, B- and Z-DNA conformations.

	DNA conformation		
	A	B	Z
Shape	Broadest	Intermediate	Narrowest
Rise per Base Pairs	2.3\AA	3.4\AA	3.8\AA
Helix Diameter	25.5\AA	23.7\AA	18.4\AA
Screw sense	Right-handed	Right-handed	Left-handed
Glycosidic bond	<i>anti</i>	<i>anti</i>	alternating <i>anti</i> and <i>syn</i>
Base pairs per helix turn	10.7	12	10
Pitch per Helix turn	25.3\AA	35.4\AA	45.6\AA
Base pair tilt (Helix axis)	19°	1°	9°
Major Groove	Narrow and very deep	Wide and quite width	Flat
Minor Groove	Very broad and swallow	Narrow and quite deep	Very narrow and deep

By the XIX century, chromosomes had been discovered in the nucleus of eukaryotic cells. Their structure and function were better understood once the DNA structure was unveiled. Chromosomes are long strands of DNA with associated proteins that bend and

pack the genetic material in order to fit it into the nucleus dimension. This *DNA + packing proteins* complex is called *Chromatin*. As mentioned before, long sequences of DNA with specific information are called genes, so every chromosome carries a determined number of genes. Most plant and animal cells are diploid, which means that chromosomes are present in pairs within the nucleus. Two chromosomes with the same array of genes are called homologous.

Nonetheless, genes represent a small fraction of the DNA content present in mammalian cells (less than 5%). This suggests that the function of the DNA primary structure, i.e. the nucleotide sequence, may be related to controlling the spatial organization of the chromatin as well [59]. Such spatial organization and its dynamics play a major role in the regulation of biological processes like gene expression, replication, recombination and repair [60, 61]. X-rays analyses of eukaryotic DNA reveal that the genetic material is highly packaged into basic chromatin units called nucleosomes [62, 63, 64]. The structure of the nucleosomes consists of two turns of DNA (containing $\sim 150\text{bps}$) wrapped around an octamer of histone proteins [65, 66]. Then, nucleosomes link to each other through shorter DNA fragments in a "bead-on-a-string" manner to subsequently arrange themselves in higher order structures [67]. This hierarchical arrangement includes the 30nm chromatin fiber, chromatin loops, intermediate segregation states and finally, the chromosomes. The mechanisms that allow such condensation involve DNA-protein interactions that depend on the local mechanical properties and composition of the DNA. For instance, a number of papers have discussed the influence of the periodicity of dinucleotides like AA/TT in the formation and positioning of nucleosomes [68, 69, 70]. Additionally, Audit *et al.* used the wavelet transform technique to compare the DNA text with bending profiles based on nucleosome positioning data. In their works, they found that there is a long range correlation (LRC) between DNA structural elements involved in the packaging of the chromatin in eukaryotic and, to some extent, archaeal genome at the $100\text{--}200\text{bp}$ range. They concluded that these LRCs are a signature of nucleosomal structure [71, 72, 73]. Nevertheless, the precise way in which the DNA primary structure contributes to chromatin organization continues to be a hot topic in molecular biology.

At the kilo base pair (kbp) scale, chromatin organizes in the so called 30nm fiber, which consists of a super-helix state with 6 nucleosomes ($\sim 1.2\text{kpbs}$) per turn. Then, the chromatin fiber forms loops and higher order structures. With regard to chromatin structure and composition, it has been observed that anchor points of chromatin loop domains are constituted by $\sim 1\text{kbp}$ AT-rich sequence patterns [74, 75]. Furthermore, other experimental works have linked the gene position and transcriptional activity with the microscopic organization of the chromatin, including the formation of loops [76, 77, 78]. Again, this speaks in favor of an influence of the primary structure of DNA in the packaging of chromatin.

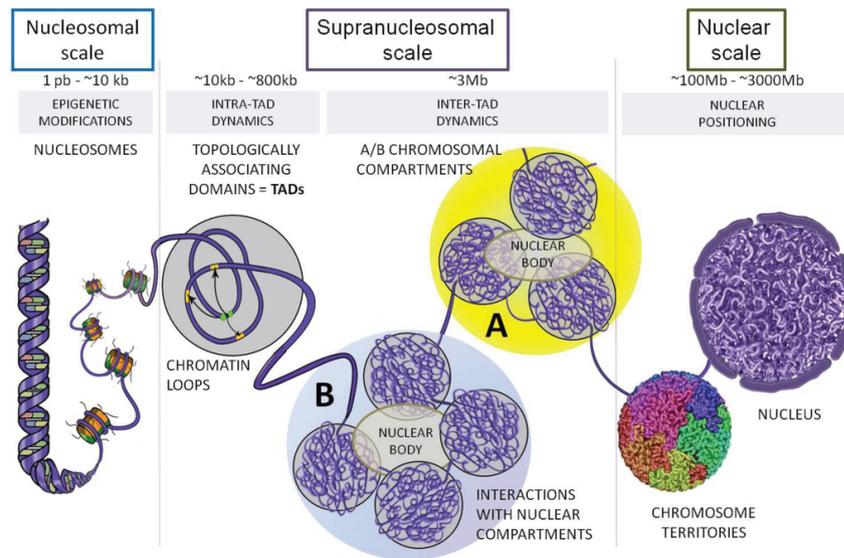


Figure 2 – A scheme of the different organizational levels of the chromatin. At the nucleosomal scale, the DNA follows scale-invariant principles along the sequence and shows long range correlations in the primary structure that are related to nucleosome formation and positioning. Before the mega base pair (Mbp) regime, the chromatin self-organizes in globular states that are in agreement with the fractal globule. From chromatin loops to TADs and compartments, there appear correlations between chromatin states, accessibility and composition (e.g. gene content or presence of promoters/enhancers), and the spatial distribution and contact profile of the chromatin. At the chromosome scale (tens of Mbps and higher), the distinction between territories is prominent. Chromosomes are organized in well-defined domains, still, some regions intermingle with other domains. Source: Vuthy et al. *Genes* (2015) [3].

Larger scales ranging from tens to thousands of kbps show intermediate compaction and segregation states of the chromatin before the identification as a chromosome. The advent of Chromosome Conformation Capture (3C) technologies enabled to study multiple loci interactions in the genome⁴. This led to the creation of interaction matrices of *all vs. all* genome regions by means of the Hi-C maps [79]. Statistical analysis of intrachromosomal interactions revealed that chromatin segregates in two types of compartments (labeled A and B) that are related to open and close states of the chromatin. In this way, folding and unfolding, which is an essential feature of chromatin dynamics, appears to be strongly correlated with the disposition of these compartments. For instance, Lieberman *et al.* [79] concluded that the A compartment is more closely associated with open, accessible, actively transcribed chromatin.

Another remarkable finding was that the interaction probability I as a function of the genomic distance s , follows a power-law scaling between $\sim 500\text{kbp}$ and $\sim 7\text{Mbp}$. Power-law dependencies of this kind are typical of polymer systems and in fact, it has

⁴ In the context of Chromosome Conformation Capture techniques, an *interaction* means a contact between loci. The technology is based on the crosslinking of cells in vivo in order to "capture" all the contacts between different regions of the genome, in close proximity.

been proposed that the chromatin can be modeled as a polymer in equilibrium [80, 81]. Notwithstanding, Hi-C data shows a scaling factor dramatically different to that predicted by the equilibrium state. In the equilibrium globule, an $I(s) \sim s^{-3/2}$ is expected while in the data, a $s^{-1.08}$ trend was found. In contrast, the model proposed by Grosberg *et al.*, called the "fractal globule" [82, 83], agrees more with the Hi-C data. The fractal globule can be described as a long-lived, non-equilibrium polymer conformation. It is highly compact and unknotted, which facilitates the folding and unfolding of regions of the polymer. In the fractal globule, the predicted scaling factor is -1 , which is very close to the value extracted from the data. Additionally, Mateos-Langerak *et al.* [81] reported a $s^{1/3}$ scaling of the three dimensional distance between loci, using 3D-FISH. This is consistent with the fractal globule that predicts the same scaling factor, unlike the equilibrium state where $s^{1/2}$ is predicted. Thus, the structure of chromatin at the Mbp scale is consistent with the fractal globule.

Chromosomes are the highest organizational level of chromatin. It is observed that they occupy well-defined territories (CTs) [84] and their distribution within the nucleus is highly non-random [85]. For instance, in lymphocytes, gene-poor chromosomes tend to reside in the predominantly hetero-chromatic nuclear periphery, while gene-rich CTs accommodate in the eu-chromatic region at the nucleus core. Interestingly, the transcriptional activity has been correlated with the gene position relative to the periphery and the bulk of CTs [86, 87]. Many factors determine the arrangement of chromosome domains within the cell nucleus; the DNA primary structure seems to play a central role at the smaller scales (e.g. DNA wrapping, nucleosome formation, loop anchoring). However, secondary and tertiary structures like the so called Topologically Associated Domains (TADs: clusters of chromatin loops with statistically significant preferred contact and border zones rich in CTCF protein.) and compartments, may also play an important role in the shaping and functioning of the chromosomes [79, 59, 88]. Figure 2 depicts our current understanding of the chromatin organizational levels. From the double helix to the chromosome territories, the relationship between the chromatin structure and its functioning is yet to be determined.

1.3 Chromosome aberrations

Nowadays it is widely accepted that Double-Strand Breaks (DSBs) in the DNA constitute the critical damage responsible for the formation of chromosome aberrations (CAs) [6, 89]. When lesions of this type are induced on the genetic material (due to IR for that matters), a number of repair mechanisms are activated. The failure of such mechanisms to reconstitute (eurepair) the created chromatin fragments, leads to different types of chromosome aberrations (see Figure 3). For instance, an un-repaired DSB appears

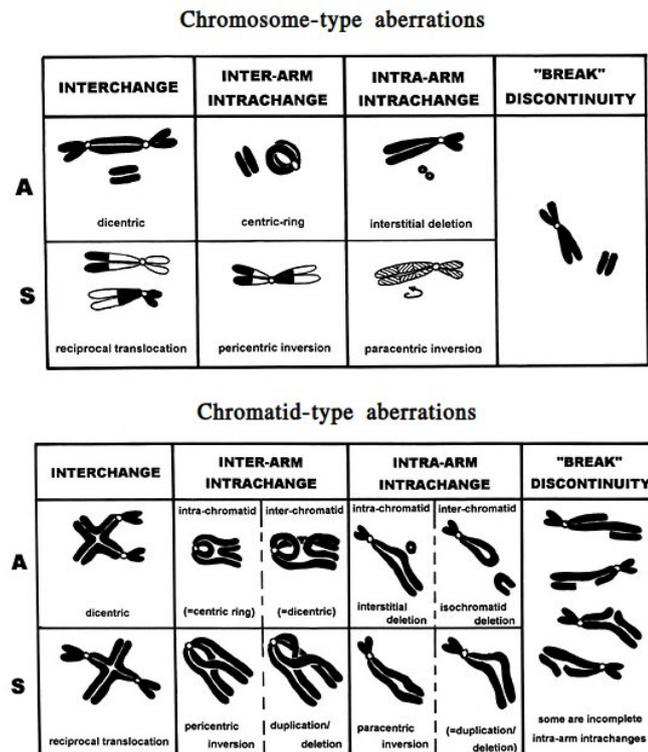


Figure 3 – Two-break types of chromosome and chromatid aberrations. exchanges can be Symmetric (S) or Asymmetric (A) depending on whether an acentric fragment is formed or not, provided that the exchange is complete (that is, all four ends rejoin). Additionally, the exchanges may occur between different chromosomes (*interchanges*), or between fragments of the same chromosome (*intra- and inter-arm intrachanges*). For chromatid-type aberrations, the intrachanges can be further classified as *intra-* or *inter-chromatid* depending on how the sister-chromatids mix. Image adapted from: Savage, J. Med. Genet. (1976) [4].

as a terminal deletion at mitosis and implies the loss of genetic material. It can also happen that two DSB interact (misrepair) forming an exchange that may lead to a mutation or cell inactivation. In any case, the fate of the cell will depend on how these processes take place.

1.3.1 Observation and classification

When DSBs are induced during the G_0/G_1 (pre-replication) phase of the cell cycle⁵, the aberrations eventually formed at this stage are referred to as *chromosomal aberrations*. If, on the other hand, such lesions and their consequences occur during G_2/S , the corresponding aberrations are termed *chromatid aberrations*. Such distinction is motivated by

⁵ There are two well defined periods in the cell proliferation cycle, namely, *mitosis* (M), where division takes place, and DNA *synthesis* (S), where DNA replication occurs. The gap between M and S is called G1 phase and the gap between S and M, G2 phase. The G_0 phase can be regarded as a resting phase in which the cell is outside of the replicative cycle due to external factors or as part of its own genetic programming.

the observation of chromosomes after the action of a clastogen⁶, at the first post-induction division when the cell has not yet undergone any subsequent division or chromosome duplication that could lead to selection or modification of some aberrations. Since chromosomes in metaphase (when observations are usually carried out) consist of two sister-chromatids, *chromosome aberrations* appear to affect both sister-chromatids while, *chromatid aberrations* involve only one, at any given locus. In this way, the chromosome products recovered at metaphase reflect the duplication state of the chromatin at the moment when the breakages were induced. The aberrations scored before replication/mitosis are thus called *primary aberrations*. In contrast, *secondary* or *derived aberrations* are those detected after several cell divisions. However, their origin is more complicated to track [4] since most of the transmissible chromatid-type aberrations will look like chromosome-type aberrations after duplication. Hence, the presence of chromatid-type aberrations in later generations of cells could be regarded as an indicator of ongoing genomic instabilities.

Chromosome rearrangements are generally visible under the microscope only when chromatin is in its condensed state. Thus, in order to observe CAs at any time after the initial damage, the chromatin needs to be forced to condense before mitosis by means of Premature Chromosome Condensation (PCC) techniques. This can be accomplished by cell fusion with mitotic cells in order to expose them to activated p34cdc2/cyclinB complex (a chromosome condensation promoter [90]), or either by treatment with Calyculin A, a toxin extracted from marine sponge *Discodermia Calyx*, that inhibits protein serine/threonine phosphatase (involved in the cell cycle progression)[91]. PCC with Calyculin A has proven to be a powerful tool for biodosimetric analysis as it allows to increase the number of spreads while avoids problems related to cell-cycle alterations and does not increase significantly the background aberrations [92].

Visualization of the structural changes in chromosomes, either during metaphase or at interphase with PCC, can be achieved through one of the two staining techniques available: *GIEMSA* staining and *FISH* (Fluorescence In Situ Hybridization). In *GIEMSA* staining, all chromosomes are dyed with the same color accomplishing high contrast with respect to the cytoplasm, and the identification of CAs is based on the number of centromeres and the shape of the final chromosome products. On the other hand, *FISH* allows for selective painting of one or more pairs of homologous chromosomes, which enables identification of a wider spectrum of aberrations with respect to *GIEMSA* staining, as will be explained.

Chromosome aberrations result from the interaction (mis/un-rejoin) of chromatin breaks. According to where those breaks are located and how they interact, aberrations can be roughly divided into *exchanges*, in which two or more breaks interact, and *discon-*

⁶ A mutagenic agent able to disrupt or induce breaks on a chromosome (e.g. ionizing radiation).

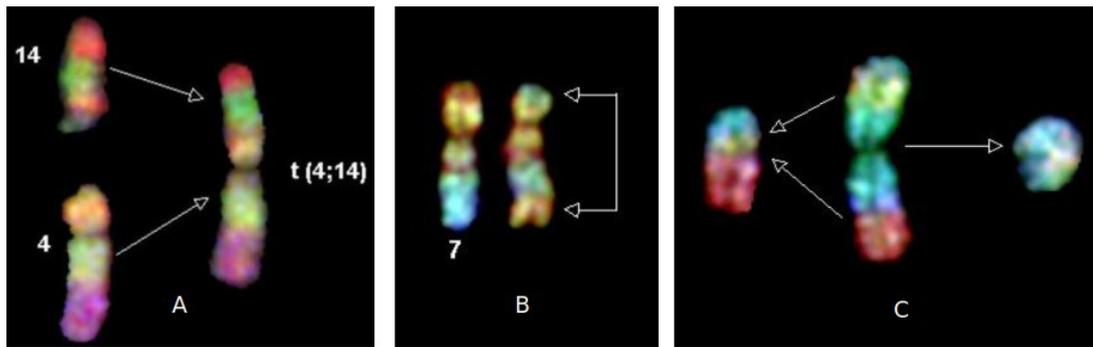


Figure 4 – Visualization of chromosome aberrations with cross-species color banding (Rx-FISH). (A) Translocation. (B) Symmetrical inter-arm exchange. (C) Asymmetrical inter-arm intrachange. Image adapted from: M. Durante *et al.* *Mutat. Res.* (2013) [5].

tinuities or *deletions*, in which a break remains open. Furthermore, when only two breaks and up to two chromosomes are involved in the formation of an exchange, it is called a *simple exchange*. Within the "simple exchange" category there are the *interchanges*, *inter-arm intrachanges* and the *intra-arm intrachanges* subcategories, which correspond to two breaks situated in two different chromosomes, in different arms of the same chromosome or in the same arm of a chromosome, respectively. Figure 3 shows some examples of the aberrations that arise from the interaction of two breaks. When the four ends of the two breaks interact in such a way that an acentric fragment is always created, the rejoining process is *Asymmetrical* (A). If, on the other hand, an acentric fragment is never formed unless the exchange is incomplete (one of the ends do not rejoin at all), the rejoining is *Symmetrical* (S).

The scoring of CAs with GIEMSA staining allows to identify only a portion of these categories since the chromosomes are colored all the same. For instance, inversions and small deletions would appear as normal chromosomes under the microscope. Additionally, the so called "complex exchanges", which by definition involve more than two breaks and at least two chromosomes, are also non-distinguishable with GIEMSA staining [93]. Nonetheless, this technique is useful to score other aberrations such as *dicentric*s, *rings* and *deletions*, since their scoring is rather based on the shape and number of centromeres of the resultant chromosomes. In contrast, FISH painting⁷ can provide more detailed information about the morphology of the rejoined chromatin fragments, and thus, it enables to distinguish aberrations that otherwise would be regarded as normal chromosomes (e.g. reciprocal translocations, insertions, etc.) [94, 95]. With FISH, aberrations would appear as multicolored "chromosomes" made up of pieces of other chromosomes (interchanges) or with altered color patterns (intrachanges) as illustrated in Figure 4.

⁷ Not only the standard technique but also its variations: Spectral karyotyping (SKY), multiplex FISH (m-FISH) and cross-species color banding FISH (Rx-FISH).

1.3.2 DSB repair and aberration formation

Although DSBs are identified as critical lesions leading to CA, the dose response curve (DRC) for cell death or CA shows, in general, a non-linear trend that differs from the prominent linearity of the DRC for DSB [96]. In this way, the relationship between CA formation and DNA damage is not directly linked to the induction but to the repair of the DSBs. There are at least three DSB repair mechanisms available in eukaryotic cells, namely, Homologous Recombination (HR), Single-Strand Annealing (SSA) and Non-Homologous End Joining (NHEJ) [97].

HR is highly accurate at restoring the original sequence at the break but is available only during the S and G2 phases of the cell cycle [98]. SSA is activated when a DSB occurs between two repeated sequences in such a way that two complementary strands can be annealed. This annealed intermediate can be processed by digesting away the single stranded tails and filling in the gaps. Both HR and SSA require regions of sequence homology and it has been shown that both repair pathways involve the products of the XRCC2 and XRCC3 genes [99, 100, 101]. This gene family is homologous to the *Saccharomyces Cerevisiae* Rad52 gene group responsible for HR in yeast (highly proficient in HR). In this way, it has been elucidated that mammalian cells may use HR [102, 103, 104] more than was suspected in the past [97]. On the other hand, NHEJ is the predominant repair pathway along the cell cycle and does not require extensive homology between DNA termini in order to bind them. Instead, in NHEJ, the DNA ends bind with heterodimers of the Ku70 and Ku80 proteins, and the catalytic subunit (DNA-PKcs) of the DNA-dependent Protein Kinase (DNA-PK) is activated. Then, the rejoining process is completed by the DNA ligase IV/XRCC4 complex. In general, NHEJ is considered error-prone since even when attempting to eu-repair the fragments, information can be lost [105].

In principle, all the repair mechanisms may lead to any kind of exchange. Nonetheless, HR can be triggered by one single DSB while NHEJ and SSA will require at least two DSB for an exchange (see Figure 5). Complex exchanges (those involving three or more DSBs and at least two chromosomes) can be formed by HR, SSA or NHEJ as well. The exact contribution of each repair pathway to a particular type of CA remains unclear as all the mechanisms compete actively for DSB processing. In any case, the resulting products may represent both small (up to the kbp scale) and large (visible with microscope) scale alterations in the DNA. Thus, chromosome aberrations are not the product of a specific cellular process but the consequence of the DSB repair dynamics [106].

DSB repair follows both slow and fast kinetics depending on the particular repair path and the stage of the cell cycle [107, 108]. In G1 phase, CAs are formed with the

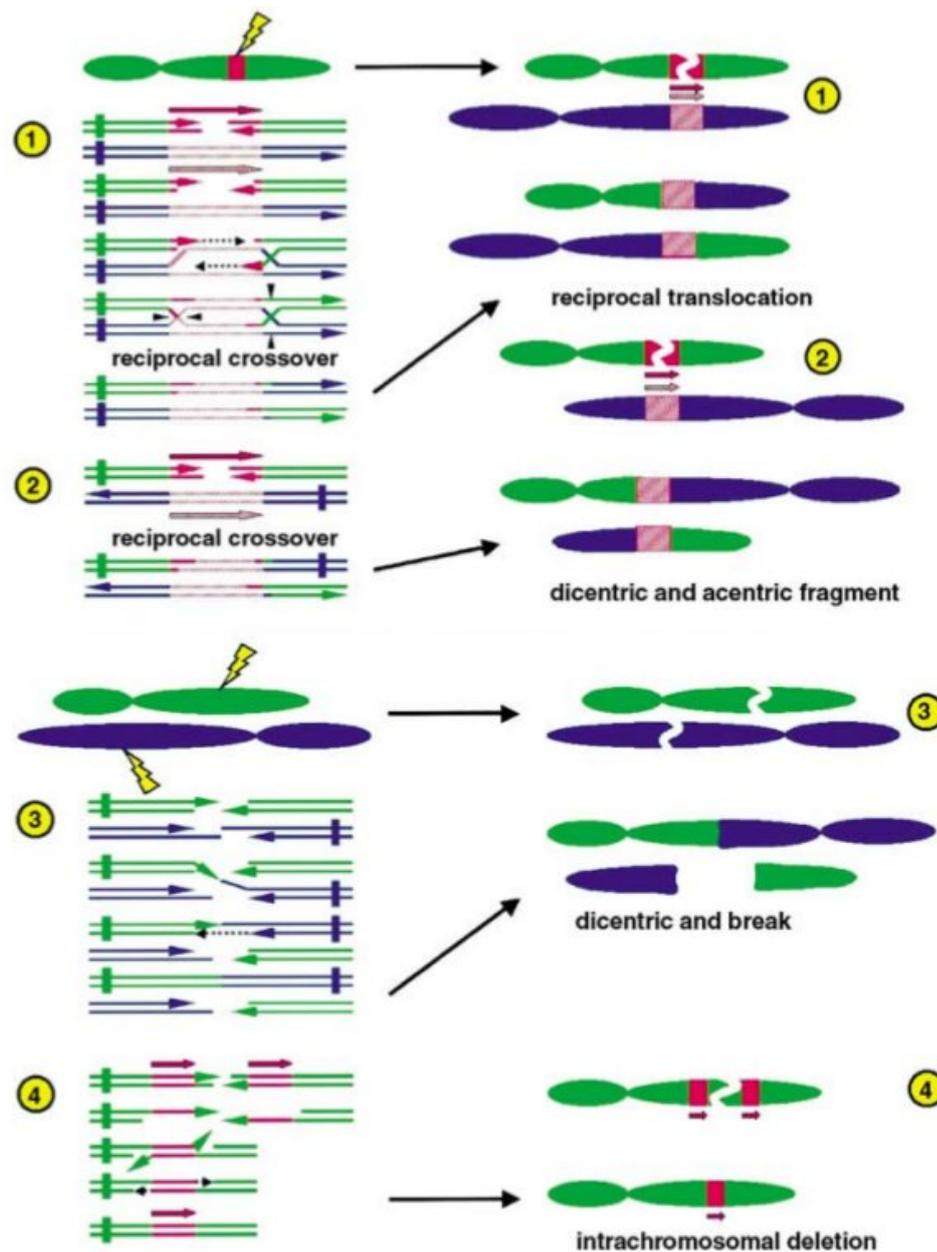


Figure 5 – Diagrammatic representation of the formation of some CAs by HR, SSA and NHEJ involving up to two DSBs and chromosomes. Red boxes and thunders represent DSBs. The green and blue horizontal lines represent chromosomes and their telomeres (vertical box). Pink arrows on top of a chromosome intend the orientation of the strand (needed by HR and SSA as they require sequence homology). (1-2) When HR occur between regions of sequence homology but located at different chromosomes (Ectopic HR), reciprocal translocations and dicentrics can arise along with acentric fragments. (3) In NHEJ, at least two DSB are needed to form an exchange. A dicentric is formed when two centromere-containing fragments are joined after being processed by the DNA-PK complex. One or two acentric fragments can also accompany the exchange, resulting in a complete or incomplete exchange respectively. (4) an example of an intra-chromosomal deletion as a consequence of SSA when the DSBs lie in an extensive region of sequence homology. Figure modified from G. Obe *et al.* *Mutat. res.* (2002) [6].

same kinetics as DSB repair, that is, fast kinetics due to NHEJ up to few hours after exposure and then slow kinetics involving alternative NHEJ paths as will be explained in brief [109, 110, 7]. On the other hand, in ref. [8], the formation of CA after γ -irradiation of G2 human fibroblasts was analyzed, showing that at this stage, all exchanges form very quickly (presumably by NHEJ). Furthermore, since no chromatid-type exchanges are observed many hours after exposure in G2 cells, it is believed that after the first stage of DSB processing, the damages are repaired via the slow error-free HR mechanism [111].

There are alternative forms of NHEJ to backup when the cell is deficient in any of the proteins and/or complexes needed for the "canonical" NHEJ (C-NHEJ) [112, 113, 114]. For instance, Yan et al. found an alternative path for rejoining in XRCC4- and LigaseIV-deficient mouse B cells, by studying the role of C-NHEJ in the context class-switch recombination⁸ [115]. Furthermore, Miladenov and Iliakis [114] described the enzymatic set-up of the so-called B-NHEJ as a true backup mechanism for C-NHEJ. They showed that this type of NHEJ does not depend on the DNA-PK complex but instead, on the LigaseIII/XRCC1 complex, and is regulated by the PARP1 polymerase. Additionally, it has been investigated the role of the Carboxy-terminal Interacting Protein (CtIP), a DNA end-resection factor, in the formation of translocations [116]. CtIP was originally identified by its participation in HR during G2 [117]. However, resection of the DNA breaks by CtIP may promote error-prone microhomology-mediated end-joining (MMEJ) during G1, which can be regarded as a form of B-NHEJ [116, 118].

Still, canonical NHEJ appears to be faster than the alternative NHEJ (B-NHEJ) pathways which show slower kinematics [119]. In this way, B-NHEJ mechanisms can not account for the formation of both fast and slow CA. Nonetheless, they may account for a portion of the CAs for example at higher doses when C-NHEJ is overwhelmed by the number of DSB sites and/or HR is not available. In any case, all the NHEJ routes contribute to the formation of CAs, despite that C-NHEJ acts quickly and that could be interpreted as lower chances to result in misrepairs (due to spatial proximity of the original DNA ends of a break). Thus, the scenario for the formation of chromosome aberrations is composed by different actors that account for different portions of the CA yield. C-NHEJ seems to be responsible for the fast component of the CA yield across the cell cycle, while B-NHEJ accounts for the slow kinetics in G1. On the other hand, if damages are induced during the S/G2 phase, still C-NHEJ is the first repair mechanism to process the DSBs due to its robust kinetics, but it may lead to early CAs. The remaining DSBs are, then, processed by HR (leading to restitution) but not by B-NHEJ at this stage. (see Figure 6).

⁸ A biological process that implies the break and rejoin of portions of DNA in order to change the production of Immunoglobulin from one type to another one.

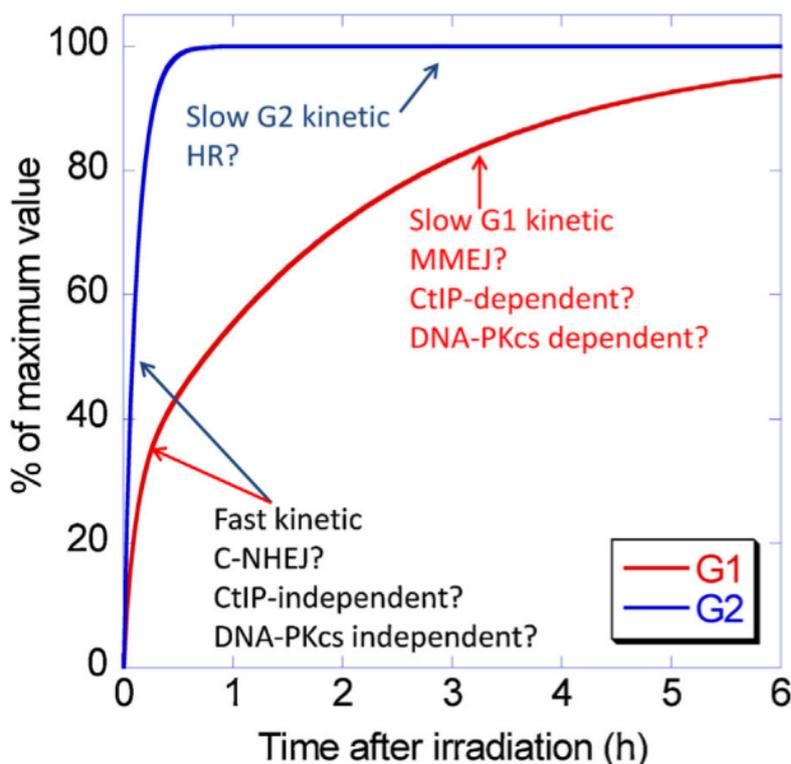


Figure 6 – Kinetics of the formation of exchanges in human cells after γ irradiation. Durante *et al.* [7] and Gotoh *et al.* [8] studied the formation of chromosome exchanges in G1 and G2 phase, respectively. They concluded that in both phases, each type of chromosome and/or chromatid exchange follow different kinematics so there is a contribution from fast and slow components to the total yield. Source: M. Durante *et al.* *Mutat. Res.* (2013) [5].

1.3.3 Cell death and carcinogenesis

Despite the vast amount of aberration categories that are identifiable at the first metaphase (or during interphase with PCC), only a fraction of such aberrations are transmissible. This means that only a portion of the primary aberrations can be recovered in future cell generations. Some of those would be changed significantly due to the influence of the cell cycle. For example, chromatid-type aberrations would be transformed into chromosome-like aberrations. Also, small deletions located at the distal regions of the chromosome could be processed by telomerase proteins leading to capping of the end by telomere sequences [120, 121].

The hypothesis that chromosome aberrations could be responsible for neoplasia⁹ was proposed at the beginning of the XX century [122]. However, the fact that cytogenetic techniques at that time were not able to identify such chromosomal changes, made it difficult that the hypothesis was accepted. The improvements in cytogenetics and molecular biology allowed the observation of recurring chromosomal aberrations, which in turn identified a large number of previously known and new proto-oncogenes. Several studies

⁹ Uncontrolled proliferation of the cells in a tissue.

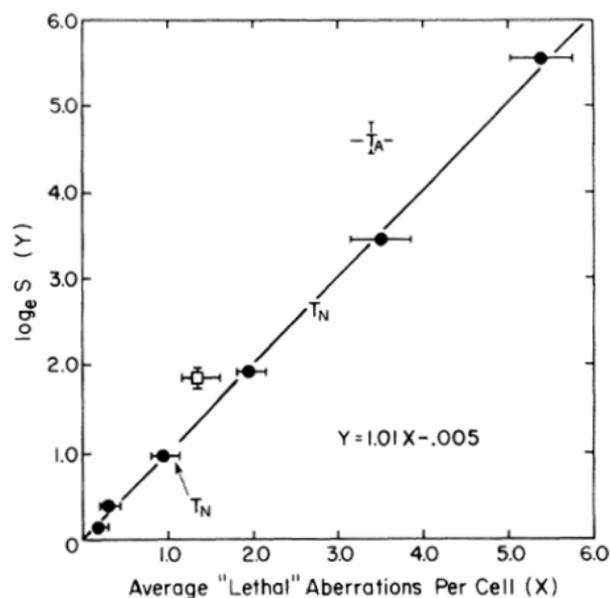


Figure 7 – Relationship between the average number of "lethal" aberrations (dicentrics + rings + large deletions) and the natural logarithm of the surviving fraction of cells. Source: Cornforth and Bedford, *J. Rad. Res.* (1987) [9].

showed that such genes are oncogenic, thus confirming the pivotal role of chromosomal aberrations in tumour development [123].

Typically, translocations, inversions and deletions are more likely to be transmitted to the cell progeny. In the case of deletions, the loss of part of the DNA sequence, while it may be compatible with the cell survival, it may also imply the loss of a tumour-suppressor gene. On the other hand, translocations and inversions can result either in the activation of a proto-oncogene (because the gene encoding for a T-cell receptor or an immunoglobulin protein comes to lie near that proto-oncogene), or the chromosome breaks occur within a gene on each involved chromosome, thus creating a fusion gene that encodes a "chimaeric protein". The c-MYC gene translocation in Burkitt's lymphoma is a paradigm of the situation in which a proto-oncogene is juxtaposed to an immunoglobulin gene, thus activating the oncogene. On the other hand, the fusion between the BCR and c-ABL genes from chromosomes 9 and 22, which leads to the so-called "Philadelphia chromosome" in Chronic Myelogenous Leukaemia, typifies the scenario in which breakage on each chromosome occurs within gene introns, producing fusion genes that subsequently lead to the expression of a fusion protein.

Radiation-induced CAs can lead to cell death as well as to the long term mutations discussed above. The most common form of radiation-induced cell death is the so-called "mitotic death", which occurs when a cell dies in attempting to divide because of damaged chromosomes. Death may occur in the first or in a later division following irradiation. Many authors reported a close quantitative relationship between cell killing and

the induction of asymmetric exchange-type chromosome aberrations such as dicentric and rings. This can be explained by the fact that during mitosis, the migration of sister chromatids towards the two opposite cell "poles" is governed by the centromeres. If a chromosome carries more than one centromere, as is the case for dicentric, the cell has a high probability to fail duplication. The linear-quadratic relationship that is characteristic of chromosome aberration induction during acute irradiations is thus carried over to the typical cell survival curve, which according to the linear-quadratic model is expressed by,

$$S = \exp(-\alpha D - \beta D^2) \quad (1.4)$$

where D is the absorbed dose, S is the fraction of surviving cells, and the parameters α and β depend on various factors including the radiation LET.

One of the most important studies relating the chromosome aberration yield with the cell killing was performed by Cornforth and Bedford [9]. They found a one-to-one correspondence between the yield of dicentric, rings and large deletions (regarded as lethal aberrations) and the negative of the natural logarithm of the surviving fraction ($-\ln S$). Such study was carried out over an AG1522 normal human fibroblast cell line irradiated with X-rays up to 6 Gy (see Figure 7). Furthermore, an excellent correlation was found between the fraction of surviving cells and the fraction of cells without visible aberrations. Additionally, an epidemiological study carried out over data coming from Italian and Nordic cohorts for various types of DNA damage, has suggested that the frequency of CAs in peripheral blood lymphocytes is a relevant bio-marker for cancer risk in humans, reflecting either early biological effects of genotoxic carcinogens or individual cancer susceptibility [124].

1.4 Radiobiological quantities of interest

Absorbed Dose, Equivalent Dose, LET and RBE.

The energy deposited locally in the medium per unit mass is called the *Absorbed Dose*. The international unit for the absorbed dose is the Gray (Gy), which equivalent to J/kg. Since, for a given dose, not all radiations induce the same biological effect, the quantity *Equivalent Dose* is introduced in radiation protection, to compare the biological effectiveness of different radiation qualities. The equivalent dose H on a tissue or organ T , is defined as

$$H_T = D * W_R \quad (1.5)$$

where D is the absorbed dose and W_R is a weighting factor that depends on the type and energy of the incident radiation R ¹⁰. The equivalent dose is usually expressed in Sievert (Sv). Table 2 shows other units used in radiobiology and their equivalences [44].

Table 2 – Summary of radiation doses and units.

Dose	SI Unit	Old Unit	Conversion Factor
Absorbed Dose	gray (Gy)	rad	100 rad=1 Gy
Equivalent Dose	sievert (Sv)	rem	100 rem=1 Sv

IRs interact with matter through various processes that result in depositions of energy along their path. The energy loss rate in a particular material depends not only on the energy and type of radiation but also on the density of the medium as shown in eq. 1.1. The LET is defined as the expected value of the energy transferred by the projectile to the medium per unit path length. This quantity accounts for the energy transferred due to electronic collisions including excitations and ionizations. For charged particles at a given energy, is defined as the quotient $-dE/dl$ where dE is the average energy transferred to the medium when the particle has traveled a distance dl . For photons, the energy depositions are due to the secondary particles created after interactions with matter. In this way, the LET of photon beams is defined by the LET of secondary electrons. The most common unit to express the LET is $keV/\mu m$. The biological effects of IRs depend partially on this quantity, thus, the LET is considered a measure of radiation quality in radiobiology. Low LET radiations like electrons, γ - and X-rays, produce sparse energy deposition patterns while ions and neutrons, regarded as high LET particles, tend to deposit their energy along their tracks. Since different radiations may show diverse biological effects, even at the same dose, the Relative Biological Effectiveness (RBE) arises as a way to compare different radiation qualities from a radiobiological point of view. The RBE quantifies the capacity of a given radiation quality to produce a certain level of biological effect relative to a radiation of reference (γ -rays and X-rays are the common reference) [45].

$$RBE = \left(\frac{\text{Dose from reference radiation}}{\text{Dose from test radiation}} \right)_{\text{given level of biological effect}} \quad (1.6)$$

In most cases, the endpoint for measuring RBE is a given level of cell survival in *in vitro* cultures.

As a first approximation, the RBE increases with LET. However, the RBE shows a peak at $\sim 100 keV/\mu m$ [2] (see Figure 8). This behavior may be attributed to saturation

¹⁰ The International Commission for Radiological Protection (ICRP) has assigned the value of W_R for the main radiation qualities [43].

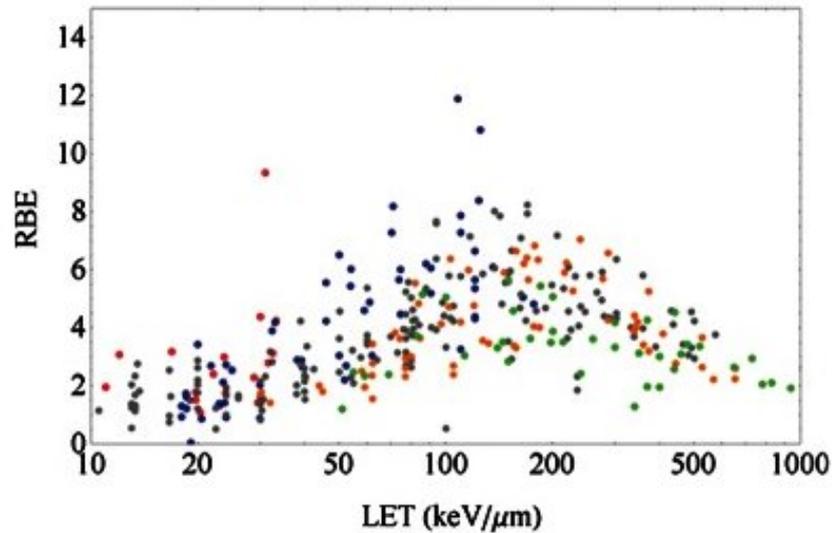


Figure 8 – RBE values for monoenergetic ion beams of different qualities and in the dose range of 4 – 30 Gy. Color code: p = red; He = blue; C = gray; Ne = orange; heavier ions than Ne = green. The RBE values show a peak around 100 keV/μm and then a decrease even at higher LET and doses due to saturation effects. Source: Friedrich et al. J. Radiat. Res. (2013) [2].

effects denominated "overkilling" that appear after this LET value. That is, at a certain point the damage on the DNA reaches a maximum so any additional injury does not really worsen the biological effect. Also, after the peak, the energy depositions are so close to each other that the number of sites affected reduces, and so the effectiveness of the radiation. High-LET radiation is able to deliver its energy in regions with sizes close to the DNA dimensions, thus, it tends to be more effective at inducing damages on the genetic material. Conversely, for low-LET radiation like photons and electrons, the energy deposition distribution is more sparse relative to this bio-molecule's dimensions ($\sim nm$).

2 Models of DNA/chromosome damage and cell death: analytical and Monte Carlo approaches

2.1 Analytical approaches

Since the beginning of the last century, when experiments on bacteria and chick embryo cells showed an exponential dependence of cell viability on radiation [125, 126], mathematical modelling has been an essential part of the fields of radiation biology and radiation therapy. The first theoretical models on radiation response were developed basing on the idea that cells have sensitive sites (or targets) that can be inactivated by radiation-induced damaging events (hits). The number of hits was suggested to be proportional to the dose delivered to the cell, and to follow a Poisson distribution with mean value $n = D/D_0$, where D_0 is the dose required to induce a mean number of hits equal to 1.

These simple considerations gave rise to a handful of models that sought to explain experimental observations regarding cell survival. For instance, in the first studies with bacteria and chick embryos, the exponential behaviour could be reproduced by assuming that only un-hit cells survive. Thus, the survival probability, S , is given by the probability of being hit 0 times, i.e.

$$S(D) = P_{hits}(0, D) = e^{-D/D_0} \quad (2.1)$$

where $P_{hits}(m, D)$ is the probability of being hit m times after a dose D . However, this description was not enough to explain the growing experimental data on mammalian cells exposed to photon irradiation. Those curves consistently showed a "shouldered" region that indicated lower sensitivity to radiation at low doses [127, 128]. Then, the multi-target single-hit approach was proposed to address this problem. The hypothesis was that cells have not only one but k sensitive targets (still with $P_{hits}(0, D) = e^{-D/D_0}$ each), which need to be inactivated in order to kill the cell. This assumption leads to the cell survival probability given by [Equation 2.2](#), which has a null initial slope that increases

gradually to $1/D_0$ at high doses.

$$S(D) = 1 - (1 - e^{-D/D_0})^k \quad (2.2)$$

Although the multi-target single-hit theory could fit part of the data, there was evidence that not all cell lines display the "zero initial slope" trend. Furthermore, experiments at high doses often showed an increasing curvature, up to the limit of detectability [129, 130]. Alternative formulations of the multi-target single-hit theory were proposed to overcome these problems. However, it appears that formulations based on the target theory lead to survival curves that are too complicated to be useful to explain the radiation response mechanisms [131]. Moreover, target theory models are also unable to describe the phenomenon of hypersensitivity to low-doses that is characterized by a V-shaped part in the 1 – 400 *mGy* range [132].

Later on, the Linear-Quadratic (LQ) model was introduced as an alternative model for cell survival. Its main characteristic is that it incorporates both a constant slope and a continually increasing quadratic curvature. Although very simple, the LQ model is nowadays the dominant mathematical model in the context of pre-clinical studies. Not only does it fit data successfully, but also offers some intuitive explanations for a range of biological effects. For instance, fractionation effects in tissues and tumors can be consistently interpreted in terms of the coefficients α and β of Equation 2.3, by noting that the ratio α/β defines a dose at which both the linear and the quadratic component contribute equally to cell killing. Thus, if cells are allowed to repair between two fractions of radiation, the shouldered portion of the curve is repeated and less killing is achieved in this way. The α/β ratio characterizes the degree of such sparing. The LQ formula, for a single acute irradiation, is given by

$$S(D) = e^{-\alpha D - \beta D^2} \quad (2.3)$$

where α and β are adjustable parameters (cell repopulation has been neglected in this case).

By the 70's, Kellerer and Rossi published their Theory of Dual Radiation Action (TDRA) in order to provide an interpretation for LQ model. They proposed that the formation of critical lesions was the result of two kinds of mechanisms. On the one hand, part of the critical lesions can be induced by the interaction of two "elementary lesions" produced by the same primary track, which is regarded as a first-order mechanism that leads to the linear term in the dose-response relation. On the other hand, a portion of the critical lesion can also be induced by two "elementary lesions" produced by independent tracks, introducing the quadratic term [133]. In this model, the nature of the "lesions"

was deliberately left unspecified. A year later, Chadwick and Leenhouts identified those critical lesions with Double-Strand Breaks (DSBs), and the elementary lesions (or sub-lesions) with Single-Strand Breaks (SSBs). Their work, known as "molecular theory of cell survival", also reproduced the general trends of the LQ model at low doses [134]. However, it is a matter of fact that DSBs tend to increase linearly with dose, instead of in a linear-quadratic fashion.

Apart from the aforementioned models, other approaches such as the "Repair Mis-Repair" (RMR) or the "Lethal Potentially Lethal" (LPL) models were also developed in the following years [135, 136]. Both RMR and LPL considered the yield of initial damage to be linear with dose. Then, the initial damage can be repaired by either linear or quadratic kinetics. Additionally, Goodhead tackled the problem from the perspective of enzyme depletion in his Saturable Repair (SR) model [137]. The basic assumption of the SR model was that as the number of lesions increases, the probability of successful repair falls, due to a reduction in the cell's ability to repair all of the induced damage.

All these models propose a mechanistic link between the LQ model and the damage induction and repair processes. However, distinguishing which models are correct (if any) remains controversial as all of them can predict dose response curves relatively well, with adequate selection of their parameters. Furthermore, at present, the very LQ model is being challenged by modern radiotherapy techniques, since there is an increasing interest in including other factors to analyze radiation responses. For instance, charged particle therapy [138], drug-radiotherapy combinations [139] and the role of tumour microenvironment after radiotherapy [140] among others. As an empirical model, the LQ parameters do not offer any sort of information about other end-points of interest in radiobiology. This motivates the development of more and more refined mechanistic models of radiation response in general.

Recently, most of the efforts, from an analytical point of view, are being focused on the kinetics of repair. Nevertheless, the induction of damage in the DNA by ionizing radiation is also an open area for theoretical and analytical developments. Specifically, the need of cross-section models at low energies for electrons and protons, is one of the bottlenecks in the field of track structure calculations nowadays [141]. With regard to repair kinetics, many approaches have been followed, including biphasic [142], biochemical kinetics [143, 144, 145], multiscale [11, 146] and statistical [147, 148] models. However, most of these approaches have met controversy, for example, biochemical models have been criticized with the argument that the biphasic model has never succeeded in providing definitive values for the repair components [149].

In general, the problem of "identifiability", that is, assigning unique values to the

model parameters, remains a hot topic as the models use too many parameters and different combinations of values can fit the data. Still, models of NHEJ have been capable of reproducing trends in both DNA repair kinetics and the role of some key genes [145]. An alternative method, which addresses the "identifiability" problem, uses Bayesian inference to assess repair mechanism activation [148]. This method relies on training a simple model against multiple datasets of DSB repair under different genetic knockouts when multiple repair mechanisms are activated. Then, with the set of most probable parameters, predictions can be made. The advantage of this model consists of having a small number of parameters whose interpretation and assignment is more robust.

2.2 Monte Carlo approaches

2.2.1 DNA damage and repair

Nowadays, Monte-Carlo (MC) simulations have become the standard when simulating radiation transport [150]. In such approach, both primary and secondary particles are followed in an event-by-event basis by modelling each event of ionization, excitation and scattering that those particles may undergo along their track. Early works on microdosimetric energy depositions at the sub-cellular level, highlighted the importance of detailed simulations in the context of radiobiology applications, as the energy distribution for different radiation qualities displays high heterogeneity even when similar average doses are delivered to the same volume [151, 152] (see Figure 9). Advances in this area have been witnessed since the 90's, when DNA structure features started to be included in the calculations [153, 154]. Thus, various aspects such as the amount of energy that must be deposited in a sugar-phosphate backbone to cause a strand break could be tuned by experimental data on measurable end-points such as SSB and DSB yields. Furthermore, the models provided means to investigate underlying mechanisms and endpoints which were not so readily measurable, such as the yield of complex, multiply-damaged sites [155].

Several MC radiation transport codes have been developed throughout the years by different groups. Typically, those codes were restricted to the energy range and particle types of interest to the research group that developed them. Among the most popular codes applied to dosimetry simulations and/or biophysical modelling, are: EGS [156], FLUKA [157], Geant4 [158], KURBUC [159], MCNP [160] and PARTRAC [161]. While many of these MC codes use "condensed history" physics models¹, which are accurate at the patient scale, it has been shown that simulation of each individual interaction is required at the cellular level [162, 163]. This resulted in the development of specific-purpose

¹ when some charged particle interactions are condensed into single steps to improve performance.

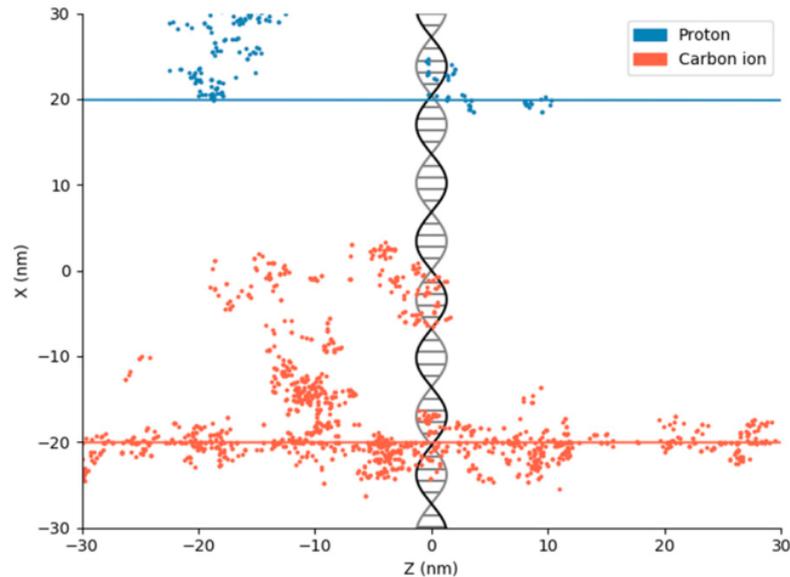


Figure 9 – Illustration of the track structure of two different radiation qualities. Blue (top): 10 MeV proton ($\sim 5 \text{ keV}/\mu\text{m}$). Red (Bottom): 200 MeV Carbon ion ($\sim 100 \text{ keV}/\mu\text{m}$). In both cases, the primary particles deposit a small fraction of their energy along their track (solid lines). Most of the energy is deposited by secondary electrons (points) near the track core, with notorious differences in their corresponding energy distribution patterns. A DNA structure is superimposed for scale. This highlights the importance of nanodosimetric calculations as densely ionizing radiation (e.g. Carbon ions in the figure) has the potential to induce more breaks in the genome than sparsely ionizing radiation (e.g. protons in the figure) at such scale. Source: S. McMahon and K. Prise. *Cancers*. (2019) [10].

MC codes that calculate track structures at lower energies, making them applicable to cellular scales. For instance, the GEANT4 MC code (one of the most widely used) has been extended to encompass ionizing events at energies relevant to cellular nanodosimetry (down to a few eV per particle) through the Geant4-DNA project [164, 165]. Additionally, the TOPAS and TOPASnBio toolkits, which are also extensions of GEANT4, are becoming popular in radiobiology research as they seek to make these Monte Carlo approaches more generally accessible [166, 167].

The success of MC codes in radiobiology research has kept growing over the years, and different codes have proven their capabilities in different fields and applications. Nevertheless, there is plenty of room for further improvements in the prediction of DNA damage from first principles. For instance, interaction cross-sections are one of the key pieces of input data needed for Monte Carlo models, describing the probability of various scattering and ionization events as a function of energy and target material. However, little data on cross-sections in organic materials at low energies (nanoscale), is available. Thus, MC calculations at the nanoscale use to treat organic materials as water-equivalents. Such approximation, while usually valid at the highest energies, at the lowest, that may not be the case [168, 169].

Additionally, among all the initial damage induced in the DNA, only a portion is due to direct interaction of radiation with the target molecules (around 30% for photon irradiation and higher for heavy charged particles). The majority of strand breaks come from chemical reactions between the DNA and the radical species and ions, created by the interaction of radiation with the water molecules surrounding the DNA. In this way, radiation chemistry becomes also an important component of MC simulations towards a better description of the radiation action in general. Thus, transport of chemical species started to be developed in most of the codes mentioned above [170, 161, 171, 172, 173]. However, in most cases these codes are still limited to simulations in liquid water, due to the lack of information on reaction rates in more complex chemical compositions.

Another open area in this field, is the modelling of DNA structures. The standard practice in MC applications in radiobiology, is to superimpose the outcomes of simulations in uniform water-equivalent volumes over a geometrical model of the genome, in order to assess the damage. Notwithstanding, which DNA models to use remains an important open question since the genome presents several organizational levels, and the damage induced by ionizing radiation may depend upon its local structure around the track [161, 174, 175, 176].

In spite of such limitations, there are many models based on the aforementioned transport codes, which predict yields of radiation-induced damages such as DSB, SSB and base lesions. Those models, in general, may incorporate different geometries and physical as well as biological models (some of them also include chemical transport). Still, they produce rather similar predictions. Such apparent flexibility in modelling can be attributed to the uncertainties associated with the biophysical parameters used in the simulations. The origin of those uncertainties comes both from experimental measurements on DNA damage, which are usually difficult to carry out, and, as already commented, from the lack of ultimate theoretical and computational models at the boundaries between physics, chemistry and biology.

Nevertheless, there is an ongoing effort in the scientific community of this field, to facilitate comparisons between models and help in their development. An example of this effort consists of adopting a Standard for DNA Damage (SDD) reporting system, which would allow modellers to test their model design and dependencies on underlying physics properties. This standard aims at promoting further interdisciplinary inter-group collaborations that will help us in advancing our understanding of DNA response to ionizing radiation exposure [177].

Regarding DNA repair, the primary mechanisms that cells possess in order to treat DNA lesions have been extensively studied (see subsection 1.3.2). Many key aspects of

DSB processing such as the proteins, complexes and other conditions, which are characteristic of a particular repair pathway (e.g. NHEJ, HR, MMEJ), have been identified from experimental studies. This has enabled the implementation of biochemistry models that aim at tracking the migration and interaction of DNA ends and proteins in detail.

For instance, Friedland *et al.* implemented a Monte Carlo model, which simulates NHEJ in a step-by-step fashion [11]. They used the PARTRAC code to provide starting conditions concerning the spatial distribution of DSBs and the characterization of lesion complexity. Then, the processing of DNA termini by the DNA-PK factory was simulated in detail, including first-order kinetics and diffusive motion. Their model required several parameters describing the association and dissociation of DNA ends with repair enzymes, as well as time constants related to the post-synaptic phase of the process, to be tuned. Such parameter-space was explored by combining both experimental data and hypotheses on the origin of the slow phase of DSB repair. This resulted in four scenarios for testing different parameter setups. The application of the model to human fibroblasts cells after low-LET radiation reproduced many features regarding rejoining kinetics, yields of misrejoining events and chromosome aberration induction. However, as the authors mentioned, further improvements are still required to better reproduce other end-points such as residual DSBs, and to incorporate more realistic assumptions with regard to DNA ends diffusion and interactions.

More recently, Henthorn *et al.* proposed a model to investigate the dependence of residual and misrepaired DSBs on the initial damage pattern created during ion therapy [178]. This approach also combines MC simulations of radiation transport with a stochastic simulation of the NHEJ process. In their work, they used the GEANT4-DNA toolkit to simulate the transport of different ion species across a water medium representing a simplified cell model. Then, the energy depositions were translated into DSBs over a geometrical model of the cell nucleus. This conversion is based on the assumption that a fraction of the nucleus is occupied by the sensitive material. Thus, to take into account indirect damage, such sensitive fraction was fitted according to experimental data that includes indirect effects. To simulate NHEJ, each DSB was allowed to move within the nucleus by sub-diffusive motion and, at each step, DSB ends underwent a series of stochastic time-constant-based state changes that represented the recruitment and action of repair protein complexes. The time constants were fitted according to experimental data on recruitment kinetics of the proteins involved in NHEJ. The authors applied their model to protons, α -particles and Carbon-ions across a range of track-averaged LET values. Interestingly, they concluded that the early biological effect of radiation can be predicted based on the extent to which DSBs are clustered. That is, it is independent of ion type, dose, LET and break complexity, when the fraction of misrepaired DSBs is considered as the end-point. However, different radiation qualities do produce different

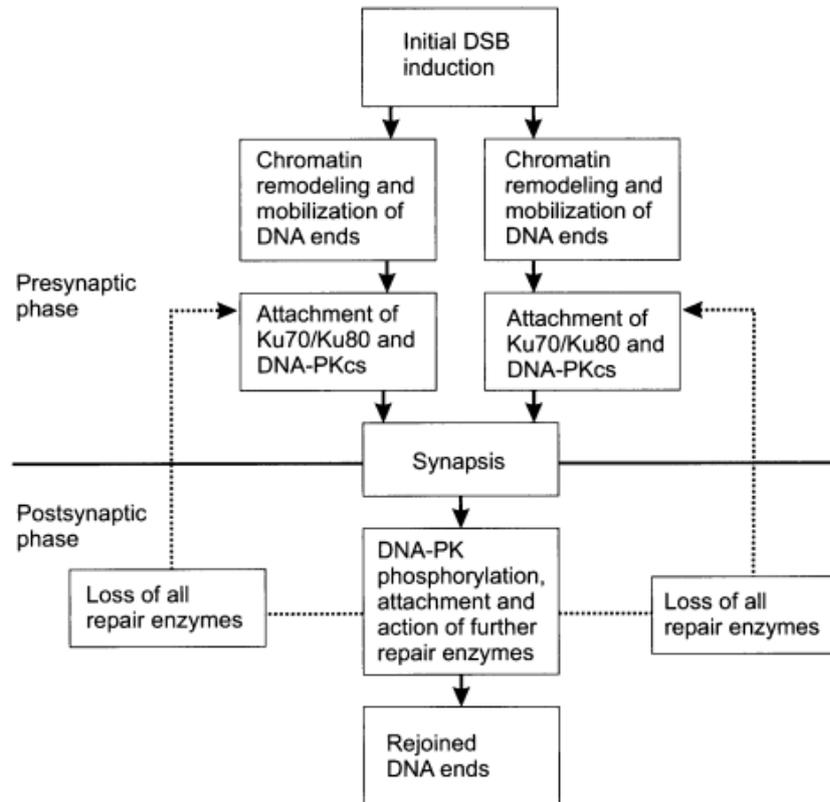


Figure 10 – Scheme of the NHEJ. The process, as implemented in Friedland *et al.* (2010), is divided in three phases: The *pre-synaptic* phase, in which DNA ends diffuse freely and, eventually, can be attached to Ku70/80 heterodimer and DNA-PKcs. The *synaptic* phase, where two DNA ends in sufficient close proximity, move together while are prepared for the next phase. In the *post-synaptic* phase, repair enzymes attempt to clean up dirty DSBs and finalize the rejoining process. Source: W. Friedland *et al.*. Radiat. Res. (2010) [11].

cluster patterns at similar LET values. In this way, they suggested that this mechanism can explain the experimentally observed differences in cell killing and mutation caused by different ion species at iso-LET.

Nonetheless, it is noticeable how the Micro-homology Mediated End Joining and Homologous Recombination repair pathways have not yet been implemented in a mechanistic fashion. Thus, their contribution to the chromosome aberration yields remains neglected in most cases. This may be attributed to the pathway complexity, in the case of HR, and the relatively lack of details of its functioning, in the case of MMEJ. Furthermore, even though NHEJ has started to be modelled from a mechanistic perspective, the extensive number of parameters that are required to reproduce experimental data, still makes it difficult to draw final conclusions regarding model testing and selection.

2.2.2 Chromosome and cell level response

Apart from the aforementioned models, which focus on detailed simulations of DNA damage and repair at the nano-scale, there exists another class of models that focus on simulations of radiation response at the micrometer-scale. Such models aim at linking early damage with chromosome aberrations and cell death, on a mechanistic way as well, but paying special attention to proximity effects as the major driver of misrepair. In exchange of the lack of details about DNA repair, some of these models have been successful in reproducing yields of chromosome aberrations and cell survival quite well, with only a few adjustable parameters. Additionally, they are computationally less expensive, which facilitates their possible use in future clinical applications.

Some of these models use CA yield as the end-point, seeking to reproduce yields of the main aberration categories observable with painting techniques. Thus, their parameters are, usually, tuned by comparison with FISH data. That is the case of, for instance, the CAS model developed by Chen *et al.* [179], which was subsequently upgraded to CAS2 [180]. Another example is the NASA Radiation TrackImage model/code published by Ponomarev *et al.* [181]. Furthermore, other models not only use CA yields as end-points, but also incorporate cell killing as in the case of BIANCA [182] and the Repair-Misrepair-Fixation (RMF) model [183]. A common feature among these approaches, apart from focusing on the already mentioned end-points, is that they are heavily based on distance dependence relationships for modelling the formation of aberrations. As a consequence, the geometry of the cells nucleus plays an important role as well. Details concerning most of these models will be presented in [subsection 3.2.3](#), in order to contrast the approach followed in this thesis work, with similar models.

Nevertheless, some generalities can be outlined here. For instance, while part of these models distinguish DSBs from clusters of DSBs, the existence of lesions severe enough to be considered "critical", is a common assumption². The distribution of initial damage is based on LET and dose, distinguishing between sparsely and densely ionizing radiation. The former, induces damages uniformly distributed over the nucleus volume, while in the latter, the lesions are placed along "tracks" in order to emulate denser patterns. The dose determines the number of initial lesions. Additionally, interphase chromosome organization within the nucleus is essential in the distinction of aberration products. Thus, these approaches implement, either implicitly or explicitly, the notion of chromosome territories. Finally, proximity effects are taken into account through misrejoining probabilities between DNA fragments (created by the critical lesions), which depend upon the initial distance between such lesions.

² e.g. "reactive DSBs" in CAS2 or "cluster lesions" in BIANCA.

Regarding cell response in terms of survival, there are extensions of many of the models cited along this section. However, when it comes about clinical RBE modelling, the Local Effect Model (LEM) [184] and the Microdosimetric Kinetic Model (MKM) [185], dominate the scene nowadays. Nevertheless, the enhancements in the BIANCA model in the last years and during this PhD project, have expanded its capabilities in such a way that clinical applications are now within the scope of BIANCA (see [subsection 4.4.2](#)).

Both LEM and MKM, predict the effects of high LET radiation basing on microdosimetric calculations and biological parameters extracted from photon exposure. In LEM, the basic assumption is that the local biological effect, that is the damage in a small cellular sub-volume, is determined by the expectation value of energy deposition in that sub-volume and is independent of the radiation type. Thus, for a fixed biological target, the only difference relies on the different radiation track structures. Additionally, this model assumes that cell survival is given by

$$S = e^{-\langle N \rangle} \quad (2.4)$$

where N is the mean number of "lethal events" induced on the cell. Specifically, it is assumed that the Linear-Quadratic relationship observed for photons, holds in those cellular sub-volumes. Thus, the total number of lethal lesions, N , is calculated as

$$\langle N \rangle = \int \frac{\alpha D_r + \beta D_r^2}{V} dV \quad (2.5)$$

where D_r is the dose delivered at a point r , V is the total volume of the nucleus, and α and β are extracted from photon survival curves irrespective of the radiation quality in question.

More recent versions of the LEM have moved to explicitly consider break repair and misrepair probabilities similar to some other models described above, but these approaches have yet to make the transition to clinical practice [186]. On the other hand, the MKM uses a similar approach with the difference that "sub-lethal events" are also considered. These "sub-lethal lesions" can either combine and form a "lethal" lesion (misrepair) or successfully repair. In this model, the average number of "lethal events" is given by

$$\langle N \rangle = AD + \beta D^2 \quad (2.6)$$

where β is independent of the radiation quality and thus, is taken from X-rays curves. In contrast, A depends not only on photon parameters but also on the dose mean specific energy by single energy deposition events in the domains. Cell survival is then calculated by means of [Equation 2.4](#).

These models allow calculation of RBE values in a variety of situations achieving good agreement with experimental data. Additionally, the MKM has been applied to brachytherapy [187], and Boron Neutron Capture Therapy [188], while the LEM has been applied to gold nanoparticle enhanced X-ray therapies [189]. Currently, both models support treatment planning optimization at carbon therapy centers in Europe (LEM) and Japan (MKM) [190, 185].

3 BIANCA model: Former versions and version developed in this work

BIANCA stands for *BIophysical ANalysis of Cell death and chromosome Aberration* and is a model, implemented as a fortran code, designed to describe the formation of chromosome aberrations and cell death. It is based on a few mechanistic assumptions that, along with a small set of adjustable parameters, investigate the interplay between ionizing radiation and DNA.

3.1 Overview

The early versions of the BIANCA model began development in 1999 by Ballarini *et al.* and continued evolving throughout the years to expand and refine its capabilities. Herein will be described the assumptions of the model, which are common to both the versions developed in this PhD project, as well as the previous. Also, a brief review of the main changes and achievements of the BIANCA versions prior to this work will be presented, as it will help the reader to gain perspective of the evolution of the model.

3.1.1 Model assumptions

Nowadays, the DNA is still considered the primary sensitive target after ionizing radiation interacts with a living organism [29]. Thus, modelling the cell nucleus as the Region Of Interest (ROI) has been the main approach over the years to study the radiobiological problem. BIANCA follows this paradigm too. To simulate the formation of chromosome aberrations (CA), BIANCA assumes that:

1. Ionizing radiation induces a number of effective DNA lesions that are severe enough to break the chromatin fibre in two independent fragments. Such lesions are referred to as *cluster lesions*, or CLs, since it is likely that they are clustered.
2. Two chromatin fragments can undergo end-joining according to the so called "proximity effects". The process of end-joining is modelled as a probabilistic event that depends on the initial distance, r , of the two involved ends.
3. Asymmetrical CA such as dicentrics, rings and large deletions, where "large" means large enough to be detected by Giemsa staining in metaphase (i.e. larger than ~ 3 Mbp, according to [9]), are assumed to lead to clonogenic cell death.

There exist experimental evidences that support assumptions 1 and 2. Such evidences show that, despite that 1 Gy of radiation produces $\sim 30 - 40$ DSBs per cell on average, only 1 or even fewer chromosome aberrations are induced in the cell. This means that only a subset of all the initially induced DSBs lead to chromosome aberrations. It is likely that these critical DSBs are of a complex type, since the complexity of the damage is a key factor in the biological expression of the radiation effect, as the more complex the damage, the higher the risk for the repair mechanisms to fail [191]. Thus, in BIANCA it is assumed that, among all the DSBs induced after irradiation, only those that are complex enough are able to lead to a chromosome aberration (assumption 1), as long as they are in spatial proximity (assumption 2). Such complex lesions are likely to be cluster lesions. Since the model does not provide an operational definition of what a cluster lesion is, the mean number of CLs per unit dose and cell is left as an adjustable parameter. Regarding assumption 3, in the work of Cornforth and Bedford [9] it was found a one-to-one relationship between the average number of lethal aberrations (where "lethal" means *dicentric + rings + large acentric fragments*) per cell and the logarithm of the cell surviving fraction, for AG1522 cells exposed to X-rays. In BIANCA, the same reasoning is extended to other cell lines and radiation qualities in order to approach clonogenic cell death. Nevertheless, the model does not exclude the fact that there are other mechanisms that may play an important role in cell killing, for instance, non-transmissible complex aberrations, apoptosis, etc.

3.1.2 Evolution of BIANCA

3.1.2.1 Early versions

From the first model versions introduced by Ballarini *et al.* [192], a pivotal role in biological radiation-induced damage has been attributed to chromosome aberrations. It was assumed that CAs arise from mis-rejoining or unrejoining of chromosome fragments produced by severe DNA damages called "Cluster Lesions". In such early versions, the localization of the interphase chromosomes was not modelled explicitly but rather introduced as a probability of intersection with a particle track whose CL yield was taken from [193]. The number of cluster lesions for a given dose was extracted from a Poisson distribution and each CL was assigned to a chromosome arm with a probability proportional to the arm DNA content. It was assumed that every CL leads to a chromosome break, i.e, it produces two independent fragments. To simulate the formation of chromosome aberrations, it was assumed that each fragment has the same probability to interact with every other possible partner, without any dependence on proximity. Subsequently, aberrations were scored and classified. The procedure was then repeated multiple times in order to achieve statistically significant results. Dose-response curves were constructed by applying the same procedure for different doses. Good agreement was found with ex-

perimental data taken from the literature; however, there was a tendency to overestimate the so-called "complex exchanges". This tendency suggested that proximity effects needed to be introduced.

Later on, the explicit modelling of chromosome territories was introduced and proximity effects were taken into account [194, 195, 196]. The interphase chromosomes were constructed within a sphere of radius $3 \mu m$ (representing a human lymphocyte), by means of a 3D mesh with 27000 cubic voxels of $0.2 \mu m$ side. The number of voxels assigned to each territory was proportional to its DNA content. The steps to create each chromosome territory consisted of selecting a starting voxel and growing the domain by adding the neighboring voxels that were available until a stop condition was reached (in principle, determined by the DNA content, however it was not always possible to meet this criteria for all the chromosomes simultaneously). In this way, it was possible to construct the 46 interphase chromosome territories (CT). Thus, when a CL hit a voxel, it was straightforward to determine which chromosome was injured. In such events, the CL was assigned to one of the chromosome arms according to a probability dependent upon their genome content.

To account for proximity effects, exchange-type aberrations were assumed to be originated only by fragments that were spatially close, either in the same territory (in-trachanges) or in adjacent ones (interchanges). This assumption was implemented in the simulation code introducing a cut-off distance d : two chromosome free-ends were allowed to undergo end-joining only if their (initial) distance was smaller than d , and within d each pair of fragments had the same end-joining probability. The value of d was set to $1.5 \mu m$. For light ions, the CL yield was taken from [193], while for γ -rays it was adjusted after comparison with experimental data. These improvements in the original model led to results in line with experimental data [197, 198] for dicentric yields, in the case of $5 keV/\mu m$ protons and $86 keV/\mu m$ α -particles, and F-ratios in the case of gamma-rays, $5 keV/\mu m$ protons and $121 keV/\mu m$ α -particles. Dose response curves for non-lethal aberrations (e.g. reciprocal translocations) were also computed and compared with experimental results obtained by FISH. The model was then applied to predict aberration yields in lymphocytes of astronauts involved in long-term missions onboard the MIR space station or the International Space Station. The calculated dicentric dose-response curves were consistent with astronauts' data [199].

In 2007, further extensions of the model were carried out in order to include the action of heavy charged particles [200, 201]. To this end, some CLs were radially "shifted" from the track core to simulate the effects of "delta rays" (c.f. subsection 3.2.2). Good agreement was found with PCC experimental data [202], measuring "simple exchanges" (dicentrics + translocations) in human lymphocytes exposed to $1 GeV/u$ Fe-ions and

290 MeV/u C-ions.

3.1.2.2 The first BIANCA

From 2010, the biophysical model thus far developed, which was focused on chromosome aberration formation, was extended to cell death¹ [203]. Since very little data is available on the survival of lymphocytes, which are generally grown as cell suspensions, fibroblast nuclei were implemented in the code. At this stage, the full set of assumptions presented in subsection 3.1.1 came into play, that is, the third statement was incorporated. As already mentioned, new geometries were introduced at this point, specifically, cylindrical nuclei representing AG1522 human and V79 hamster fibroblasts (elliptical and circular base cylinders, respectively).

Concerning the simulation of irradiation, for each cell line and radiation quality the average number of CLs per Gy and per cell was fixed as an input parameter. In the case of ions, the mean number of lesions per micron instead was used (see Equation 3.3 in subsection 3.2.2). For heavy ions, the probability for a lesion to be at the core of the track was taken as 50%. Those lesions lying at the penumbra (out of the core) of the track were radially shifted by means of Equation 3.7 (see subsection 3.2.2). As before, the actual number of lesions was extracted from a Poisson distribution for each dose of interest. Once the lesions were placed within the nucleus, the formation of aberrations was still modelled using a step-like probability function. That is, chromatin ends separated at a distance smaller than a threshold distance d , had the same chance to rejoin to any potential partner. The scoring of aberrations was based on the products visible under GIEMSA staining, namely, dicentrics, rings and deletions, which can be identified by their shape and number of centromeres. All the steps were repeated until achieving statistical significance. For a single dose D , the cell surviving fraction was then calculated as $S(D) = e^{-\langle LA \rangle}$, where $\langle LA \rangle$ was the average number of "Lethal Aberrations" per cell at a dose D . Cell survival curves were constructed by running simulations with several doses.

In the first BIANCA, both the average number of CL per Gy and per cell (CLy) and the threshold distance (d) needed to be adjusted by comparison with experimental data. To find those values for a given cell line, a reference radiation curve (typically photons) was used to first tune d . Next, keeping unchanged the value of d , CLy was found for ions at a specific dose, and then both parameters were fixed throughout the simulations with other doses. The motivation for this procedure came from the observation that d would represent a characteristic of the target cell (independent of the radiation quality), while CLy reflects the ability of a given radiation to induce damage. However, with regard to the CLy , some modulation is also expected by the considered cell line,

¹ From this extension and on, the model receives its acronym.

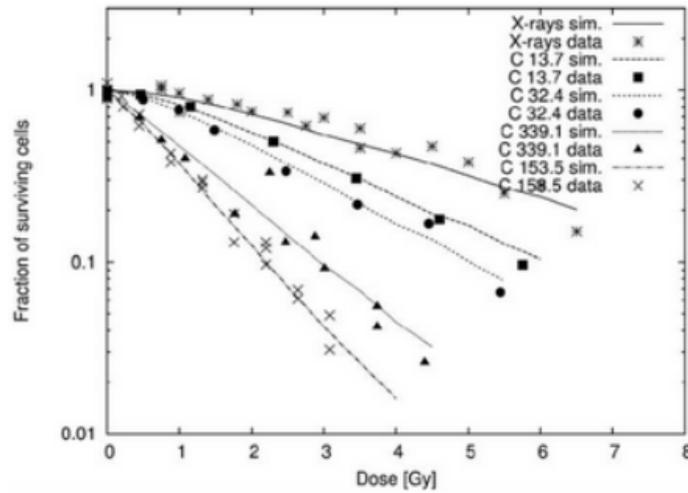


Figure 11 – Survival curves for V79 cells exposed to X-rays and to four carbon-ion beams of differing energy. Simulations were carried out using the first version of BIANCA. Source: Ballarini *et al.* Rad. Res. (2013) [12].

depending on the specific features of the cell repair mechanisms.

Simulated survival curves were compared with survival data on V79 cells exposed to different mono-energetic beams of protons or Carbon ions, as well as a reference radiation (X-rays) [12]. From the X-rays data, it was found $d = 5 \mu\text{m}$ and $CLy = 1.8 \text{ Gy}^{-1}\text{cell}^{-1}$. Using this d , the values for CLy were in the range of $2 - 4 \text{ Gy}^{-1}\text{cell}^{-1}$ for protons with LET between 7.7 and $34.6 \text{ keV}/\mu\text{m}$. On the other hand, for carbon ions, the value of d was fixed to $4 \mu\text{m}$ as the cells used in the experiment considered for comparison were smaller. Concordance with the X-ray experimental curve was found with $d = 4\mu\text{m}$ and a CL yield of $1.3 \text{ Gy}^{-1}\text{cell}^{-1}$. The difference in the CLs yields for the two experiments may reflect the differences in radiosensitivity between the two cell lines. The Carbon ions had LET values in the range of 13.7 to $339.1 \text{ keV}/\mu\text{m}$ and the CLy were in the range of $2 - 6 \text{ Gy}^{-1}\text{cell}^{-1}$ (see Figure 11).

Furthermore, the model was applied to the simulation of human fibroblasts exposed to X-rays and $116 \text{ keV}/\mu\text{m}$ α -particles as in [9, 20]. For this cell line, the threshold distance was found to be $3 \mu\text{m}$, while the CLy were 2 and $9.4 \text{ Gy}^{-1}\text{cell}^{-1}$, respectively. With these values, a good agreement with the data (considering the experimental error bars) was found. The approach was then extended to other two AG1522 data sets taken from the literature. One data set consisted of gamma rays, C, He and Ne data [13], while the other one comprised gamma and α -particles data [14]. In this case, all curves were obtained with $d = 5 \mu\text{m}$. The difference with respect to the other experimental data set was attributed to important differences in the experimental techniques carried out by the corresponding laboratories. Nonetheless, the simulation outcomes were supported by the experimental data in most cases, as can be seen in Figure 12 [15].

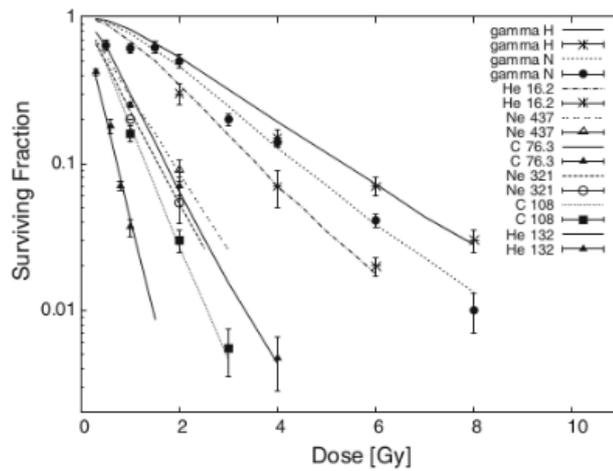


Figure 12 – From top to bottom: survival of AG1522 cells exposed to γ -rays (two curves: the upper one compared with data from Hamada et al. 2006 [13], and the lower one compared with data from Neti et al. 2004 [14]), $16.2 \text{ keV}/\mu\text{m}$ He ions, $437 \text{ keV}/\mu\text{m}$ Ne ions, $76.3 \text{ keV}/\mu\text{m}$ C ions, $321 \text{ keV}/\mu\text{m}$ Ne ions, $108 \text{ keV}/\mu\text{m}$ C ions, and $132 \text{ keV}/\mu\text{m}$ α -particles. The lines are simulation results, whereas the points are experimental data taken from Hamada et al. 2006 (upper gamma curve, He ions, C ions, and Ne ions) or Neti et al. 2004 (lower gamma curve and α -particles). Source: Ballarini *et al.* Rad. Res. (2014) [15].

3.1.2.3 BIANCA II

Despite the positive results obtained with BIANCA, the fact that the threshold distance was $5 \mu\text{m}$ appeared to be incompatible with experimental evidences regarding chromatin mobility. For instance, $1 - 2 \mu\text{m}$ DSB migration distances have been estimated for MCF10A human epithelial cells [204]. Additionally, it has been suggested that repair mainly takes place in small channels separating adjacent chromosome domains [205]. To encompass these observations within the framework of BIANCA, a newer version of the model was developed between 2015 and 2017, BIANCA II. In this version, the threshold distance was fixed *a priori* to the average distance between adjacent territories². For human AG01522 fibroblasts, the average distance was found to be $3 \mu\text{m}$, while for V79 hamster fibroblasts it was found to be $3.6 \mu\text{m}$. It is important to point out that in BIANCA II the threshold distance d was fixed *a priori* on the basis of the cell geometry, so it was no longer an adjustable parameter. Notwithstanding, the proximity effects remained to be modelled with a step-like function.

On the other hand, the first BIANCA model tended to overestimate the survival fraction at low doses. This was probably a consequence of assuming a 100% rejoining probability within the threshold distance d . For low doses, when isolated lesions can be induced, eu-rejoin is always available for the fragment pair created by the lesion. In

² The distances were computed from the center of mass of each chromosome territory.

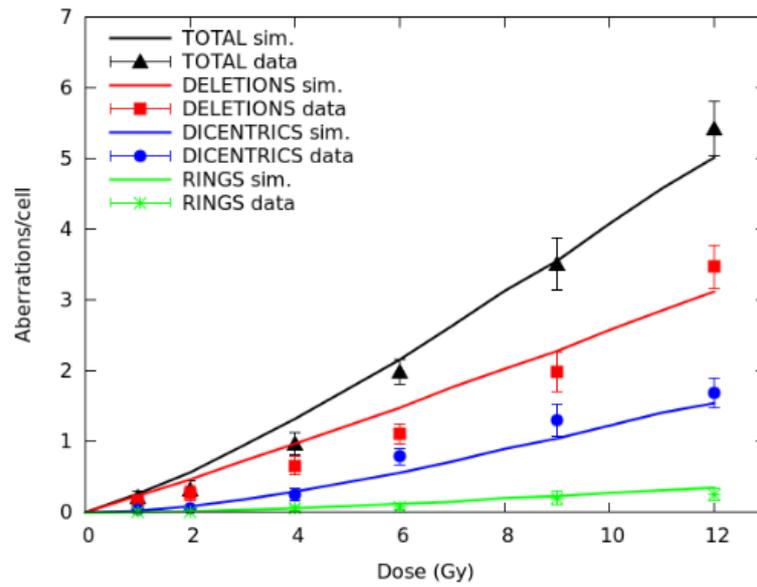


Figure 13 – Yields (mean number per cell) of different aberration types (dicentric, rings and deletions, as well as total aberrations) in AG1522 primary normal human fibroblasts exposed to different doses of X-rays. The lines are simulation outcomes, the points are experimental data taken from [9]. Source: Carante *et al.* Rad. Res. (2016) [16].

BIANCA II a more realistic scenario was considered in which a chromosome fragment, even if one or more potential partners are available within d , has a probability f of remaining un-rejoined. This assumption is consistent with studies indicating that a certain fraction of exchange-type chromosome aberrations are "incomplete", i.e. not all the involved chromosome fragments are finally rejoined [206, 207, 208]. Such probability of remaining un-rejoined is expected to be cell-line-dependent, as more radiosensitive cells show higher yields of deletions with respect to normal and radioresistant cell lines [209]. Thus, in BIANCA II, f was left as an adjustable parameter. In this way, there remained only two free parameters in the model.

The data set for gamma rays and $116 \text{ keV}/\mu\text{m}$ α -particles was revisited using the BIANCA II version [16]. The f value was adjusted from the gamma data and then, it was kept fixed for the simulations with α -particles. For $f = 0.18$, an acceptable agreement was found with the experiments even when the aberration categories were separated into dicentric, rings and deletions (see Figure 13). The same procedure was carried out for V79 hamster fibroblasts with data at 3 Gy . For this cell line, $f = 0.8$ was found. The two values of f obtained by these comparisons, both for AG1522 and for V79 cells, were used as input parameters for the simulation of cell survival curves by different ion species in subsequent works in hadrontherapy, carried out by Ballarini *et al.*

3.2 The version developed in this work

In order to further investigate the formation of chromosome aberrations, in this thesis work, BIANCA has been upgraded to include new proximity functions to describe the rejoining probability of chromatin free ends. Also, an explicit representation of chromosome arm domains has been included in the geometrical model, thus improving the distinction of intra-/inter-arm intrachanges products. This section, provides a description of the upgraded version of BIANCA from a computational point of view, as the whole architecture of the code that implements the biophysical model was revisited. Thus, the new structure of the code constitutes a major part of the work carried out during this PhD project. A summary of the technical improvements with respect to the previous version of BIANCA will be presented in [chapter 4](#).

The rationale of code consists of creating a geometrical model of the cell nucleus, then, generating critical lesions within the genetic material to finally assess the number and type of aberrations induced by those critical lesions, as well as the probability of cell survival. The outputs of BIANCA are two .txt files, one containing the yield of different aberration types and survival fraction, and the other one showing the aberration distribution, both as a function of dose. By default, BIANCA exports and renders a pov file with the last geometry used along the simulation.

Two adjustable parameters, namely, the *cluster lesion yield* **CLy** and the *typical interaction distance* σ (in microns) need to be passed to the code in order to run a simulation. BIANCA now implements three different biophysical models to determine chromosome aberrations. Such models are labeled as the *Gaussian-*, *exponential-* and *step-like-* model and are controlled by the *typical interaction distance*. The *cluster lesion yield* defines the effective number of lesions distributed within the nucleus.

The user has to provide an input file that encloses the information about the radiation, doses and cell line of interest. This file also specifies the value of the two adjustable parameters and the selected model for the simulation. [Figure 14](#) shows how an input file looks like. Keywords and their values will be explained along the chapter. In general, the file accommodates the cell culture information (both cell line and number of irradiated cells), characteristics of the particle like its LET, energy and a descriptor of the track structure along with the type of beam, the doses to be delivered and the model parameters to be used.

3.2.1 Geometry

In this work, the cell nucleus content was modelled by means of a discretized 3D grid made up of cubic voxels of $0.1 \mu\text{m}$ side. Currently, the code is able to generate three

```

Input.dat ✕
1 Cell_Dish: V79 5000          ! AG/lymph/V79   Number_of_cells
2 Beam: broadbeam beam.dat    ! broadbeam/microbeam  beam.dat(only for microbeam)
3 Particle: heavyion 137. 12.9 ! photon/lightion/heavyion LET Energy
4 Cluster_Yield: 6.7 GyCell    ! Number GyCell/GyDa/um (if GyDa write E-13)
5 doses: '0.5,1.,2.5,3.,4.'    ! 'd1, d2, ..., dn'
6 mydist: 0.7                 ! r0 (characteristic distance for all the models)
7 Model: StepModel           ! ExpModel/GaussModel/StepModel
8

```

Figure 14 – Input file of BIANCA. The admitted values of each key are shown on the right as comments. The target information is controlled by the *Cell_Dish* keyword. *Beam*, *Particle* and *doses* determine the irradiation setup. The adjustable parameters, namely, the CLy and σ and the selection of the model are passed through the *Cluster_Yield*, *mydist* and *Model* keywords, respectively. The order of the lines does not matter but for a given line, the key values order does.

types of geometries, namely, spherical-, cylindrical with circular base- and cylindrical with elliptical base-nuclei (suitable for lymphocytes, V79 and AG01522 cell lines, respectively). The shape of the nucleus is decided according to the specifications in the input file given by the *Cell_Dish* keyword. The values of the key can be *AG*, *Lymph* and *V79* for human fibroblasts, lymphocytes and hamster fibroblasts, respectively. Also, the number of cells is passed.

For each cell line, there is a .dat file in the source environment that defines the shape, size and genome content of the nucleus. Table 3 shows the information contained in the header of each .dat file. Then, the file deploys a list with the label and the number of Mega-base pairs (Mbps) in the long and short arm of each chromosome. The shapes selected for each cell line correspond to the typical growing techniques applied in experiments with these cells. For instance, the nuclei of AG1522 fibroblasts grown as monolayers will show an elliptical base with fixed height, while lymphocytes grown in suspension will approximate spheres. The semi-axes determine the size and number of voxels NV_{tot} of the grid. A number of voxels NV_{chro} proportional to the DNA content is assigned to each chromosome. NV_{chro} represents the size of the chromosome territory (CT). The corresponding number of voxels for each arm of each chromosome is also computed, to explicitly model chromosome arm domains.

Table 3 – Shape, size and number of chromosomes of the cell lines implemented in BIANCA.

Cell line	Shape	No. Chro	semi-axis (μm)		
			x	y	z
AG	Elliptical cylinder	46	10	5	2
Lymph	Spherical	46	3	3	3
V79	Cylinder	22	6	6	2

Following the work performed in this PhD, the construction of a nucleus with N_{chro} CTs follows three major steps, assuming no significant intermingling during interphase. First, N_{chro} voxels that define the origin of the CTs are selected at random in such a way that every CT could have, at least, 50% of its NV_{chro} voxels available for the subsequent steps. This is accomplished by sampling the voxels one by one, rejecting and re-sampling those that lie within a sphere with center in a previously accepted CT origin. The radius of such spheres is calculated from NV_{chro} . This 50% initially reserved volume proved to be enough for optimal performance of the algorithm. Larger values could restrict the number of configurations that fit the CTs inside the grid, while smaller values may conflict with the disposition of the chromosome arms as it will be clear in brief. Subsequently, a few voxels around the origin are labeled as pertaining to the domain, and a randomly oriented plane is attributed to the CT. This plane splits the CT into two sub-domains, i.e. the chromosome arms. Points in the positive side of the plane (defined by the normal) correspond to the short arm, while points in the negative side belong to the large arm.

Once the origins of the CTs are in place, all domains are grown simultaneously by adding to the CT one of the neighbors of a randomly selected voxel of the domain. To do so, first, a CT and arm is selected, then, a voxel of the sub-domain with at least one non-assigned neighbor is chosen (it could happen that the selected voxel is surrounded by voxels that already belong to a CT) and finally, one of the non-assigned neighbors is added to the CT. This is repeated until either all chromosome domains reached at least 80% of their volume, or a trapped domain is detected. To guarantee that all CT volumes grow at the same rate, the sampling of the CTs is weighted by the current number of already assigned voxels. A trapped domain occurs when a given CT does not have non-assigned voxels to add available. One scenario for this is, for instance, when a small chromosome is surrounded by larger chromosomes. When a trapped domain is detected, the whole nucleus is discarded and a new one is constructed, meaning that the origins are sampled again. The same considerations apply to the intra-chromosome domains, i.e. the algorithm guarantees that also the chromosome arms grow at the correct rate. A simple but efficient solution to the trapped domain problem was the *initially reserved volume* discussed above. The algorithm performs better in terms of speed and low chances to have trapped domains when this value is kept around 50%.

As the domains grow, the chances to select voxels with no non-assigned neighbors increase. For this reason, it becomes more practical to switch from the growing process to a capture process, in which the remaining non-assigned voxels are sampled and designated to the closest CT. This is the final step of the construction of the nucleus. A given non-assigned voxel is rejected if it is not in direct contact with a CT. If the voxel is accepted, one of its faces is chosen at random until the designation occurs. The changing from the growing to the capture process at the 80% of the CT volume comes from test of

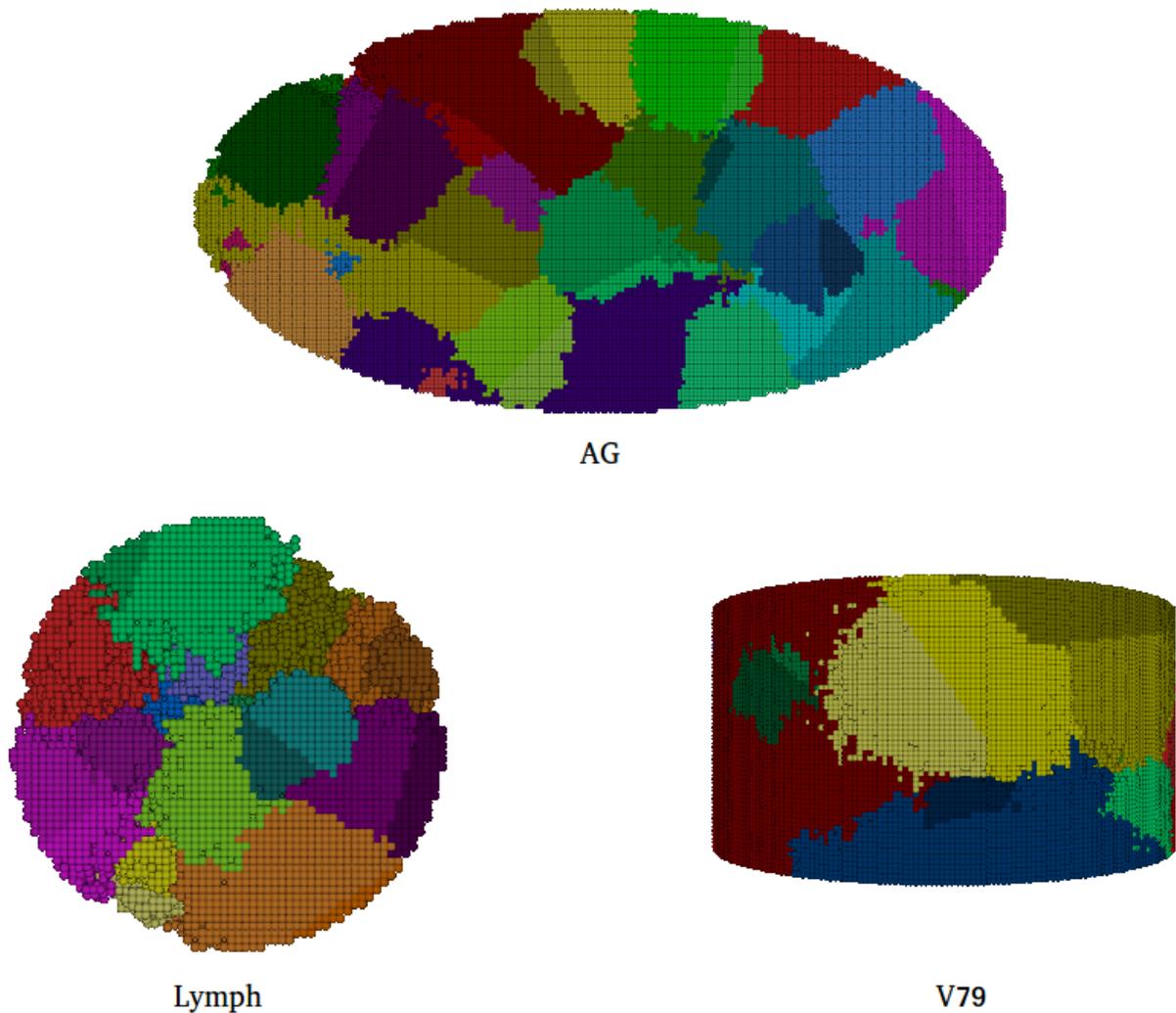


Figure 15 – An example of the cell models available in BIANCA. On top, an AG fibroblast grown as monolayer modelled as an elliptical cylinder. At the bottom, on the left, a lymphocyte grown in suspension simulated as a sphere. V79 cells are shaped as circular cylinders. Dimensions of each geometrical model are those shown in **Table 3**. Each color represents a chromosome territory. Changes of tonality within a chromosome are intended to highlight its arm domains.

performance as well. At this stage, the nucleus is finished and needs to be accepted or rejected. To this end, the actual number of voxels assigned to each chromosome arm of each chromosome is calculated. A nucleus is rejected if any arm could not reach at least 70% of its size, otherwise, it is accepted. All models created in this manner, that is, after tuning the parameters of the nucleus generator, have a fidelity of 94% or superior in terms of genome content. **Figure 15** shows an example of the geometries produced by BIANCA for each available cell line. The coordinates are exported in pov format and rendered automatically. The fidelity of the nucleus and the time required to create it, among other information, are printed in a log file. In this example, the AG cell took 95.3 s and achieved 97.3% of fidelity, while the V79 and the lymphocyte took 337.6 s and 4.23 s, respectively, and their fidelity was 97.9% in the case of V79 and 100% for the lymphocyte.

An internal random seeder is called from the beginning of the nucleus construction in order to ensure repeatability of the structures. This is important since the interaction models invoke random number generators as well. Thus, it is convenient to detach the geometry from the simulation itself for validation purposes. Another advantage of using an internal random seeder is that it allows to speed up future simulations, as the seed that creates an accepted nucleus is automatically stored in the BIANCA include folder. Every time the nucleus generator is called it first checks for the existence of seeds that produced accepted geometries. The random seed generator implemented in the module is the example shown in [210].

3.2.2 Irradiation

BIANCA distinguishes three types of particle sources according to their track structure, namely, photons, light ions and heavy ions. Distinction of these categories comes from the fact that their energy deposition patterns are very different due to the physical mechanisms that govern their interactions. LET, energy and production of secondary particles are the most relevant physical quantities and mechanisms that account for such differences, as discussed in chapter 1.

Instead of simulating the transport of primary particles, BIANCA models the immediate consequences of the irradiation of the target as a distribution of DNA cluster lesions that reflects the track of the primaries and, when needed, the secondaries as well. For photons, the CLs are placed uniformly within the nucleus since they are considered sparsely-ionizing radiation. In the case of low-energy light ions, the lesions are distributed along straight lines that cross the ROI. This pattern reflects the fact that most of the energy depositions of the primary particle (e.g. a proton) occur along its track, and so is for the effective damage that it causes. Heavy ions also travel through the nucleus in nearly straight lines, nevertheless, they are more likely to produce secondary electrons with enough energy to escape and deposit their energy away from the primary track. This is modelled in BIANCA as a cylindrical track in which part of the cluster lesions are placed in the track core and part in the penumbra [211]. The *Particle* keyword in the input file controls the type, LET and energy of the particle to be used. The first entry may take the values *photon*, *lightion* or *heavyion* and the second one specifies the LET in $keV/\mu m$. There is a third value that is read for the key and corresponds to the energy, in MeV/u , of the particle. For *photons* and *lightions* this entry is ignored, while for *heavyions*, a non-zero value is expected. In the case of heavy ions, the energy per nucleon E determines the maximum radial distance at which a lesion could be found. This radial distance is given by [212]

$$r_{max} = 0.062E^{1.7} \quad (3.1)$$

where r is expressed in μm and E in MeV/u .

In BIANCA, the source of particles is placed at the bottom of the ROI and emulates normal irradiation of the target. The user has the option to select between a broad- or a micro-beam. The keyword *Beam* expects two strings, the first one having the value *broadbeam* or *microbeam* and the second one being a txt file that contains the position of the source in the case of micro-beam simulations. The .txt file must contain the information about the number of beams and particles per beam, the FWHM of the beam and the coordinates of each shooting point. Doses are passed through the *Doses* keyword between quotation marks.

The Cluster Lesions Yield is an adjustable parameter of the model that accounts for the mean number of DNA lesions per unit dose and per cell that a given radiation is able to induce. The value of this parameter is passed through the *Cluster_Yield* keyword and consists of a float and a string that specifies the units of CLy. *Gycell* means that the value is given per unit dose in Gray and unit cell, while *GyDa* refers to a value that is given per atomic unit mass. The CLy can also be provided per micron (*um* in the input file) if the simulated particle is an ion. In this case, its value represents the mean number of lesions that a single particle can produce along its track. Conversions between CLy units are given by equations 3.2 and 3.3,

$$CLy[(GyDa)^{-1}] = CLy[(Gycell)^{-1}] / (6.6 \times 10^8 \times Mbp_{tot}) \quad (3.2)$$

$$CLy[\mu m^{-1}] = CLy[(Gycell)^{-1}] \times 0.16 \times LET \times V_{nucl}^{-1} \quad (3.3)$$

where Mbp_{tot} is the total number of Mega-base pairs in the genome, V_{nucl} is the volume of the nucleus in μm^3 and the LET is given in $keV/\mu m$. Numerical factors come from the conversion of the *J* and *kg* units implicit in the *Gy* unit.

3.2.2.1 Initial damage

Although the calculation, for a given dose, of the mean number of lesions (\overline{CL}) follows the same general principles for all types of particles, i.e. it is determined by the dose and the CLy, there are important differences of implementation in each case. It is important to notice that the calculations of the actual values of the following stochastic variables, are carried out for each irradiated cell.

Photons

The cluster yield is converted into $(Gycell)^{-1}$ if necessary, then the \overline{CL} is given by

$$\overline{CL} = CLy \times Dose \quad (3.4)$$

Once \overline{CL} is determined, the actual number of lesions, CL , is extracted from a Poisson distribution with $\lambda = \overline{CL}$. The coordinates of the CLs are uniformly generated over the nucleus volume, and every CL position is checked in order to ensure that it hits an assigned voxel of the nucleus. If a CL does not hit the genome because it fell in an empty voxel (i.e. a voxel not assigned to any chromosome), its coordinates are re-calculated. If the Poisson sampling returns a zero value for CL, the cell is considered as spared and the code continues execution.

Light ions

First, the cluster lesion yield is converted to μm^{-1} if necessary, then the mean number of tracks/primaries, $\overline{N_{tr}}$, is calculated by means of equation 3.5 and an actual number of tracks (N_{tr}) is sampled from a Poisson distribution.

$$\overline{N_{tr}} = \frac{Dose \times S_{nucl}}{0.16 \times LET} \quad (3.5)$$

S_{nucl} in eq. 3.5 is the cross sectional area of the nucleus in μm^2 , whereas the *Dose* is given in *Gy*. The mean number of cluster lesions per track, $\overline{CL_{tr}}$, is computed as

$$\overline{CL_{tr}} = CLy \times H_{nucl} \quad (3.6)$$

where H_{nucl} is the thickness of the nucleus (in μm) as the irradiation is along the z -axis. For every track, the x - and y -coordinates of the shooting point are uniformly generated on the bottom of the ROI, and the actual number of lesions for each track, CL_{tr} , is Poisson-sampled. Then, the z -coordinate of the CL_{tr} lesions are uniformly sampled, one by one, along the line that defines the path of the particle, and the position of each CL is checked. If the CL does not hit the target, z is re-calculated. Every time $N_{tr} = 0$ the cell is considered as spared.

Heavy ions

For heavy-ion irradiation with a broad beam, the process to compute the effective number of lesions to be distributed within the cell nucleus is, essentially, the same as that for light ions. The only difference is that the x - and y -coordinates of the particle track define the core of the track instead of the coordinates of all the induced lesions. Some lesions remain along the core, while some others lie at the penumbra. The probability in each case is 50%[211]. The location of a CL that lies at the penumbra is obtained by sampling its z -coordinate uniformly over the line defined by the position of the track core and then, by computing the radial displacement with respect to this line. The radial shift is sampled from the probability function

$$P(r) = \begin{cases} h, & r < r_{min} \\ k\frac{1}{r}, & r_{min} < r < r_{max} \\ 0, & otherwise \end{cases} \quad (3.7)$$

where r is the radial shift, $r_{min} = 2 \text{ nm}$, r_{max} is given by eq. 3.1 and

$$k = \frac{1}{1 + \ln \frac{r_{max}}{r_{min}}}, \quad h = \frac{k}{r_{min}} \quad (3.8)$$

micro-beam

In the case of micro-beam irradiation, N_{tr} and the x - and y -coordinates are specified by the user. The actual number of cluster lesions and their positioning are calculated as described above for light and heavy ions, but taking into account the user-provided parameters. For instance, the actual x - and y -coordinates of the track core are sampled from a Gaussian distribution with the FWHM given in the corresponding .txt file.

3.2.2.2 Chromatin fragments

Every time a cluster lesion hits a voxel, the genomic position of the break is determined in accordance with the chromosome arm location. To this end, the index of the Mbp that was hit is sampled uniformly among the length (in genomic distance) of the chromosome arm containing the hit. Assumption 1 of the model establishes that each lesion produces two independent chromatin fragments. Thus, the information about the two new pieces is stored in an array. The process considers the case where a chromosome is broken in more than two parts by checking the stock of fragments in the array. The information of every chunk of chromatin is encapsulated in a class that saves the location of the break, the length of the fragment, the presence of centromere and telomeres and which arm or arms compose the piece. Most of this information is encoded in a string to make it easy the search for structural characteristics of the fragments.

3.2.3 Interaction models

Assumption 2 of BIANCA says that free ends in close proximity may undergo rejoining via proximity effects. This means that the probability of rejoining of two given ends depends on their initial distance, r . Currently, three functions of r model are implemented in BIANCA to model this mechanism:

$$P(r) = \begin{cases} e^{-r/\sigma}, & \textit{Exponential} \\ e^{-r^2/2\sigma^2}, & \textit{Gaussian} \\ 1 \textit{ if } r < \sigma \textit{ and } 0 \textit{ otherwise} & \textit{Step-like} \end{cases} \quad (3.9)$$

where σ is the adjustable typical interaction distance that controls the shape of $P(r)$ in each model. Mis-rejoining and un-rejoining lead to chromosome aberrations that are scored according to the criteria adopted in the experimental works chosen for comparison.

The focus of this thesis is on the *Exponential* and *Gaussian* models, as they constitute the original contribution of the project. Nevertheless, some remarks on the *Step-like* model will be presented too, since this model is still used in studies aimed to model cell death, with applications in tumor therapy.

3.2.3.1 Probability functions

Step-like probability

The first interaction model implemented in BIANCA consisted of a step-like function in which σ was typically chosen in the order of the linear dimensions of the chromosome territories. Experimental evidences suggesting that the repair of DNA damages may occur within the boundary regions of chromosome territories, motivated this choice [213]. Additionally, some works in the area propose the existence of DNA repair centers that, in combination with DNA damage-sensing proteins such as 53BP1, drive the interaction [204].

Free ends separated by a distance smaller than or equal to σ are assumed to rejoin with a 100% chance in this model. This implies that, eventually, all ends will rejoin with another ends or with their original partners. Thus, for low doses where isolated CLs may be induced in the nucleus, the only possible outcome is faithful repair. This situation tends to overestimate cell survival in these cases. Since there are evidences that not all chromosome exchanges are complete, that is, not all the free ends produced after irradiation will "find" a partner [206, 208, 207], the model was revisited by Ballarini and Carante and the f parameter was introduced [182]. f represents the probability for a free end to remain un-rejoined even in the presence of potential partners within the threshold distance σ . Regarding the interaction mechanism, the step function was still the probability function used, however, σ was fixed to the mean distance between adjacent chromosome territories, instead of being an adjustable parameter. Thus, the model parameters were only two in any case.

Good agreement with experimental data was achieved by the *Step-like* model in both modes (either adjusting σ or f), especially for the production of centric rings and dicentric aberrations [203, 15, 214, 16, 182]. The introduction of f improved the results on the so-called "excess acentric fragments" (EAF) category, in which interstitial and terminal deletions are included. Nevertheless, the yield of inter- and intra-arm intra-changes that represent the formation of interstitial deletions, centric and acentric rings, still showed important discrepancies in terms of the relative frequencies observed experimentally [215, 216]. For instance, despite the total EAF yield was consistent with experiments, interstitial deletions were underestimated while terminal deletions were overestimated.

On the other hand, while the ratio of dicentrics to centric rings (F-ratio) showed

the correct tendency, the production of acentric and centric rings was very similar. Such closeness is expected, for instance, in the case of random interactions as it was studied in [217]. However, experiments indicate that the ratio of intra- to inter-arm exchanges (G-ratio) is substantially larger than 1.

Exponential and Gaussian probability

The success of the *Step-like* model in reproducing inter-chromosomal to inter-arm intra-chromosomal exchange ratios can be attributed to the fact that the dimensions at which these interactions occur is in the order of microns, which is the typical order of *sigma* that fits well the absolute yields. Notwithstanding, this is also the reason why the distinction between inter-arm and intra-arm intra-changes is not satisfactory with the step-like model. Since these exchanges take place within a given chromosome territory, it is reasonable to think that they are more sensitive to proximity effects. Thus, a monotonically decreasing function appears to be a better candidate over the step-like one in order to further improve the outcomes of the model.

In the past, Kellerer and Rossi proposed their *Theory of Dual Radiation Action* (TDRA) in which they identified the DNA as the sensitive target for ionizing radiation. They proposed that IRs are able to produce sub-lethal lesions that have a certain probability to interact and create lethal lesions that may kill or inactivate the cell. They also suggested that such probability can be described by a Gaussian or an exponential function of the distance between sub-lethal lesions [133]. There are some authors that have implemented this type of functions to investigate the formation of radiation-induced aberrations. For instance, in [180], FISH data on human fibroblasts exposed to 2, 4 and 6 Gy of X-rays was compared with simulated yields of several types of chromosome aberrations (including acentric rings, centric rings and dicentrics). The probability of mis-rejoining and unrejoining was modelled with the function

$$P(r) = \frac{A \Sigma \exp(-r/r_0)}{N + A \Sigma \exp(-r/r_0)} \quad (3.10)$$

where r_0 , A and a third parameter Ω were adjustable parameters of the model and N represented the number of free ends available at each step of the simulation. The model, called CAS2 (Chromosome Aberration Simulator), was in good agreement with the data with the exception of acentric rings, which were underestimated. Nevertheless, they succeeded in obtaining G-ratios closer to the experimental values (they found G-ratios ~ 4). Among the drawbacks, commented by the authors, there is the fact that CAS2 required three adjustable parameters and the dose response curve was steeper than the experimental one. Furthermore, the approach was not extended to other cell lines and/or other radiation qualities.

Other works including the slowly-varying functions proposed by Kellerer and Rossi are [181, 218]. In [181], the probability function was

$$P(r) = \frac{1}{Z} \exp(-r^2/\sigma^2) \quad (3.11)$$

where Z and σ were adjustable model parameters. The authors tuned these parameters using Giemsa experimental data of human fibroblasts exposed to γ -rays and α -particles [20]. Then, they used the fitted model to predict yields of dicentric and rings after 1 GeV/u Fe-ions. They decided to include all the fragments created after exposure, even the smaller ones that are undetectable experimentally. These made difficult the interpretation of their results when compared with the experimental values. For instance, dicentric and centric ring yields suggest that their predicted F-ratio was about 30, while in the experimental data set is around 7. On the other hand, their G-ratio was considerably higher than the experimental one by three orders of magnitude.

In subsection 3.1.1, the probability functions implemented in BIANCA were presented. These include a Gaussian and an exponential dependence on the initial distance of the free chromatin ends. There are substantial differences with respect to the aforementioned works, despite the probability functions being similar. For example, the geometrical models are completely different in all cases. Additionally, the normalization factors in eqs. 3.10 and 3.11 include more than one adjustable parameter. In BIANCA, the probability of interaction is controlled by a single parameter, namely, the typical or characteristic interaction distance σ . The other model parameter, CLy, controls the formation of cluster lesions exclusively. In this way, the interpretation of the results becomes more transparent.

In BIANCA, the user has the freedom to define the interaction model to be used in the simulation. This is done through the *Model* keyword in the input file. The accepted values of the key are: *ExpModel*, *GaussModel* and *StepModel*. σ (or alternatively f) is passed through the *mydist* keyword. A float is expected. When the selected model is the step-like, the code sets the value of σ to the mean separation distance between two adjacent chromosome domains.

3.2.3.2 Formation of Aberrations

After the cluster lesions are placed within the nucleus, an array with the chromatin fragments created is passed to the *interaction* module of the code. At the beginning, all free ends in the array are considered *reactive*³ and all combinations of reactive ends are *valid*. Free ends created by the same CL are referred to as *conjugate*. To simulate the formation of chromosome aberrations, the algorithm follows the steps bellow:

³ In the Step-like model it is first decided, by means of f , which ends will remain unjoined. Those ends are considered "neutralized" from the beginning.

1. A valid pair of reactive ends is selected randomly, and their distance r is calculated.
2. A random number r_n is generated uniformly over the interval $[0, 1)$. If $P(r) < r_n$, the ends are considered rejoined. Else, the ends remain free and the same pair is excluded in future samplings, i.e. is no longer valid. If the pair is not conjugate, a second pair with the conjugates is formed. If the new pair is valid, step two is repeated.
3. The array of fragments is updated. If the fragment or fragments created do not have free or reactive ends, they are removed from the stock and stored in a new array for subsequent scoring.
4. All available ends are checked before a new loop starts. If it is found that a free end has no more partners to make a valid pair, it is considered as "neutralized". If both ends of a fragment result neutralized, the fragment is removed from the stock and stored for subsequent scoring.
5. Cycle

Step 4 is the realization of the implicit modelling of interstitial and terminal deletions. Neutralized free ends are nothing but un-rejoined ends that arise as a consequence of the sampling process. When a given end rejoins with another free end created at another location, immediately, its conjugate now has a smaller likelihood to undergo rejoining. This is more evident at low doses and low LET, for instance. Interstitial deletions occur when both ends of a fragment are neutralized, while a terminal deletion may happen to a fragment containing a telomere.

On the other hand, step 2 implies that, though not all exchanges are complete, they are prioritized since the conjugate pairs are evaluated at the same step as well. That is, the implicit modelling of un-rejoining is moderated by this bias in the sampling of the valid pairs.

3.2.4 Scoring

Once the interactions have been carried out, the resulting fragments are analyzed by the *Scoring* module. As the structural information of each fragment is encoded in a string, their classification is straightforward. It was decided to exclude from the analysis all the fragments of size smaller than the cutoff length $l_c = 3 \text{ Mbps}$ [9].

3.2.4.1 Classification of the fragments

Fragments are classified according to their shape and number of centromeres and telomeres before the chromosome aberrations are computed. The classification steps are the following:

1. All fragments with size $s < l_c$ are ignored.
2. Rings are separated from linear fragments by looking at the presence of the character "R" at the first and last position of each string describing the fragment structure.
3. Fragments containing centromeres are distinguished within each group (*Rings* and *Linear*). This is carried out by searching for the occurrence of letter "c" in the fragment's string code. In this manner, a distribution of the type of fragments based on their shape and number of centromeres is constructed.
4. A distribution of the number of centromeres vs. telomeres is also computed.

3.2.4.2 Chromosome aberration assessment

The code is ready to assess the type of chromosome aberrations after classification of the fragments, following the protocols and conventions used in experiments. All the information about the structure of chromosome exchanges is contained in the fragments's string code so, in principle, any type of exchange may be identified, even those involving more than two chromosomes (complex exchanges). The main categories of aberrations identified by Giemsa staining are dicentrics, centric rings, acentric rings and acentric fragments. To each dicentric and ring chromosome, an acentric fragment is associated as if they had come from a complete exchange. [Figure 16](#) shows a schematic representation of the aberration scoring adopted in BIANCA. When more than one CA category is presented, dicentrics are scored first, then centric rings and finally acentric rings. The remaining acentric fragments complete the scoring process. In order to compare with experiments that do not distinguish complex exchanges, the presence of more than two centromeres was translated into dicentrics or centric rings according to the shape of the fragment.

Other protocols like FISH or m-FISH could be included in the *Scoring* module with minor modifications of the code. This would be the case if the user is interested in following the fate of a specific portion of the genome. In such situation, the presence of a given chromosome or arm could be easily discriminated in the fragments. In experiments with telomere probes like in [\[219, 216\]](#) for instance, the existence of telomeres and/or centromeres is readily assessed and, in fact, a distribution of the linear fragments in terms of telomere and centromere content is determined by default (see [subsubsection 3.2.4.1](#)).

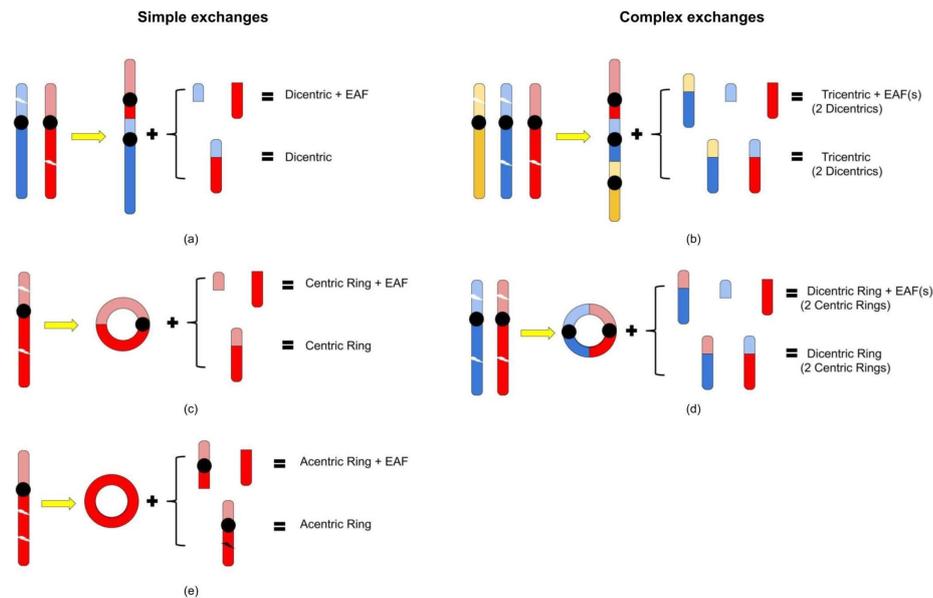


Figure 16 – Scoring scheme adopted in BIANCA. For simple exchanges (panels (a), (c) and (e)), each aberration must be accompanied by at least one linear fragment. (b) multicentric linear fragments must be associated to at least $n - 1$ linear fragments (where n is the number of centromeres) in order to be scored. (b) In the case of multicentric circular fragments, at least n linear fragments are expected. Complex exchanges as in (b) and (d) can be distinguished and scored separately or translated to simple exchanges. The number of "equivalent simple exchanges" is equal to the minimum number of linear fragments associated to the multicentric fragment. Acentric linear fragments smaller than 3 *Mbps* are not considered during the assessment of chromosome aberrations. When one or more exchanges are not complete (illustrated in the upper case in the brackets), one or more Excess Acentric Fragments (EAFs) may be counted as well. The order of the panels reflects the priority of the aberration categories, when there is more than one type of aberration presented.

3.3 Chromatin modelling using HiC data

In [section 1.2](#), our current understanding of chromatin organization in space was outlined. To date, the most robust and detailed information about chromatin 3D structure comes from the application of Chromosome Conformation Capture (3C) techniques such as HiC. In a nutshell, HiC maps provide statistical data on contact frequency between genomic regions up to the kilobase-pair resolution. Such contact maps cover all vs. all *in vivo* contact frequencies at a given resolution, including both intra- and inter-chromosome data. [Figure 17](#) illustrates the concept of a HiC map. In the example, is shown an all vs. all inter-chromosomal contact map between chromosomes A and B. The matrix of the lower panel can also be represented as a heat map to facilitate visualization. A comprehensive explanation of the HiC protocol can be found in [\[220\]](#).

Such a high quantity of information not only has helped to better understand the organization of chromatin in 3D space, but also has opened the possibility to im-

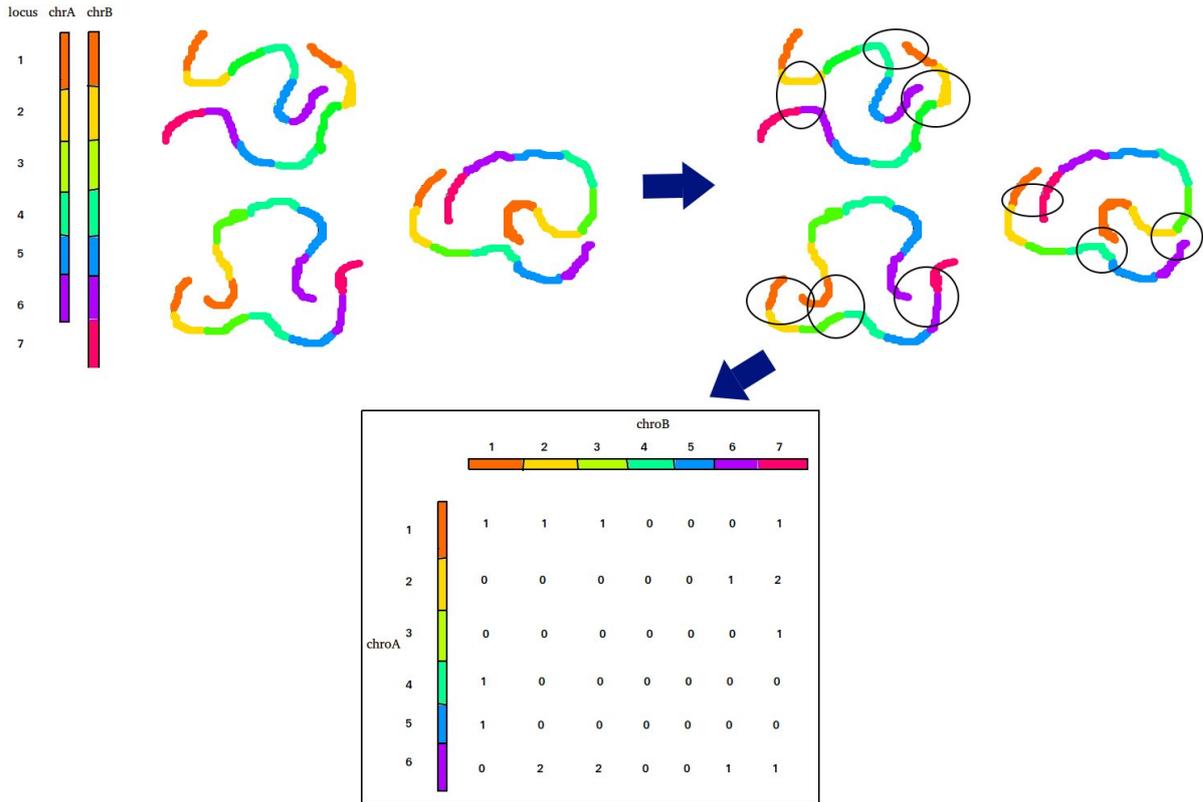


Figure 17 – Simplification of the HiC map concept. The entries of the matrix represent the number of times that a pair of chromosome loci were in close proximity to each other. In the figure, the construction of an inter-chromosome HiC map is illustrated.

prove the geometrical modelling of the entire genome [17]. For instance, from the polymer physics point of view, HiC maps offer a direct way to test polymer models used to create 3D conformations of the DNA [79, 221, 222, 223]. On the other hand, HiC data can also be used as an input for the so-called restraint-based models, which aim at creating population-averaged 3D structures from the data set by solving a computational optimization problem [224, 225, 226, 28]. This section will introduce an ongoing complementary project, which implements a restraint-based model in order to improve the geometrical modelling performed in BIANCA.

3.3.1 Restraint-based modelling with HiC data

HiC maps provide statistical data on the co-location frequencies of loci in the entire genome, and it has been observed that such interaction frequencies (IFs) are related to the inverse of their average relative spatial distance [227, 79, 228, 229]. In this way, IFs or the relative distances inferred from them, can be used as constraints to build 3D models that aim at locating the loci in space. To do so, the 3D model is subjected to forces seeking consistency with the interaction probabilities that come from the data. Thus, the models can be generated by minimizing a scoring function that is proportional to the violation of the imposed restraints. Typically, the steps to construct the 3D structures consist of

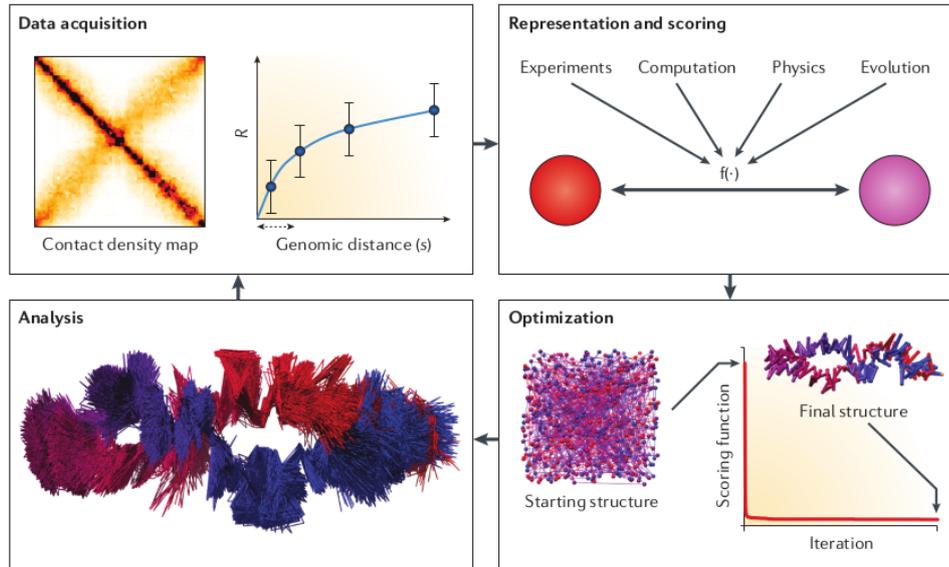


Figure 18 – Iterative and integrative process for model building. The iterative process consists of data acquisition, model representation and scoring, model optimization and model analysis. Source: Dekker *et al.*. Nat. Rev. Gen. (2013) [17].

1) selecting the appropriate data set for the resolution and genomic regions of interest, 2) generating the models and computing the scoring function, 3) applying an optimization method, and 4) comparing the generated structures against the data (see Figure 18). These steps are repeated until a consensus structure is achieved, usually, when a local minimum (or maximum) of the scoring function is reached.

Several restraint-based methods have been developed in the recent years to address the problem of 3D modelling. Some of them require a function to translate interaction frequencies into average distances, while others do not. For instance, Varoquaux *et al.* implemented four methods based on statistical analysis to infer the 3D structure of the genome [225]. In two of them, the IFs were modelled as Poisson random variables and the optimization process aimed at maximizing the likelihood to observe those IFs. A third method directly inferred the coordinates of loci from their pairwise Euclidean distances, while the last method relied on the assumption that, if $IF_{ij} > IF_{kl}$ (IF between loci i and j), then $d_{ij} < d_{kl}$, in order to derive an objective function to optimize.

ChromSDE is another method which applies semi-definite programming techniques to find the best structure fitting the observed data, and uses golden section search to find the correct parameter for converting the contact frequency to spatial distance. Interestingly, with this method, the authors proved that the conversion factor from IFs to spatial distances changes with scale [224]. Additionally, LorDG (Lorentzian 3D Genome) is also a restraint-based method able to reconstruct 3D genome structures using HiC data and a Lorentzian function as the scoring function [28]. It represents the chromosome/genome

structures as a series of beads with each bead representing a chromosome fragment of a specific length (e.g. 1 Mbp or 500 kbp). The initially random position in space of each bead is iteratively corrected by a gradient-ascend method until convergence to a local maximum of the scoring function. Currently, LorDG is one of the most successful methods to produce geometrical models of the entire genome as it is relatively fast, robust and computationally efficient at processing data. For these reasons, it was decided to build upon this method within this project as a means to further improve BIANCA. The main aspects of LorDG will be discussed in the following sections, while the current state of the fortran implementation of the method performed in this thesis work, will be presented in [section 4.5](#).

3.3.2 The LorDG method

3.3.2.1 Algorithm

The method starts by reading the intra- and inter-chromosome HiC data. Then, the mean adjacent interaction frequency IF_{adj} is computed, that is, the average of all the intra-chromosome contacts for which $|i - j| < 1$. At the same time, the overall mean interaction frequency IF_{ave} and the IF_{max} are determined as well. Next, the data set is completed by adding the adjacent contacts that were missing in the original HiC data set, setting their value to IF_{adj} . Once the data has been pre-processed this way, the interaction frequencies are translated to spatial distances. A function commonly used to convert IF into Euclidean distances is of the form,

$$wd_{ij} = IF_{ij}^{-\alpha} \quad (3.12)$$

where wd_{ij} is the theoretical or expected "wish" distance between i and j , IF_{ij} is the corresponding interaction frequency and α is a conversion factor [[230](#), [226](#), [225](#), [224](#)]. Since the conversion function plays an important role in the quality of the reconstructed models, some methods try to optimize α instead of using a fixed value [[230](#), [225](#), [224](#)]. In LorDG, α is searched over the interval $[1.0 - 3.0]$ in steps of 0.1, as this range covers most of the previously used values [[230](#), [226](#), [225](#)].

The model is initialized by randomly distributing the beads within an expandable box of unit side. Then, the method aims at rearranging the beads within the box in such a way that their relative distances satisfy as much wish distances as possible. As mentioned, LorDG uses a Lorentzian function ([Equation 3.13](#)) as the scoring function in order to measure the satisfaction of the restraints. In [Equation 3.13](#), Γ determines the width of the curve, while each wd_{ij} defines a local peak of the function. Additionally, d_{ij} represents the distance between beads i and j in the current model. The adjacent contact

terms were separated from the rest as they dominate due to the natural disposition of contiguous loci, which make them have significantly higher IF values.

$$f = \sum_{|i-j|=1} \frac{\Gamma \cdot IF_{max}}{\Gamma^2 + (d_{ij} - wd_{ij})^2} + \sum_{|i-j|\neq 1} \frac{\Gamma \cdot IF_{ij}}{\Gamma^2 + (d_{ij} - wd_{ij})^2} \quad (3.13)$$

Thus, the optimization process tries to maximize f by reducing the difference between d_{ij} and wd_{ij} . This function offers a number of advantages over other alternatives like the traditional squared-error function ($\sum IF_{ij}^2 (d_{ij} - wd_{ij})^2$). For instance, its derivative is very small almost everywhere, except along the two slopes of the curve centered at the wish distances. Furthermore, the violation of restraints with low IF does not dramatically affect the value of the function, which implies that the method tends to reward the satisfaction of consistent restraints (those with higher IF s) instead of severely penalizing the violation of the inconsistent ones.

At every step, the code computes the relevant distances between loci, that is, between pairs of loci for which the corresponding constraint is present, and input them to [Equation 3.13](#). Also, the gradient of the function with respect the spatial coordinates of each bead is calculated by means of [Equation 3.14](#),

$$\nabla_i f = \sum_{|i-j|=1} \frac{2 \cdot \Gamma \cdot IF_{max} (wd_{ij} - d_{ij})}{d_{ij} \cdot [\Gamma^2 + (d_{ij} - wd_{ij})^2]^2} (\mathbf{r}_i - \mathbf{r}_j) + \sum_{|i-j|\neq 1} \frac{2 \cdot \Gamma \cdot IF_{ij} (wd_{ij} - d_{ij})}{d_{ij} \cdot [\Gamma^2 + (d_{ij} - wd_{ij})^2]^2} (\mathbf{r}_i - \mathbf{r}_j) \quad (3.14)$$

where ∇_i means the gradient of f with respect to \mathbf{r}_i (the coordinate vector of bead i). Then, the gradient-ascent method is applied in order to correct the position of the beads in the current model. The gradient-ascent method consists of taking steps in the direction of the gradient of the objective (i.e scoring) function to displace the coordinates of the beads until a local maximum of f is achieved, that is,

$$\mathbf{r}_i \rightarrow \mathbf{r}_i + \gamma \nabla_i f \quad (3.15)$$

where γ is the size of the step. The step size is fixed along the optimization ($\gamma = 0.1$) but within an iteration, it is allowed to vary if at some point the scoring function ceases to increase. In that case, the new value of γ is determined by line search using Wolfe conditions.

In summary, the optimization performs the following steps:

1. The current score and gradient of f is calculated (Equation 3.13 - 3.14)
2. The new positions are computed using Equation 3.15 with the default value of γ . The score of the model with the new positions is assessed and compared with the current value.
If: $new\ score > current\ score$
the new positions are accepted and the model is updated.
else:
A new step size is computed by line search⁴. This is accomplished by doing $\gamma \rightarrow \gamma/2$ iteratively, while $(new\ score - current\ score) < 0.5\gamma\|\nabla f\|^2$ and $\gamma > \epsilon$. Once the new step size is found, the model is updated accordingly.
3. Steps 1 and 2 are repeated until the scoring function converges to a maximum ($\|\nabla f\| < \epsilon \cdot f$) or a maximum number of iterations is reached.
4. The structure is tested (see below).

When the code runs with a given data set for the first time, the value of α in Equation 3.12 is searched over the interval [1.0, 3.0] in steps of 0.1 as mentioned. To do so, three structures for each value of α are generated and tested. Then, the best α is selected as the conversion factor for that particular HiC data set (i.e. cell line and resolution). Additionally, Γ is set equal to the average wish distance. In this way, several consensus structures with the selected α can now be created and studied. For instance, Trieu and Cheng [28] analyzed the relative location of four 3D-FISH probe positions (L1,L2,L3,L4) in chromosome 14 of the cell line GM06990 with models generated by LorDG. According to [231], L1 and L3 tend to be closer even though L2 lies in the middle of both probes. Additionally, L2 tends to be closer to L4, despite L3 being in between. The distances L1-L3, L2-L3 and L2-L4 were measured in the reconstructed models, showing that L2-L3 was consistently larger than L1-L3 and L2-L4 in all cases. LorDG is available at: <https://missouri.box.com/v/LorDG>.

3.3.2.2 Input data

Raw HiC data can be found in the [GEO Accession viewer](#) for several human and mouse cell lines. The data set used by Trieu and Cheng [28] and also in this project, correspond to that submitted by Rao *et al.* [232] regarding the GM12878 and Lieberman *et al.* [79] for the GM06990 human lymphocyte cell lines. The GEO data set of Rao *et al.* provide not only the raw reads at several resolutions (from 5 *kbp*s to 1 *Mbp*s) of

⁴ The first inequality is the Wolfe condition in which $\|\nabla f\|$ is the norm of the gradient of all the beads added up, while the second condition uses the overall convergence criteria to decide when the scoring function has stopped to increase.

the whole genome, but also normalization vectors that allow correction of the data according to different strategies. Such corrections are necessary because usually HiC data suffers from several sources of biases and experimental noises due to chromatin accessibility, nucleosome occupancy, alignability or restriction site density at a locus, among other factors [232, 223]. The normalization vectors included in the Rao *et al.* data are those for Vanilla Coverage (VC), VC square rooted (VCSQRT) and Knight-Ruiz (KR) normalization. These methods assume one-dimensional biases, which means biases that are a function of the locus itself and influence the contact frequency between that locus and any other locus. Thus, if there is a bias towards observing, say, locus i , then all entries of IF_{ij} (for any j) will tend to be higher as if loci i and j would have interacted more.

In VC normalization, a row specific R_i term is calculated by totaling the counts in row i and taking the inverse. Similarly, the inverse of a column specific C_j term is determined for column j . Then, each raw entry of the HiC map is normalized by doing $[IF_{ij}] \rightarrow R_i[IF_{ij}]C_j$. When the square root of the R and C vectors is used instead of the original values, the method is called VCSQRT normalization. VCSQRT is preferred over VC because the latter tends to over-correct the raw IF s. Also, as discussed in [232], VCSQRT provides values that are very close to those of much more sophisticated and computationally intensive algorithms of matrix balance. VC normalization can be viewed as a single iteration of the classic matrix balancing algorithm of Sinkhorn and Knopp [233].

KR normalization is a generalization of the Sinkhorn and Kopp method based on inner-outer iteration schemes [234]. It combines the inexact Newton's method with a pre-conditioned conjugate gradient iteration, achieving faster convergence since it demands less matrix-vector operations. KR normalization always results in a balanced matrix as long as the original matrix is not too sparse, which typically happens in very high resolution maps ($\sim 5 - 10$ *kbps*).

3.3.2.3 Testing the structures

An indicator commonly used to measure similarity between 3D structures is the root mean square error (RMSE) given in Equation 3.16. In order to compare two different structures, they need to be at the same scale. This can be partially achieved by scaling all the pairwise distances in a way that both models have the same average pairwise distance.

$$RMSE = \sqrt{\frac{1}{n} \sum (d_{ij} - d'_{ij})^2} \quad (3.16)$$

In Equation 3.16, n is the number of pairwise distances, and d_{ij} and d'_{ij} are the distances between loci i and j in each model.

Another test that is performed over the structures to assess validity of the results, is the Spearman correlation test. This test consists of calculating the Pearson coefficient between the rank values of the reconstructed distances and the *IFs* or wish distances. The Spearman correlation coefficient (ρ) is the criteria used to select the α value. Models with the highest ρ scores are considered best.

4 Results

During this PhD project, the BIANCA code was re-designed and expanded. Technical improvements were achieved as a result of such re-structure but also, new capabilities that make it possible to further investigate the mechanisms behind the formation of chromosome aberrations were incorporated. This chapter highlights those technical improvements with respect to the previous version, and the results of the works carried out during the PhD project using the new code.

The new *Geometry* module is able to model chromosome sub-domains that allows identification of the arms, telomeres and centromeres. As a consequence, the *Scoring* unit was also upgraded in order to discriminate chromatin re-combinations with more detail. Into the irradiation setup, the biophysical model was expanded to include exponential and Gaussian probability functions that are, presumably, linked to different modes of diffusion, namely, Brownian motion and active transport of the free ends, respectively.

After the code was upgraded, a systematic study of the formation of CA in human lymphocytes and fibroblasts was performed. Experimental data on a wide range of LET values and doses for both cell lines was compared against simulations. The performance of each model was analyzed in terms of absolute aberration yields and aberrations ratios in a way that their main advantages and drawbacks could be elucidated. The new version of BIANCA was also applied in a number of collaborations that combined both experiments and simulations.

Finally, the current state of the ongoing project aiming at including HiC data in the BIANCA framework will be presented. The research carried out during this PhD project led to some publications in indexed journals [19, 21, 24, 235].

4.1 BIANCA upgrade

BIANCA has evolved along the time both in structure and capabilities. This has allowed to approach the simulation of the formation of chromosome aberrations in a consistent and more detailed fashion throughout the time. During this project, a new version of BIANCA has been coded, translating the original *FORTRAN77* version into a modern *fortran2008* one. This effort represented more than a stylistic improvement. since the advantages of the modern fortran could now be appreciated. Among such advantages there is the possibility to define classes, pointers and external procedures. Also, it is possible

to split the code into modules that allow for code re-usability and easier maintenance and extension. Furthermore, inter-operability with C is a feature that could be of great value for the future of the code as well. The first stage of the PhD work was devoted to re-designing the code. Results on this matter are worth to be mentioned in this chapter, as they were not trivial and represented a notorious improvement in performance and capabilities.

4.1.1 Technical aspects

The original version of BIANCA consisted of a single, nearly ten thousand lines source file. Declaration of variables and implementation of procedures followed the F77 standard mainly. The execution flow allowed to run the simulations for only one dose for each calling of the main and, though in principle it was possible to loop over some components of the algorithm, the control parameters to, say, create different geometries were not set. Additionally, the typical execution times were given in hours for single-dose runs with 5000 – 10000 cells. A list of the technical improvements and their main influence in performance follows.

- The source code was rewritten from scratch. The different components of the simulation were distributed into modules implementing separately the geometry, irradiation, interactions and the scoring. Each module of the new source file follows the F2008 standard. This enables the use of modern compilers with faster libraries, which means faster execution of the code. Also, debugging, maintenance and extension of the code become easier.
- The *Geometry* module implements a new method to construct the cell nucleus. Chromosome arms are modelled explicitly, which means that the geometry is more realistic now. Structures can be saved for future simulations reducing the overall execution time.
- The *Scoring* module implements new classes and procedures that allow to record the aberrations according to several protocols at the same time (Giemsa, FISH, Telomere probes).
- Single-dose and multi-dose simulations with one or several cell configurations can be readily carried out. This is controlled by the user's input file.
- Automatic generation of scripts for post-processing of the outcomes was included.
- The user's input file controls the selection of models eliminating the need of re-compilation. This is a consequence of the use of pointer procedures.

As a result of this upgrade, the execution times decreased down to a few seconds for single-dose simulations and minutes for multi-dose runs.

4.1.2 Model improvements

Once the code was ready, the *Interaction* module was expanded by implementing an interface to virtually any function of r (initial distance of the chromosome fragment free-ends). Thus, the exponential and Gaussian probability function were included. The sampling of the candidate free-ends to rejoin was conceived in a way that allows implicit modelling of un-rejoining instead of leaving this mechanism to be controlled by another parameter. Such conception encompasses with the hypothesis that all the free ends follow the same dynamics so, un-rejoining is a consequence rather than a different aspect of chromatin mobility. On the other hand, the irradiation setup was also expanded to include micro-beams. Thus, the model improvements of the new version of BIANCA are the following.

- Gaussian and exponential probability functions to model free-end re-joining. This functions are consistent with active transport channels or free diffusion modes for the chromatin, respectively.
- Implicit modelling of the production of terminal and incomplete interstitial deletions (un-rejoined ends).
- Micro-beam irradiation included.

4.2 Proximity effects in chromosome aberration induction by low-LET ionizing radiation

The model was applied to simulate chromosome aberration induction in low-LET irradiated lymphocytes (doses: 1–4 *Gy*) and fibroblasts (1–6.1 *Gy*). Good agreement with experimental yields of dicentrics and centric rings, and thus their ratio ("F-ratio"), was found by both the exponential model (with $\sigma = 0.8 \mu m$ for lymphocytes and $0.7 \mu m$ for fibroblasts) and the Gaussian model (with $\sigma = 1.1 \mu m$ for lymphocytes and $\sigma = 1.3 \mu m$ for fibroblasts). While the former also allowed reproducing dose-responses for excess acentric fragments, the latter substantially underestimated the experimental curves. Both models provided G-ratios (ratio of acentric to centric rings) higher than those expected from randomness, although the values calculated by the Gaussian model were lower than those calculated by the exponential one. For lymphocytes the calculated G-ratios were in good agreement with the experimental ones, whereas for fibroblasts both models substantially underestimated the experimental results, which deserves further investigation. This work

suggested that, although both models performed better than a step model (which previously allowed reproducing the F-ratio but underestimated the G-ratio), an exponential function describes proximity effects better than a Gaussian one. These results will be presented in detail in [subsection 4.2.1](#) (lymphocytes) and [subsection 4.2.2](#) (fibroblasts).

4.2.1 Human lymphocytes

4.2.1.1 Dose-response curves

[Figure 19](#) reports yields (mean number per cell) of different aberration types (dicentric, centric rings and excess acentric fragments, as well as their sum) for human lymphocytes exposed to different doses of ^{137}Cs γ -rays. The full symbols represent experimental data taken from the literature [18], whereas the empty symbols, connected by lines to guide the eye, are simulation outcomes obtained by the exponential model with $\sigma = 0.8 \mu\text{m}$ and $\text{CL} = 3.1 (\text{GyCell})^{-1}$. While σ mainly controls the aberration ratios (dicentric to centric rings and acentric rings to centric rings), the CL yield mainly controls the absolute yields of the various aberration categories. Interestingly, $0.8 \mu\text{m}$ is the same value found in [180] for X-irradiated human fibroblasts. Concerning the CL yield, $3.1 (\text{GyCell})^{-1}$ is very small (less than 10%) compared to the consensus value of about 40 DSB per Gy per cell. In principle, this may imply either that only a small fraction of all initial DSBs are candidates for large-scale misrejoining leading to chromosome aberrations, or that all initial DSBs are candidates, but restitution "eliminates" most free-end pairs. This issue is still an open question in radiobiology, and both hypotheses have been tested in the modelling works available in the literature. However, it is worth mentioning that in a previous work [214], CL yields for different radiation qualities showed good agreement with yields of kilobase-pair (kbp) size DNA fragments, suggesting that DSB clusters at the kbp scale, possibly in addition to other levels of clustering, may play a relevant role in the induction of chromosome aberrations and cell death. This would speak in favour of the first hypothesis, i.e. the existence of a sub-class of severe DSBs that are candidates for large-scale mis-rejoining, as also suggested by old and recent experimental works (e.g. [236, 191]).

Of course the considered experimental data set represents only a small, rather old, fraction of the experimental data available in the literature. However, we chose this particular paper because, in addition to photon data, it also reports data on intermediate- and high-LET radiation obtained by the same group of authors applying the same experimental conditions and protocols as explained in [section 4.3](#). This was very useful for the extension of our modelling approach to higher LET radiation. Another reason for choosing this particular work is that, for each considered radiation quality and dose, it reports both the F-ratio and the G-ratio, which are very useful to test the assumed

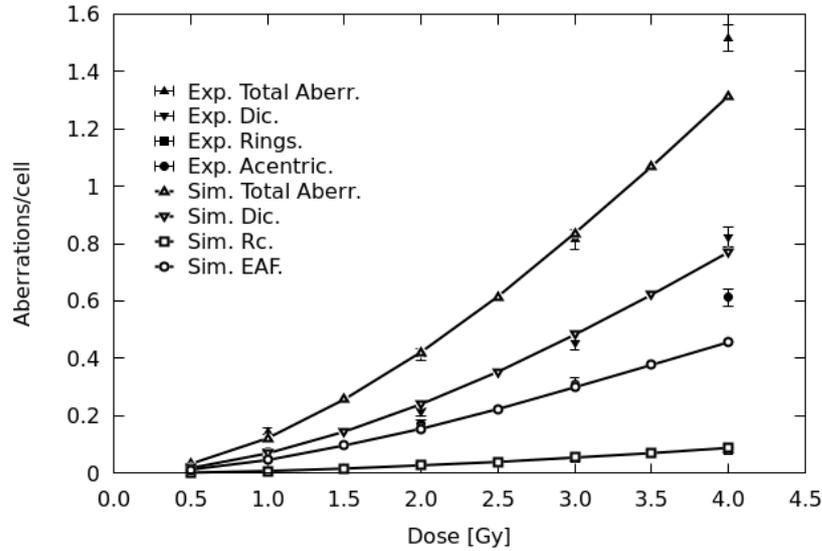


Figure 19 – Aberrations in lymphocytes calculated by the exponential model. Full symbols are experimental values taken from [18]; the error bars represent one standard deviation from the mean, calculated from the raw numbers (both ^{137}Cs γ -ray data sets pooled) assuming Poisson statistics. Empty symbols, connected by lines to guide the eye (from top to bottom: total aberrations, dicentric, excess acentric fragments and centric rings), are results of simulations performed by the exponential model, with $\sigma = 0.8 \mu\text{m}$ and $\text{CL} = 3.1 (\text{GyCell})^{-1}$. Source: Tello *et al.* DNArepair (2017) [19].

distance-dependence of the free-end rejoining probability (see below).

In addition to the doses considered in the experimental work, which were 1, 2, 3 and 4 Gy , simulations were also carried out at other doses (0.5, 1.5, 2.5 and 3.5 Gy), for which therefore the simulation outcomes are full predictions. With the exception of excess acentric fragments at 4 Gy , for which the simulation outcome was significantly lower than the experimental value, the simulated yields were within the experimental error bars for each considered dose and aberration category. At least for dicentric, it can be useful to fit the simulation outcomes at the various considered doses by a linear-quadratic function of the form $Y(D) = \alpha D + \beta D^2$, where $Y(D)$ is the dicentric yield at dose D and α and β are the well-known linear and quadratic parameter, respectively. In the range 0 – 4 Gy , such parameters resulted to be $\alpha = 0.026$ and $\beta = 0.044$. For comparison, we also performed a best fit of the experimental data, which provided $\alpha = 0.016$ and $\beta = 0.046$.

The same kind of simulations were performed applying a Gaussian model. Figure 20 reports results obtained with $\sigma = 1.1 \mu\text{m}$ and $2.3 \text{CL}(\text{GyCell})^{-1}$, compared with the same experimental data reported in Figure 19. Calculating the linear and quadratic coefficients for the simulation outcomes with the Gaussian model, we obtained $\alpha = 0.025$ (experimental value: $\alpha = 0.016$) and $\beta = 0.038$ (experimental value: $\beta = 0.046$). While

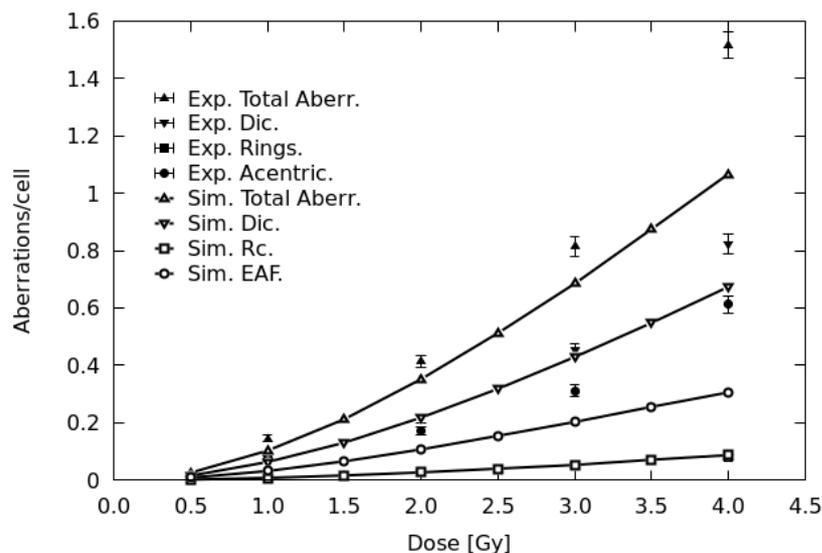


Figure 20 – Aberrations in lymphocytes calculated by the Gaussian model. Like in Figure 19, full symbols are experimental values taken from [18], with standard deviations calculated assuming Poisson statistics. Empty symbols, connected by lines to guide the eye (from top to bottom: total aberrations, dicentric, excess acentric fragments and centric rings), are results of simulations performed by the Gaussian model, with $\sigma = 1.1 \mu\text{m}$ and $\text{CL} = 2.3 (\text{GyCell})^{-1}$. Source: Tello *et al.* DNArepair (2017) [19].

good agreement between simulations and data was found for dicentric and centric rings, simulated yields of excess acentric fragments were much lower than the experimental yields, suggesting that, for low-LET irradiated lymphocytes, an exponential model can better reproduce the data than a Gaussian model. This may be explained taking into account that, at the small distances that are involved for intra-arm exchanges, which represent an important contribution of excess acentric fragments, the exponential function decreases more rapidly than the Gaussian function; this should allow better taking into account the observed bias for intra-arm aberrations with respect to inter-arm and inter-chromosome aberrations. More generally, the exponential model combines a fast decay of the (mis-) rejoining probability at small distances, where diffusion of free-ends may dominate, with a slower decay at larger distances, where also other mechanisms (e.g. some form of active transport) may play a role. To compare the performance of the two applied models, chi-square tests were performed, from which the chi-square values for the exponential model were found to be significantly smaller than those for the Gaussian model for all considered aberration categories.

4.2.1.2 Aberration ratios

Calculated values of F-ratio (ratio of dicentric to centric rings) and G-ratio (ratio of acentric rings to centric rings) at the doses considered in the experiments are reported

in Table 4, with the corresponding experimental values for comparison. For both models, with a few exceptions the calculated F-ratios were in good agreement with the experimental ones. Concerning the G-ratio, the simulation outcomes at 2, 3 and 4 Gy were significantly higher than the experimental ones. However, the values obtained in [18] are among the lowest ones reported in the literature, and other authors reported higher values. For instance, applying a combination of pan-centromeric and pan-telomeric probes to gamma-irradiated human lymphocytes, in [216] the authors found G-ratios of 3.4, 2.2 and 2.9 after doses of 2, 4 and 6 Gy, respectively, with a mean value of 2.8. It cannot be excluded that some of the acentric fragments classified as "linear" in [18], in which telomeric probes were not used, derived from two (un-rejoined) chromosome breaks induced in the same arm, which would increase the G-ratio.

Table 4 – Experimental ("Exper.") F- and G-ratios for human lymphocytes exposed to different doses of ^{137}Cs γ -rays compared with values calculated by the exponential model ("Exp.") or the Gaussian model ("Gaus."). The errors associated to the experimental values represent standard deviations calculated applying Poisson statistics to the raw numbers reported in the experimental paper ([18], both γ -ray data sets pooled).

Dose(Gy)	F-ratio			G-ratio		
	Exper.	Exp.	Gauss	Exper.	Exp.	Gauss
1.0	13.2 ± 6.1	10.0	8.0	2.4 ± 1.3	1.6	1.3
2.0	7.7 ± 1.6	9.5	8.1	0.9 ± 0.3	1.6	1.4
3.0	8.8 ± 1.5	8.4	8.1	0.8 ± 0.2	1.6	1.5
4.0	9.4 ± 1.3	8.8	7.7	0.8 ± 0.2	1.7	1.4

Since the G-values calculated in the present work are intermediate between those reported in [18] and those reported in [216], this suggested that an exponential distance-dependence for the fragment rejoining probability can better model proximity effects with respect to a step function, which could reproduce yields of dicentrics and centric rings, and thus the F-ratio [196], but could not reproduce the deviation from randomness often observed experimentally for the G-ratio. Since the simulations were performed assuming that fragments smaller than a threshold of 3 Mbp were not visible, possible changes in the results after eliminating this threshold were investigated. While for dicentrics and centric rings the calculated yields did not change significantly, a significant increase in excess acentric fragments was found. However, such increase was slight for all considered doses: for instance, after 4 Gy the yield of acentric rings calculated by the exponential model increased by a factor ~ 0.08 , and that of acentric linear fragments increased by a factor ~ 0.04 . Very similar results were found for fibroblasts exposed to 6 Gy γ -rays. With the exponential model, both F and G showed a dependence on the value of σ . As expected, the F-ratio increased with increasing σ ; concerning G, the dependence on σ was

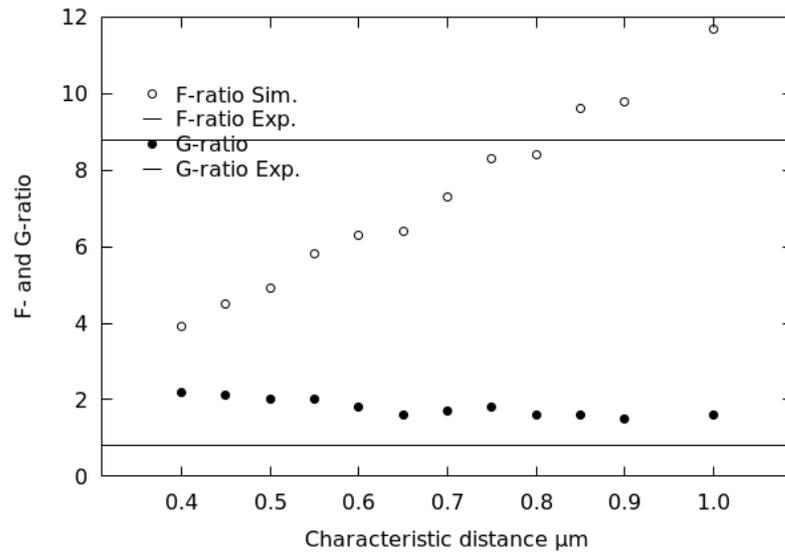


Figure 21 – F- and G-ratio for lymphocytes irradiated by 3 Gy ^{137}Cs γ -rays. Points represent values of F (open symbols) and G (closed symbols) calculated with the exponential model for different values of σ , whereas the upper and lower solid lines represent the (mean) experimental value of F and G, respectively, according to [18]. Source: Tello *et al.* DNArepair (2017) [19].

less pronounced than the dependence shown by F, but there was anyway a tendency to increase with decreasing σ .

This is shown in Figure 21, where calculated values of F and G are plotted for different values of σ in the range 0.4 – 1.0 μm (in steps of 0.05 μm) in the case of a γ -ray dose of 3 Gy. For each considered σ value, the CL yield was adjusted to the experimental yield of dicentrics. Although this allowed reproducing the experimental dicentric yields for all considered values, σ values lower than 0.8 μm lead to an underestimation of F, whereas σ values higher than 0.8 μm lead to an overestimation.

4.2.2 Human fibroblasts

4.2.2.1 Dose-response curves

Afterwards, the approach was extended to gamma-irradiated human fibroblasts, to investigate the role played by the cell type and/or the cell nucleus shape and dimensions. Analogous to Figure 19, Figure 22 reports yields of dicentrics, rings and excess acentric fragments (as well as their sum) for human fibroblasts exposed to different doses of ^{137}Cs γ -rays. Closed symbols represent experimental data on gamma-irradiated AG1522 normal human fibroblasts [20], whereas open symbols, connected by dashed lines to guide the eye, are simulation outcomes obtained with $\sigma = 0.7 \mu\text{m}$ and $\text{CL} = 4 (\text{Gycell})^{-1}$

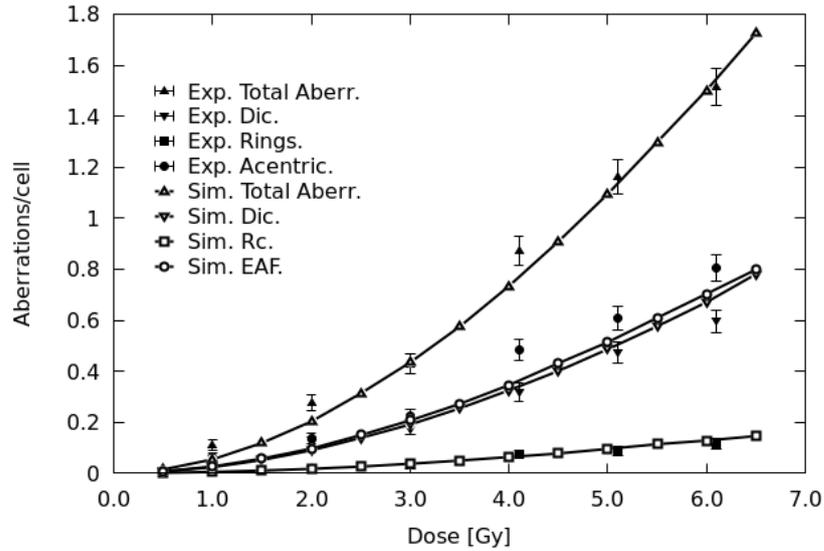


Figure 22 – Aberrations in AG1522 fibroblasts calculated by the exponential model. Full symbols represent experimental values taken from [20] ("Day 0" data set, with error bars representing one standard deviation calculated from the raw numbers assuming Poisson statistics), whereas the empty symbols, connected by lines to guide the eye (from top to bottom: total aberrations, dicentrics, excess acentric fragments and centric rings), are results of simulations performed by the exponential model (with $\sigma = 0.7 \mu\text{m}$ and $\text{CL} = 4.0 (\text{GyCell})^{-1}$). Source: Tello *et al.* DNAREpair (2017) [19].

following adjustment by trial-and-error to the experimental data. Like for lymphocytes, while σ mainly controls the aberration ratios, the CL yield mainly controls the absolute yields of the various aberration categories. For the sake of completeness, it is worth mentioning that the "ring" experimental category also included some large acentric rings (i.e., where the "hole" was clearly visible) in addition to centric rings. However, also other authors (e.g. [181, 237]) compared calculated yields of centric rings with the "ring" category reported in [20]. Like for lymphocytes, in addition to the doses considered in the experimental work (1, 2, 3, 4.1, 5.1, and 6.1 Gy), full predictions were also performed at other doses (0.5, 1.5, 2.5, 3.5, 4.5, 5.5, and 6.5 Gy).

Good agreement was found with the experimental data on dicentrics and rings, for which the simulation outcomes were within the experimental error bars with the exception of dicentrics at the highest dose, for which the calculated yield was higher than the experimental one. This indicates that, at least for low-LET radiation, an exponential function can reproduce the proximity bias of centric rings with respect to dicentrics not only in lymphocytes but also in fibroblasts. The (slightly) smaller σ value used for AG1522 fibroblasts with respect to lymphocytes ($0.7 \mu\text{m}$ instead of $0.8 \mu\text{m}$) is consistent with the reduced interphase chromatin mobility observed in these cells with respect to other cells [238]. Concerning the CL yield, the difference with respect to the value used for

lymphocytes (4.0 instead of $3.1 \text{ CL}(\text{Gycell})^{-1}$) may depend on the different characteristics of the target cell, including the fact that the volume of AG1522 nuclei ($628 \mu\text{m}^3$) was much larger than that of lymphocyte nuclei ($113 \mu\text{m}^3$). However, the ratio between the two nuclear volumes does not imply an equal ratio between the two CL yields, because AG1522 cells are less radiosensitive than lymphocytes.

Concerning excess acentric fragments, which are the sum of interstitial and terminal deletions, with the exception of the data at 4 Gy a reasonable agreement was found with the experimental yields of total (i.e., interstitial plus terminal) deletions, for which the calculated values were significantly but slightly lower than the experimental ones.

However, when considering interstitial and terminal deletions separately, the simulated yields of interstitial deletions were substantially lower than the experimental ones, and thus the calculated G-ratio, which was around 2 , was smaller than the experimental one, which was between 6 and 7 (see Table 5). From the literature, the G-ratio seems to be higher in fibroblasts than in lymphocytes; the possible reasons, which may be related to chromatin mobility and/or interphase chromatin organization, are still under investigation. However, the fact that the G-ratio calculated by the exponential model was significantly higher than the ~ 1.2 value expected from randomness, indicates that also for fibroblasts an exponential model can better reproduce proximity effects with respect to a step model.

Like for lymphocytes, also for fibroblasts a Gaussian model was also tested against the same data set. Figure 23 shows that parameter values of $\sigma = 1.3 \mu\text{m}$ and $\text{CL} = 2.0 (\text{Gycell})^{-1}$ allowed obtaining good agreement with the experimental yields of dicentric and rings, and thus with the F-ratio. However, the calculated yields of excess acentric fragments, compared with the experimental category of interstitial plus terminal deletions, were substantially lower than the experimental ones. Like for lymphocytes, also for fibroblasts chi-square tests were performed to compare the exponential and the Gaussian model. While for dicentric and rings the chi-square values calculated for the two models were similar, for excess acentric fragments and total exchanges the chi-square values for the Gaussian model were about a factor of 10 larger than those for the exponential model.

4.2.2.2 Aberration ratios

Table 5 reports values of F-ratio and G-ratio (experimental and calculated according to either the exponential model, or the Gaussian model) for gamma-irradiated human fibroblasts, at the doses considered in [20]. Concerning the F-ratio, the values obtained by both models were in good agreement with the experimental ones. Good agreement with dicentric and (centric) rings was also obtained by a step model adopted in previous

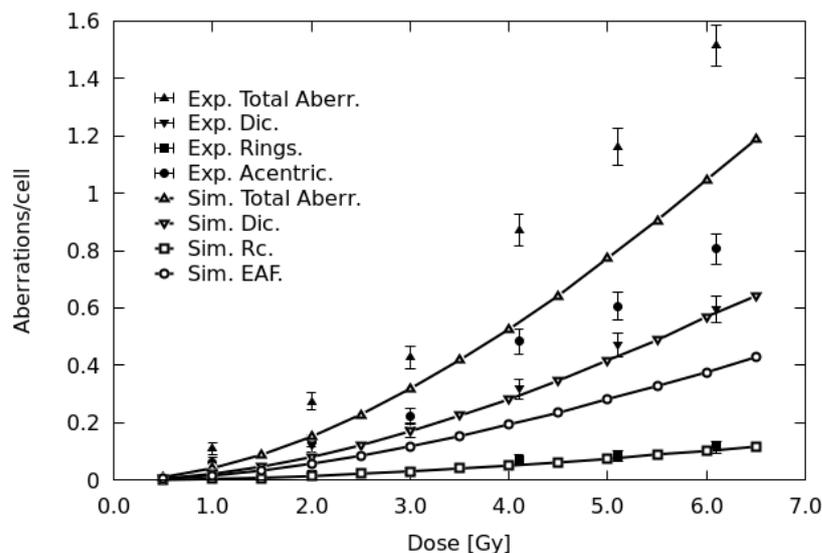


Figure 23 – Aberrations in AG01522 fibroblasts calculated by the Gaussian model. Like in Figure 22, full symbols represent experimental values taken from [20]; empty symbols, connected by lines to guide the eye (from top to bottom: total aberrations, dicentric, excess acentric fragments and centric rings), are results of simulations performed by the Gaussian model with $\sigma = 1.3 \mu\text{m}$ and $\text{CL} = 2.0 (\text{GyCell})^{-1}$. Source: Tello *et al.* DNArepair (2017) [19].

works (e.g. [182]). Concerning the G-ratio, both the exponential and the Gaussian model (as well as the previous step model) provided values that were substantially lower than the experimental ones, whose mean value was close to 7. Nevertheless, different than the step models, these values were significantly higher than that expected from randomness. Overall, the results reported in Figure 22 and Figure 23 and Table 5, suggest that also for fibroblasts an exponential model may reproduce proximity effects better than a Gaussian model or a step model, although the underestimation of acentric rings requires further investigations in the future.

The same experimental data set [20] was considered for comparison by Friedland and Kundrat [146, 239], who developed a very detailed chromosome-aberration model in the framework of the PARTRAC biophysical simulation tool, combining radiation track structure, multiscale models of DNA and chromosomes, kinetics of enzyme attachment and DSB repair, and chromatin mobility. The simulated DNA damage following energy deposition by radiation represents the starting point for modelling DSB repair by Non-Homologous End Joining (NHEJ), which in turn is the major component of the chromosome aberration simulations. After DSB recognition and chromatin remodeling, the DSB is assumed to form two mobile DNA ends, which are followed separately. Such DNA ends are allowed to diffuse in a semi-confined fashion (where confinement is due to anchoring of the chromatin fibre to nuclear attachment sites), in parallel with the association of

DNA-PK repair enzymes. If two ends with attached DNA-PK happen to be in close vicinity, they form a synaptic complex leading to final ligation of the two fragments. Further details on the methods, including assumptions and parameters, can be found in [239]. Calculated dicentric yields were compared with experimental data reported in [20], both for γ -rays and for α -particles. The shape of the simulated dose-response was in agreement with the experimental ones, i.e. linear-quadratic for γ -rays and linear for α -particles. However, the absolute dicentric yields were overestimated by a factor ~ 5 for γ -rays and a factor ~ 2 for α -particles; according to the authors, this suggests that the long-range mobility and the number of free chromatin ends at later repair times may have been overestimated. Interestingly, the DSB mis-rejoining probability resulting from these simulations is predicted to decrease exponentially with increasing initial distance between the two involved DNA ends, consistent with the conclusions of the present thesis work.

Table 5 – Experimental ("Exper.") F- and G-ratios for AG1522 human fibroblasts exposed to different doses of ^{137}Cs γ -rays compared with values calculated by the Gaussian model ("Gaus.") or the exponential model ("Exp."). The errors associated to the experimental values represent standard deviations calculated applying Poisson statistics to the raw numbers reported in the experimental work ([20], "Day 0" data set).

Dose(Gy)	F-ratio			G-ratio		
	Exper.	Exp.	Gauss	Exper.	Exp.	Gauss
1.0	6.0 ± 4.6	5.0	5.3	9.5 ± 7.1	2.0	1.6
2.0	6.8 ± 3.3	5.3	5.8	6.6 ± 3.2	2.1	1.7
3.0	5.9 ± 2.3	5.2	5.6	7.1 ± 2.7	2.1	1.6
4.1	4.4 ± 1.1	5.1	5.6	6.4 ± 1.6	2.1	1.7
5.1	5.6 ± 1.3	5.1	5.7	6.8 ± 1.5	2.1	1.7
6.1	5.2 ± 1.0	5.2	5.6	6.8 ± 1.3	2.2	1.7

4.3 Dependence of proximity effects on radiation quality, cell type and dose

Afterwards, the investigation was extended to intermediate- and high-LET. Since the σ values ($0.8 \mu\text{m}$ for lymphocytes and $0.7 \mu\text{m}$ for fibroblasts) were taken from the low-LET study [19] (see section 4.2), the results were obtained by adjusting only one model parameter, i.e. the yield of "Cluster Lesions" (CLs). In lymphocytes, the exponential model allowed reproducing both dose-response curves for different aberrations (dicentrics, centric rings and excess acentrics), and values of F-ratio (dicentrics to centric rings) and G-ratio (interstitial deletions to centric rings). In fibroblasts, a good correspondence was found with the dose-response curves, whereas the G-ratio (and, to a lesser

extent, the F-ratio) was underestimated. With increasing LET, F decreased and G increased in both cell types, supporting their role as "fingerprints" of high-LET exposure. A dose-dependence was also found at high LET, where F increased with dose and G decreased, possibly due to inter-track effects. We therefore concluded that, independent of radiation quality, in lymphocytes an exponential function can describe proximity effects at both inter- and intra-chromosomal level; on the contrary, in fibroblasts further studies (experimental and theoretical) are needed to explain the strong bias for intra-arm relative to inter-arm exchanges. The details of this investigation will be presented in [subsection 4.3.1](#) (lymphocytes) and [subsection 4.3.2](#) (fibroblasts).

4.3.1 Human lymphocytes

4.3.1.1 Proton aberration yields

Figures [24-26](#) show the results obtained for lymphocytes irradiated with protons of 3.5, 5.3, and 19.0 $keV/\mu m$, respectively. Each figure includes the calculated absolute yields of dicentrics (DIC), centric rings (CER) and excess acentric fragments (ACE) using the exponential model, as well as experimental data taken from the literature [[18](#)]. Since the σ parameter should not depend on radiation quality, the same value used in the previous study for photons (see [section 4.2](#)) was applied, i.e. $\sigma = 0.8 \mu m$. This choice is related to the fact that σ , which represents the characteristic distance governing the (mis-)rejoining probability, should be mainly related to interphase chromatin organization and mobility. As shown for low-LET radiation, the value of σ has a strong influence on the F-ratio, which increases with σ . Some further tests confirmed this tendency, showing that larger values of σ lead to an over-estimation of the F-ratio, whereas smaller σ values lead to an under-estimation of F. The CL yield was adjusted separately for each LET value basing on the dicentric experimental yields. For all simulations, the relative uncertainty was smaller than 5%; the simulation outcomes are therefore reported without error bars, which would be smaller than the symbol size.

For protons of 3.5 $keV/\mu m$ ([Figure 24](#)), the calculated yields of both dicentrics and centric rings, and thus the F-ratio, were in good agreement with the data, since in most cases the simulation outcome was within the experimental error bar. Concerning acentrics, the simulations were systematically lower than the data and appeared very similar to dicentrics. However, this was not the case neither for the photon results discussed before, nor for the 5.3 $keV/\mu m$ results (see [Figure 25](#) below), where the dicentric yields were higher than the acentric yields, not only in the simulations but also in the data. In general, both theoretical and experimental works suggest that at low LET the yield of acentrics tends to be lower than that of dicentrics, whereas at high LET the opposite scenario occurs; this issue will be discussed in later. The fact that the simulated G-ratio

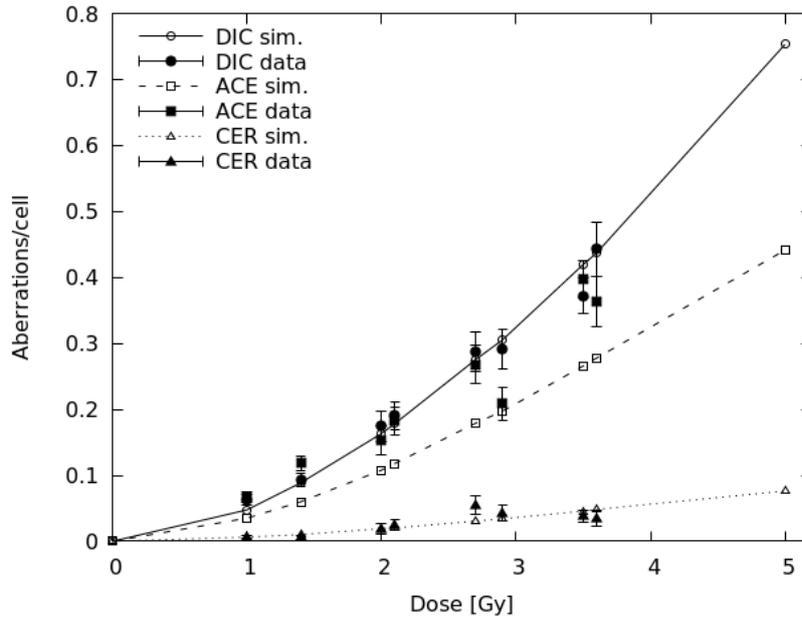


Figure 24 – Aberrations in human lymphocytes exposed to $3.5 \text{ keV}/\mu\text{m}$ protons. Empty symbols, connected by lines to guide the eye (solid line for dicentrics, dashed line for excess acentrics and dotted line for centric rings), represent simulation outcomes obtained with the exponential model, taking $CL = 0.02 \mu\text{m}^{-1}$. Full symbols are experimental values taken from Bauchinger and Schmid [18]; the error bars represent one standard deviation from the mean, calculated from the raw numbers assuming Poisson statistics. Source: Tello *et al.* DNArepair (2018) [21].

was consistent with the corresponding experimental value, suggests that the underestimation of total acentrics found at $3.5 \text{ keV}/\mu\text{m}$ is mainly related to an underestimation of (large) linear fragments, rather than interstitial deletions. Figure 25 shows the results obtained for protons of $5.3 \text{ keV}/\mu\text{m}$. At lower doses (up to 2 Gy), the absolute aberration yields show very good agreement with the experimental data for all three aberration categories. For the two highest doses (4.9 and 5.1 Gy), the dicentric yields were (slightly) overestimated, whereas the yields of acentric fragments were (slightly) underestimated. The results obtained at $19 \text{ keV}/\mu\text{m}$ are reported in Figure 26. Concerning dicentrics, if one excludes the experimental point at 3.6 Gy , which is much lower than that at 3.7 Gy , a good correspondence between simulations and data was found, although the simulations (slightly) underestimated the data at the lower doses and (slightly) overestimated the point at the highest dose (3.7 Gy). Analogous conclusions can be drawn for acentrics if one excludes the experimental point at 1.1 Gy , which is as high as that at 1.4 Gy , and the point at 3.6 Gy , which is much lower than that at 3.7 Gy .

4.3.1.2 Alpha-particle aberration yields

Simulations were then run for $150 \text{ keV}/\mu\text{m}$ alpha-particles, for which the paper by Bauchinger and Schmid [18] also contains experimental data. This data set is particularly

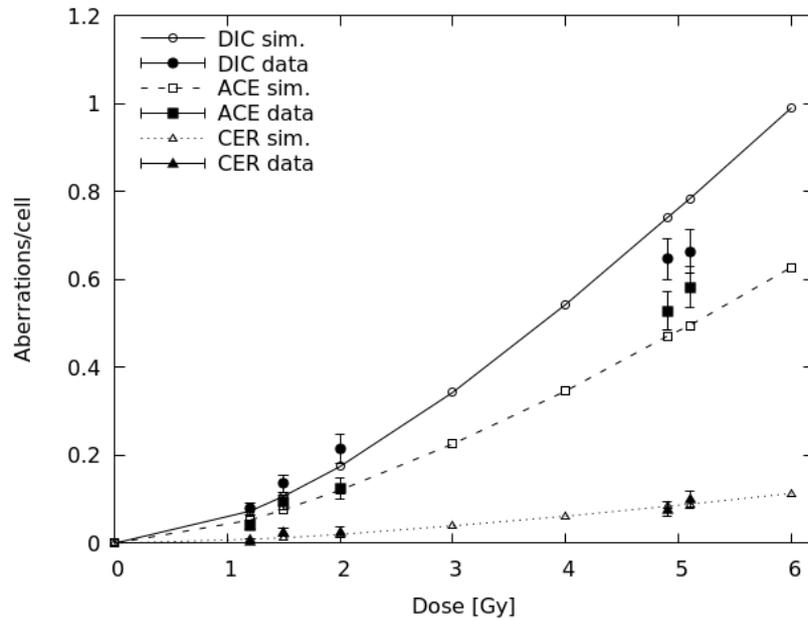


Figure 25 – Aberrations in human lymphocytes exposed to $5.3 \text{ keV}/\mu\text{m}$ protons. Empty symbols, connected by lines to guide the eye (solid line for dicentrics, dashed line for excess acentrics and dotted line for centric rings), represent simulation outcomes obtained with the exponential model, taking $CL = 0.03 \mu\text{m}^{-1}$. Full symbols are experimental values taken from Bauchinger and Schmid [18]; the error bars represent one standard deviation from the mean, calculated from the raw numbers assuming Poisson statistics. Source: Tello *et al.* DNArepair (2018) [21].

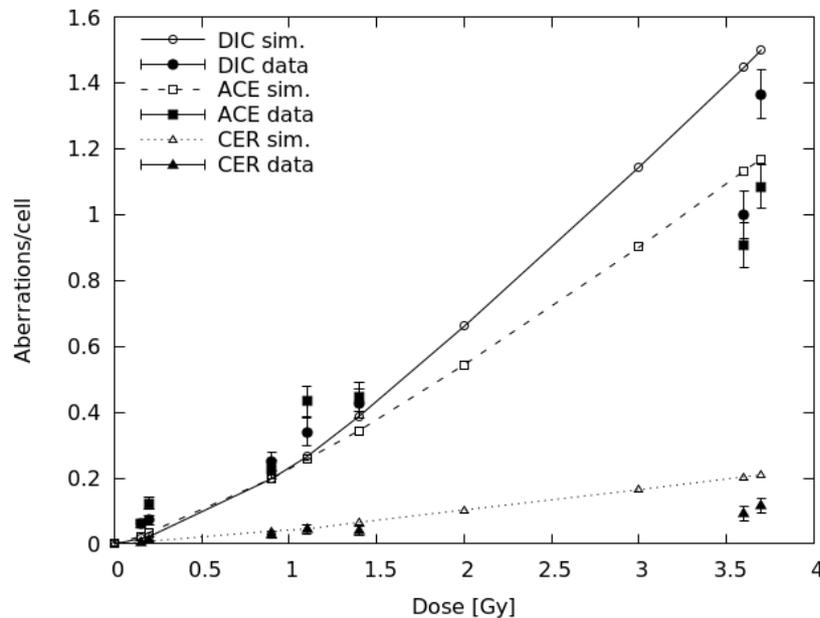


Figure 26 – Aberrations in human lymphocytes exposed to $19.0 \text{ keV}/\mu\text{m}$ protons. Empty symbols, connected by lines to guide the eye (solid line for dicentrics, dashed line for excess acentrics and dotted line for centric rings), represent simulation outcomes obtained with the exponential model, taking $CL = 0.21 \mu\text{m}^{-1}$. Full symbols are experimental values taken from Bauchinger and Schmid [18]; the error bars represent one standard deviation from the mean, calculated from the raw numbers assuming Poisson statistics. Source: Tello *et al.* DNArepair (2018) [21].

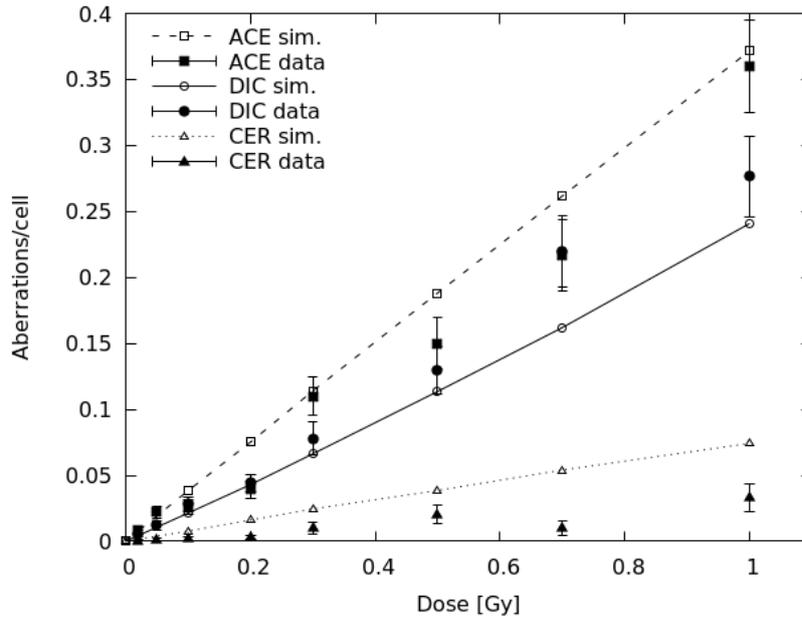


Figure 27 – Aberrations in human lymphocytes exposed to $150 \text{ keV}/\mu\text{m}$ α -particles. Empty symbols, connected by lines to guide the eye (solid line for dicentrics, dashed line for excess acentrics and dotted line for centric rings), represent simulation outcomes obtained with the exponential model, taking $CL = 1.06 \mu\text{m}^{-1}$. Full symbols are experimental values taken from Bauchinger and Schmid [18]; the error bars represent one standard deviation from the mean, calculated from the raw numbers assuming Poisson statistics. Source: Tello *et al.* DNArepair (2018) [21].

interesting because, due to the high LET and the small nucleus cross-sectional area, especially at relatively low doses it was very unlikely that a cell was hit by more than one particle: representing a lymphocyte nucleus as a sphere with a $3 \mu\text{m}$ radius, even at the highest dose (1 Gy) the mean number of particle traversals is 1.2. These data thus help clarifying what happens when a single α -particle traverses the cell nucleus, which cannot be easily derived from lymphocyte data at higher doses and/or fibroblast data, for which an inter-track action due to the presence of multiple traversals cannot be excluded. More specifically, comparing with these data allowed estimating the yield of critical DNA damages (where "critical" means severe enough to lead to chromosome aberrations) induced by a single α -particle, which at $150 \text{ keV}/\mu\text{m}$ resulted to be about $1 \text{ CL}/\mu\text{m}$ (see below). This is consistent with experimental results obtained by Costes *et al.* [240]: measuring the induction of γH2AX , ATMP and 53BP1 foci in human epithelial cells exposed to $1 \text{ GeV}/\text{amu}$ Fe ions, which also have a LET of about $150 \text{ keV}/\mu\text{m}$, these authors found a (maximum) number of foci that was $\sim 0.7 - 0.8 \text{ foci}/\mu\text{m}$. Concerning the effectiveness of a single particle traversal, further experimental data are desirable on different particle types and/or different energies, especially data obtained by micro-beam irradiation, which allows delivering an exact number of particles rather than a mean number.

The results obtained with $1.06\text{ CL}/\text{micron}$ are reported in [Figure 27](#), which also shows the corresponding experimental data. Good agreement between simulations and data was found for both dicentrics and acentrics, since the discrepancies were within the experimental error in most cases. Interestingly, while at low and intermediate LET the dicentric yields were higher than the acentric yields, at $150\text{ keV}/\mu\text{m}$ the yields of acentrics became higher than those of dicentrics, both in the simulations and in the data. A cell-type dependence may also play a role, because in fibroblasts the ratio between total acentrics and dicentrics seems to be higher than in lymphocytes (see below). The interpretation of this phenomenon in terms of mechanisms is not trivial, because one would need to distinguish between terminal deletions, which are produced by a single, un-rejoined chromosome break, and interstitial deletions, which are produced by two breaks induced in the same chromosome arm. If the increase in (total) excess acentrics results to be due to an increase in interstitial deletions, then it may be explained considering that high-LET radiation, which is densely-ionizing, is more likely to induce two breaks in the same chromosome arm with respect to low-LET radiation. In some works, including that by Bauchinger and Schmid [18], all acentrics were sub-divided into linear or circular fragments. Notwithstanding, it is not automatic that the fragments scored as linear are terminal deletions, because some fragments scored as (small) linear forms may actually be small acentric rings where the "hole" is not visible. Experimental studies making use of pan-telomeric probes, like that performed by Mestres et al. [219], can be of great help in this regard. Concerning centric rings, the simulations were systematically higher than the data. Of course this implies that the simulated F ratio was much lower than the experimental one. This issue will be discussed in the *Aberration ratios* section.

Analogous to the low-LET work ([section 4.2](#)), the simulations for α -particles were also repeated by applying the Gaussian model. Similarly to the exponential model, the yields of dicentrics and acentrics were in good agreement with the data, whereas those of centric rings overestimated the data; however, the G-ratio was lower than that found with the exponential model. The lower G-ratio may be a consequence of the fact that, at the small distances characterizing the intra-chromosomal level, the Gaussian function decreases less rapidly than the exponential one, thus leading to a less pronounced bias for intra-arm exchanges versus inter-arm exchanges.

4.3.1.3 Aberration ratios

[Table 6](#) summarizes the values of F-ratio and G-ratio (" F_{sim} ." and " G_{sim} .", respectively) obtained with the exponential model for lymphocytes exposed to the four considered radiation qualities as well as γ -rays, which have been discussed in [section 4.2](#). The corresponding experimental data (" F_{data} ." and " G_{data} .", respectively), taken from Bauchinger and Schmid [18], are also reported for comparison. As already discussed in

section 4.2, the simulated G-ratio for γ -rays was substantially higher than the experimental one, which was even smaller than the value expected assuming randomness. However, higher values can be found in other works: for instance, a mean G-ratio of 2.84 ± 0.61 is reported in Benkhaled et al. [216], and a G-ratio of about 2 is reported in Deng et al. [241].

Table 6 – F-ratio (total dicentrics/total centric rings in the whole dose range) and G-ratio (total interstitial deletions/total centric rings in the whole dose range) in human lymphocytes, compared with data taken from Bauchinger and Schmid [18].

Radiation	Dose range (Gy)	F_{sim}	F_{data}	G_{sim}	G_{data}
^{137}Cs γ – rays	1.0 – 4.0	8.7	9.1 ± 0.8	1.6	0.9 ± 0.1
p 3.5 keV/ μm	1.0 – 3.6	8.7	8.8 ± 1.0	1.8	1.5 ± 0.2
p 5.3 keV/ μm	1.2 – 5.1	8.7	7.5 ± 1.0	1.8	1.5 ± 0.2
p 19.0 keV/ μm	0.15 – 3.7	6.7	10.2 ± 1.2	2.4	2.0 ± 0.3
α 150 keV/ μm	0.1 – 1.0	3.0	10.1 ± 1.9	3.4	3.1 ± 0.6

The only large discrepancy was related to the F-ratio for 150 keV/ μm α -particles; however, our results are not so inconsistent with other data reported in the literature. Using pan-centromeric and pan-telomeric probes in lymphocytes exposed to 150 keV/ μm α -particles (dose range: 0.2–1.0 Gy), Mestres et al. [219] found a mean F-ratio of 5.47 ± 0.36 . Even lower values have been observed following in vivo α -particle exposure, for which Sasaki et al. [242] reported 5.0 ± 0.3 , and Tawn et al. [243] reported 4.5 ± 2.0 . Interestingly, Mestres et al. [219] noted that, when tricentrics and tetracentrics were considered as dicentric equivalents and included in the F-ratio calculation, their ratio increased to 7.16, suggesting that the inclusion of higher-order multacentrics may be significant. Also our simulations showed an increase of F if higher-order multacentrics were included (counting a tricentric as 2 dicentrics, a quadricentric as 3 dicentrics, etc.), although the increase was less pronounced than that found by Mestres et al. [219]: for instance, at 1 Gy the simulated F-ratio increased from 3.2 to 3.5. A similar trend was also found at 19 keV/ μm , where, for instance, at 3.7 Gy the simulated F-ratio increased from 7.2 to 7.7 after including higher-order multacentrics. Concerning the dependence of F and G on radiation quality, the simulations confirmed the increase of G with LET shown in various experimental works, as well as the decrease of F reported in some other works (e.g. [219, 242, 243]). Furthermore, the simulations at the higher LET values (19 keV/ μm and 150 keV/ μm) showed an increase of F with increasing dose. This is possibly related to the fact that, with increasing dose, there is an increased probability for a dicentric to be induced by two different particles (inter-track effect).

More generally, the results obtained for lymphocytes exposed to protons and

α -particles suggest that, in human lymphocytes, the dependence of the fragment (mis-)rejoining probability on the fragment initial distance can be described by an exponential function not only at low LET, but also at intermediate and high LET. On the contrary, the Gaussian model provided a good description of the data at high LET, but did not perform well at low LET. We therefore conclude that, at least for lymphocytes, the exponential model is more realistic than the Gaussian one, independent of radiation quality. The fact that a good agreement with the data was obtained without changing the value of σ is consistent with the biophysical meaning of this parameter, which should be related to the chromatin mobility features independent of radiation quality. Many authors (e.g. [238]) found that chromatin free-end mobility is mainly governed by diffusion, whereas others hypothesize an active transport mechanism. Interestingly Friedland and Kundrát [146, 239], in an independent simulation work assuming free-end diffusion, found a posteriori that the mis-rejoining probability decreased exponentially with increasing the initial distance between the two involved DNA ends. This implies that an exponential decrease is consistent with a diffusion mechanism, at least for chromosome aberration production by Non-Homologous End Joining. Of course this does not exclude that in other repair pathways an active transport can play a role. To investigate this issue, it is desirable that further experiments are performed where the time-dependent motion of DNA free-ends is followed, e.g. by time-lapse observation of repair proteins foci. Performing these observations in different cell types which adopt different repair pathways (and/or at different cell-cycle stages where different pathways are activated), should help identifying possible scenarios where an active transport may play a role.

4.3.2 Human fibroblasts

The results described in subsection 4.2.2 suggest that, at least for low-LET radiation, an exponential model of the form $\exp(-r/\sigma)$ can reproduce the proximity bias for inter-arm exchanges with respect to inter-chromosome exchanges not only in lymphocytes but also in fibroblasts. However, different from lymphocytes, in fibroblasts such exponential model underestimates the bias for intra-arm versus inter-arm exchanges. A Gaussian model provided a worse performance, since it allowed obtaining good agreement for dicentrics and rings, and thus for the F-ratio, but substantially underestimated total acentrics. In the following, the results obtained for fibroblasts at high LET will be presented and discussed.

4.3.2.1 Alpha-particle aberration yields

To extend the investigation to high-LET radiation, we took the data for comparison from the same paper considered for photons [20], which also reports data on AG1522 cells exposed to $116 \text{ keV}/\mu\text{m}$ α -particles. Figure 28 displays the outcomes obtained by

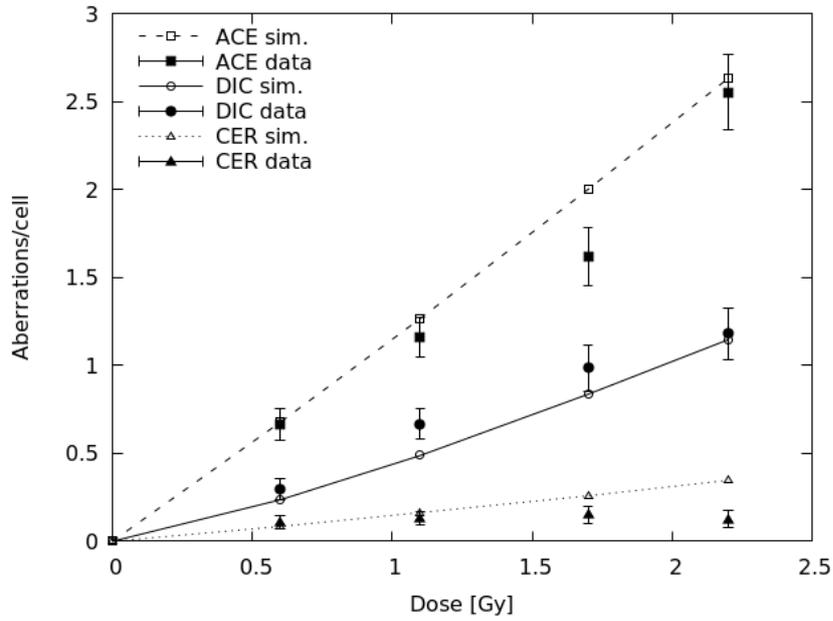


Figure 28 – Aberrations in human fibroblasts exposed to $116 \text{ keV}/\mu\text{m}$ α -particles. Empty symbols, connected by lines to guide the eye (solid line for dicentrics, dashed line for excess acentrics and dotted line for centric rings), represent simulation outcomes obtained with the exponential model, taking $CL = 0.47 \mu\text{m}^{-1}$. Full symbols are experimental values taken from Cornforth et al. [20]; the error bars represent one standard deviation from the mean, calculated from the raw numbers assuming Poisson statistics. Source: Tello *et al.* DNArepair (2018) [21].

applying the exponential model with $\sigma = 0.7 \mu\text{m}$ (the same value used for photons) and $0.47 \text{ CL}/\mu\text{m}$; the data reported in Cornforth et al. [20] is also shown. Analogous to what found for photons, the simulations compare reasonably well with the data both for dicentrics and for excess acentrics. On the contrary the experimental rings were overestimated, implying that the simulated F-ratio was smaller than the experimental one (see next section).

The results obtained with the exponential model for AG1522 cells exposed to high-LET radiation thus confirmed the main features of those obtained at low LET, since the simulations were in agreement with the data for dicentrics and total acentrics, whereas interstitial deletions were underestimated. Furthermore, at high LET the experimental yield of rings was overestimated. This implied that, at high LET, the G-ratio and the F-ratio were both underestimated. Reducing the value of σ allowed to obtain higher values for G but even lower values for F: for instance, using $\sigma = 0.5 \mu\text{m}$, at 2.2 Gy the G-ratio increased from 5.2 to 6.4, but the F-ratio decreased from 3.3 to 2.6. Since for α -particle irradiated lymphocytes the performance of the Gaussian model was not very different from that of the exponential model, also for α -irradiated fibroblasts simulations were run applying the Gaussian model. Like at low LET, the exponential model and the Gaussian one behaved rather similarly for dicentrics and centric rings (and thus for the

F-ratio), but the exponential model was closer to the data both for total acentrics and for the G-ratio.

4.3.2.2 Aberration ratios

Table 7 reports the values of F-ratio and G-ratio (" F_{sim} " and " G_{sim} ", respectively) obtained with the exponential model for fibroblasts exposed to $116 \text{ keV}/\mu\text{m}$ α -particles, as well as γ -rays for comparison. The corresponding experimental data (" F_{data} " and " G_{data} ", respectively), taken from Cornforth et al. [20], is also reported for comparison. As mentioned above, a major discrepancy is related to an underestimation of G for both radiation types, and possibly an underestimation of F for α -particles. It is interesting to note that also in the simulations G increased with LET, consistent with the data. Concerning the dose-dependence, in the simulations F increased with dose, like for α -particle-irradiated lymphocytes; furthermore, G significantly decreased with increasing dose. Possible explanations for such a decrease of G , which was found for fibroblasts but not for lymphocytes, might include a more pronounced role of inter-track effects, related to the higher number of particles involved: while in lymphocytes the (mean) fluence at the highest dose was $0.04 \text{ particles}/\mu\text{m}^2$, in fibroblasts it was $0.12 \text{ particles}/\mu\text{m}^2$.

Table 7 – F-ratio (total dicentrics/total centric rings in the whole dose range) and G-ratio (total interstitial deletions/total centric rings in the whole dose range) in AG1522 human fibroblasts, compared with data taken from Cornforth et al. [20].

Radiation	Dose range (Gy)	F_{sim}	F_{data}	G_{sim}	G_{data}
$^{137}\text{Cs } \gamma - \text{rays}$	1.0 – 6.1	5.1	5.3 ± 0.6	2.2	6.8 ± 0.8
$\alpha \text{ } 116 \text{ keV}/\mu\text{m}$	0.6 – 2.2	3.2	5.7 ± 1.0	5.5	10.3 ± 1.8

The substantial underestimation of the G-ratio suggests that even an exponential function of the form $\exp(-r/\sigma)$ is not sufficient for reproducing the large bias for intra-arm exchanges to inter-arm exchanges observed in fibroblasts, which deserves further investigation in the future. The fact that in fibroblasts such bias seems to be more pronounced than in lymphocytes might be related to many biological factors, including not only a reduced mobility of chromatin [238] but also its 3D organization in the cell nucleus during interphase, as well as the dimensions of the various chromosome territories. Since the volume of fibroblast nuclei is much larger than that of human lymphocyte nuclei, the (mean) volume of a chromosome territory is much larger in fibroblasts than in lymphocytes. Assuming that, after a given time interval, a chromosome free-end induced in a fibroblast has travelled a similar distance as a free-end induced in a lymphocyte, then for fibroblasts a free-end initially induced in a given chromosome arm may be more likely to stay within that arm domain, and possibly join with another free-end induced in the same arm, thus giving rise to an interstitial deletion (or a paracentric inversion). On the

contrary, for lymphocytes a free-end initially induced in a given arm may have a higher probability of reaching the domain occupied by the other arm and joining with a free-end induced there, thus producing an inter-arm exchange like a centric ring (or a pericentric inversion). This may explain why the values of G-ratio experimentally observed in fibroblasts tend to be higher than in lymphocytes, and why an exponential function allows to reproduce the G-ratio in human lymphocytes but underestimates it in human fibroblasts.

More generally, these findings support a scenario where the intrinsic mobility properties of chromatin free-ends have a strong influence on the distance-dependence of the (mis-)rejoining probability. In turn, such properties might be related to interphase chromatin organization, including the existence of attachment points to the nuclear matrix: one possibility is that, after a (complex) DSB has been induced, the maximum distance that a free-end can travel is given by the length of the chromatin loop between two subsequent attachment points. For instance, considering a loop having a size of $50 - 200 \text{ kbp}$ (e.g. [244]), and assuming that each 1.2 kbp structure corresponds to 10 nm , one would obtain a distance of $\sim 0.4 - 1.6 \mu\text{m}$, which is consistent with the characteristic distance for the mis-rejoining probability found in this and other works.

4.4 Applications for hadrontherapy

This section introduces the results of two collaborations with regard to applications useful for hadrontherapy, in which the BIANCA model served as a means to interpret experimental data. The radiation qualities and doses used in the experiments span a wide range of setups that are of interest in hadrontherapy. In both works, the simulations were carried out using the step-like proximity function. In the collaboration with the BioQuaRT project (subsection 4.4.1), the experiments were performed with the microbeam available at the Physikalisch-Technische Bundesanstalt (PTB) in Braunschweig, Germany. The upgraded version of BIANCA allowed to precisely simulate this irradiation setup, according to the microbeam specifications (e.g. beam positions, FWHM and number of particles). In a second collaboration (subsection 4.4.2), experimental data on V79, AG01522 human fibroblasts and U87 cells exposed to protons, C-ions and He-ions, was analyzed in terms of CL yields per micron. Within this work, an approach was proposed to predict the ion-response of cell line(s) of interest from the ion-response of a reference cell line and the photon responses of both cell lines.

4.4.1 Analysis of radiation-induced chromosome aberrations after alpha particle microbeam irradiation

As a part of the EMRP joint research project, Biologically Weighted Quantities in Radiotherapy (BioQuaRT), experimental and theoretical analyses on chromosomal aberrations were performed in Chinese hamster ovary cells (CHO-K1) exposed to α particles with final energies of 5.5 and 17.8 MeV (absorbed doses: ~ 2.3 Gy and ~ 1.9 Gy, respectively), which were generated by the microbeam at the Physikalisch-Technische Bundesanstalt (PTB) in Braunschweig, Germany. In line with the differences in linear energy transfer (approximately 85 keV/ μ m for 5.5 MeV and 36 keV/ μ m for 17.8 MeV α particles), the 5.5 MeV α -particles were more effective than the 17.8 MeV α -particles, both in terms of the percentage of aberrant cells (57% vs. 33%) and aberration frequency. The yield of total aberrations increased by a factor of ~ 2 , although the increase in dicentric plus centric rings was less pronounced than in acentric fragments. The experimental data was compared with Monte Carlo simulations based on the BIANCA model. This comparison allowed interpretation of the results in terms of critical DNA damage [cluster lesions (CLs)]. More specifically, the higher aberration yields observed for the 5.5 MeV α particles were explained by taking into account that, although the nucleus was traversed by fewer particles (nominally, 11 vs. 25), each particle was much more effective (by a factor of ~ 3) at inducing CLs. This led to an increased yield of CLs per cell (by a factor of ~ 1.4), consistent with the increased yield of total aberrations observed in the experiments. Details of the experimental setup can be found in [235], whereas the simulation work will be reported below.

4.4.1.1 Absorbed dose estimation

To estimate the absorbed dose to the cell nucleus, an *ad hoc* simulation code was developed in which the cell nucleus was represented as a cylinder of circular base, with a variable base area and a fixed height. The mean base area was assumed to be 68 μ m², based on the average value found in measurements performed at PTB. This value is very close to the 70 μ m² value reported by Nasonova *et al.* [245] for the same cell line (CHO-K1), as well as the 69 μ m² value reported by Carpenter *et al.* [246] for a similar cell line (CHO-10B). For each cell, the actual nucleus area was taken from a Gaussian distribution that had a mean value of 68 μ m² and was truncated at $\pm 1 \sigma$ (30 μ m²). The nucleus thickness was assumed to be 4 μ m, as reported, for example, by Carpenter *et al.* [246] based on light microscopy measurements.

Like in the experiments, each nucleus was irradiated in a perpendicular direction from the bottom (foil side); the number of particles depended on the particle energy (11 or 25 particles for 5.5 or 17.8 MeV, respectively). Whereas for 5.5 MeV, all particles were

nominally targeted at the nucleus center, for 17.8 *MeV*, five (nominal) positions were targeted in each cell nucleus, corresponding to the following coordinates in a system centered in the nucleus center: $(0, 0)$, $(2\mu\text{m}, 2\mu\text{m})$, $(-2\mu\text{m}, 2\mu\text{m})$, $(-2\mu\text{m}, -2\mu\text{m})$, $(2\mu\text{m}, -2\mu\text{m})$. To account for the intrinsic targeting uncertainties, the actual position where a given particle entered the nucleus was "shifted" radially with respect to the nominal position by random sampling from a rotationally symmetric, bi-dimensional Gaussian distribution with FWHM of 4.5 μm . In this way, on average, approximately 2% of the particles were found to fall outside the cell nucleus after 5.5 *MeV* irradiations, where all particles were targeted at the nucleus center, while this percentage was approximately 10% after 17.8 *MeV* irradiations, where the particles were targeted according to the cross pattern described above.

For each cell nucleus, the energy deposited by each particle's traversal was calculated by multiplying the particle's LET at the cell center by the nucleus' traversal length; the specific energy imparted to that nucleus was calculated by dividing the total energy deposited by all particle traversals by the mass of that nucleus (assuming 1 g/cm^3 as a density). Finally, the absorbed dose was calculated by adding all specific energies and dividing them by the number of cells, which was 10,000. For 5.5 *MeV* irradiations, these calculations provided an absorbed dose of approximately 2.3 *Gy*, whereas for 17.8 *MeV* particles, the calculated value was approximately 1.9 *Gy*. In the latter case, the absorbed dose was lower because a higher percentage of particles did not hit the nucleus. Specifically, the calculations for 85 $\text{keV}/\mu\text{m}$ α -particles, by which 11 nucleus traversals provided an absorbed dose of approximately 2.3 *Gy*, are consistent with those reported by Nagasawa and Little [247], who exposed the same cell line (CHO-K1) to a 112 $\text{keV}/\mu\text{m}$ α -particle broad beam. Based on CR-39 fluence measurements and a mean nuclear area of 62.2 μm^2 , these authors calculated an average of 11 traversals per nucleus and a corresponding absorbed dose of 2 *Gy*.

4.4.1.2 Calculation of aberration yields.

For the two energies considered, Table 8 shows calculated yields of dicentrics plus centric rings, as well as EAF. The corresponding experimental yields, with uncertainties obtained as the square root of the empirical sample variance, are reported in parentheses. Columns 2 and 3 list the mean number of cluster lesions per particle track and per cell, respectively, that were used as code inputs to perform the calculations. All calculations were performed assuming $f = 0.08$ for the chromosome-fragment non-rejoining probability, based on the experience with V79 cells [16].

With the exception of the EAF category at 36 $\text{keV}/\mu\text{m}$, the CL values reported in the table enabled us to obtain simulation results that were within the experimental

uncertainties. While the different yields of CL per track reflect the different LET values, the yield of CLs per cell depends both on LET and on the number of nucleus traversals. Moving from the first to the second row of Table 8, the (nominal) number of particles per cell, which for a given absorbed dose is inversely proportional to the radiation LET, decreases by a factor of ~ 2.3 ; at the same time the yield of CLs per track increases by a factor of ~ 3.1 . This explains why the yield of CLs per cell increases by a factor of ~ 1.4 (which can be obtained by dividing 3.1 by 2.3). Such an increase in the yield of CLs per cell may explain the increase in aberrations observed in the experiments, which resulted in $\sim 2.3 \pm 0.1$.

Table 8 – Calculated aberration yields. The corresponding experimental values for comparison are reported in parenthesis.

LET($keV/\mu m$)	CLs/track	CLs/cell	(D + R)/cell	EAF/cell
36	0.17	4.2	0.349(0.36 \pm 0.04)	0.483(0.21 \pm 0.03)
85	0.53	5.8	0.605(0.60 \pm 0.06)	0.659(0.69 \pm 0.07)

CLs = cluster lesions; D + R = dicentrics plus centric rings; EAF = excess acentric fragments.

This interpretation is in line with DNA damage investigations performed within the same research project by other researchers, who investigated the induction of 53BP1 foci in primary HUVEC cell cultures under the same irradiation conditions, i.e., by α -particle microbeam performed at the PTB facility [248]. The results showed that, when the LET increased from approximately $37 keV/\mu m$ to approximately $90 keV/\mu m$ (nominal values, corresponding to our $36 keV/\mu m$ and $85 keV/\mu m$ values, respectively), the ratio between the mean number of observed foci and the number of particle tracks increased by a factor of ~ 1.5 . The fact that the increase in the number of foci was less pronounced than the increase in the number of cluster lesions (which increased by a factor of ~ 3) may be explained by considering that there is not a one-to-one correspondence between foci and (complex) DSBs, and that foci within the same particle track may not be distinguishable. This explanation is supported by simulations performed independently by Villagrasa *et al.* [248] using the Geant4-DNA code, according to which the mean number of DSBs per particle track increased by a factor of ~ 4 between the two LET values, and the probability for a single particle to induce at least one focus increased by a factor of up to ~ 2.4 .

The simulations performed with BIANCA also enabled investigation of the possible consequences of adopting two different irradiation geometries (all particles to the nucleus center for $5.5 MeV$ particles, cross pattern for $17.8 MeV$ particles). For both energies, delivering all particles to the nucleus center provided higher aberration yields with respect to adopting the cross pattern, because in the latter case there is a higher probability of ions not hitting the nucleus, owing to the $4.5 \mu m$ FWHM of the microbeam width. However, the differences were significant but not dramatic. If the $5.5 MeV$ particles were delivered according to the cross pattern (delivering 3 particles to the nucleus center and 2 particles to each of the four vertices),

the (total) aberration yield per cell was 1.03 times smaller than the yield obtained by delivering all particles to the nucleus center; if the 17.8 MeV particles were all delivered to the nucleus center, the aberration yield per cell was 1.16 times higher than the yield obtained with the cross pattern. For both energies, the difference between the two configurations was more pronounced for dicentrics plus centric rings than for EAF; this may be due to an increased probability of inter-track interactions when all particles were targeted to the nucleus center.

4.4.2 Application of BIANCA to cell survival data by protons, C-ions and He-ions at energies and doses used in hadrontherapy

The model version developed in this work was applied to V79 and AG01522 cells exposed to protons, C-ions and He-ions over a wide LET interval ($0.6 - 502 \text{ keV}/\mu\text{m}$), as well as proton-irradiated U87 cells. In this application, proximity effects were modelled with the step-like function. This time, only the cluster yields were adjusted since the f value was taken from previous works. As will be shown in the following, good agreement between simulations and experimental data was obtained, suggesting that BIANCA II is suitable for calculating the biological effectiveness of hadrontherapy beams. For both V79 and AG01522 cells, the mean number of CLs per micrometer was found to increase with LET in a linear-quadratic fashion before the over-killing region, where a less rapid increase, with a tendency to saturation, was observed. Although the over-killing region deserves further investigation, the possibility of fitting the CL yields is an important feature for hadrontherapy, because it allows performing predictions also at LET values where experimental data are not available. Finally, an approach was proposed to predict the ion-response of the cell line(s) of interest from the ion-response of a reference cell line and the photon response of both cell lines. A pilot study on proton-irradiated AG01522 and U87 cells, taking V79 cells as a reference, showed encouraging results. These results will be presented and discussed in sections 4.4.2.1 - 4.4.2.4.

4.4.2.1 Protons

A comparison of BIANCA II simulations with survival data on V79 cells exposed to different proton beams, taken from Folkard *et al.* [249] and Belli *et al.* [22], has been published in Carante and Ballarini [16]. In the present work, those simulations were repeated applying the BIANCA II upgraded version, where chromosome-arm domains are modelled explicitly and cell survival is calculated on a cell-by-cell basis. Figure 29(a) reports simulated survival curves for four LET values (7.7, 11.0, 20.0 and $30.5 \text{ keV}/\mu\text{m}$), as well as X-rays as a reference radiation. The experimental data for comparison was taken from [22]. Simulations at 10.1, 17.8 and $27.6 \text{ keV}/\mu\text{m}$ were also performed, and the outcomes showed good agreement with the data reported in [249]. More specifically, at 10.1 and $17.8 \text{ keV}/\mu\text{m}$ the value of the reduced chi-square was 0.6 and 1.9, respectively. At $27.6 \text{ keV}/\mu\text{m}$ the agreement for the entire curve, which extends to doses that are much higher than those used in hadrontherapy fractions, was worse (with a reduced chi-square of 11.8); however, limiting the comparison to doses not higher than 4 Gy the reduced chi-square was 1.7. These results were not included in Figure 29(a) to avoid making the figure

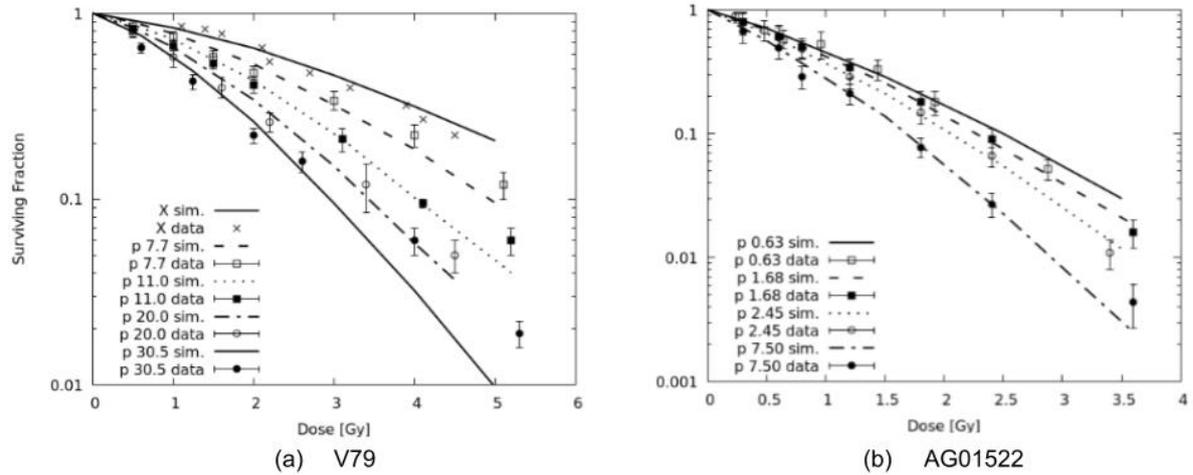


Figure 29 – (a) Survival of V79 cells irradiated with protons of different LET (7.7, 11.0, 20.0 and 30.5 $keV/\mu m$), as well as X-rays. The lines are simulation outcomes, the points are experimental data taken from Belli *et al.* (1998) [22]. (b) Survival of AG01522 cells irradiated with protons of different LET (from top to bottom: 0.63, 1.68, 2.45 and 7.50 $keV/\mu m$). The lines are simulation outcomes, the points are experimental data taken from Marshall *et al.* (2016) [23]. Source: Carante *et al.* PMB. (2018) [24].

too "crowded". All the curves for V79 cells were obtained with $f = 0.08$, based on a previous work [16]. The yield of cluster lesions was adjusted separately for each LET value. For the x-ray curve a yield of 4.3 $CL/(Gy \cdot Cell)$ was used, whereas the proton curves were obtained with CL yields in the range $\sim 0.004 - 0.025 CL/\mu m$, increasing with LET. Except for the highest LET value, where the slope of the simulated curve at high doses is more pronounced with respect to the data, the agreement was satisfactory, since in most cases the simulations were within the experimental error, and the reduced chi-square was between 1.1 and 1.9.

The underestimation of the surviving fraction at the higher doses (especially at the higher LET values) is suggestive of an over-estimation of the so-called "inter-track effects". This might be related to the choice adopted for the fragment end-joining probability, which was modelled by a step function with threshold distance in the order of the linear dimensions of interphase chromosome territories (that is, $\sim 3.8 \mu m$ for V79 cells). At the higher doses, where a cell nucleus is traversed by many primary particles, such choice may lead to an over-estimation of the interaction between two CL free-ends induced by two different primary particles. Adopting an end-joining probability that decreases monotonically with increasing the (initial) fragment distance, as we did in [19, 21] (section 4.2 and 4.3), might help overcoming this problem in future cell survival studies.

In Carante and Ballarini (2016) [16], the model has been also applied to AG01522 normal human fibroblasts exposed to a proton beam available at the CATANA ocular melanoma facility of INFN-LNS in Catania, Italy [250], where the cells were exposed at different depth positions corresponding to LET values in the range 1.1 – 22.6 $keV/\mu m$ [27]. Like for V79 cells, in the present work those simulations were repeated making use of the upgraded BIANCA II, and the

results were in line with those reported in [16].

Furthermore, a subsequent work by the same group of authors was considered, in which AG01522 cells were irradiated at four key depth positions of a clinical SOBP available at the Proton Therapy Center in Prague, Czech Republic [23]. The corresponding LET values were 0.63, 1.68, 2.45 and 7.50 $keV/\mu m$. Figure 29(b) reports simulated survival curves for these four LET values. The corresponding experimental data, taken from Marshall *et al.* [23], are also reported for comparison. Based on [16], all simulations for AG01522 cells were performed with $f = 0.18$. Again, the yield of cluster lesions was adjusted separately for each LET value. The curves reported in Figure 29(b) were obtained using CL yields in the range $\sim 0.001 - 0.011 CL/\mu m$, increasing with LET. With these values, the agreement between simulation outcomes and experimental data was quite satisfactory, since all simulations were within the error bars reported in the experimental work, and the value of the reduced chi-square was smaller than 1 for all curves.

The CL yields (mean number of CLs per unit length along the primary-particle traversal) used to simulate the survival of proton-irradiated V79 and AG01522 cells are reported in Figure 30 as a function of LET. For both cell lines the CL yield increased with LET, consistent with the increasing clustering of energy deposition in the target. For a given LET value, a higher CL yield was used for AG01522 cells with respect to V79 cells, reflecting the higher radiosensitivity of AG01522 cells. For both cell lines the LET-dependence of the CL yield was well described by a linear-quadratic function of the form $Y(L) = aL + bL^2$, where Y is the CL yield expressed in $CL/\mu m$, L is the radiation LET in $keV/\mu m$, and a and b are fitting parameters. To evaluate the fit goodness, we assigned to each CL yield a 5% relative error, equal to the maximum relative error of the simulated surviving fraction. Such CL fitting allows deriving CL yields, and thus predicting survival curves, even at LET values where experimental data are not available. This is important in view of applications in hadrontherapy, where the biological effectiveness along the beam profile should be known at as many depth positions as possible, that is as many LET values as possible.

Interestingly, the two fitting curves showed a very similar shape: a posteriori, we realized that by multiplying the CL fitting function of V79 cells by a fixed factor, one obtains a function that fits reasonably well the CL yields of AG01522 cells. The possible implications will be discussed in subsection 4.4.2.4.

4.4.2.2 Carbon ions

The approach was then extended to Carbon ions. For V79 cells the data reported in Furusawa *et al.* were considered, which cover a very wide LET range (22.5 – 502 $keV/\mu m$) [25]. Keeping $f = 0.08$ and adjusting the CL yield for each LET value, simulated survival curves were obtained for all 24 LET values reported in the experimental paper. To avoid making the figures too difficult to read, only some of them were reported in Figure 31. Figure 31 also shows the x-ray curve as a reference (in panel (a)). CL yields in the range $\sim 0.009 - 0.556 CL/\mu m$ provided

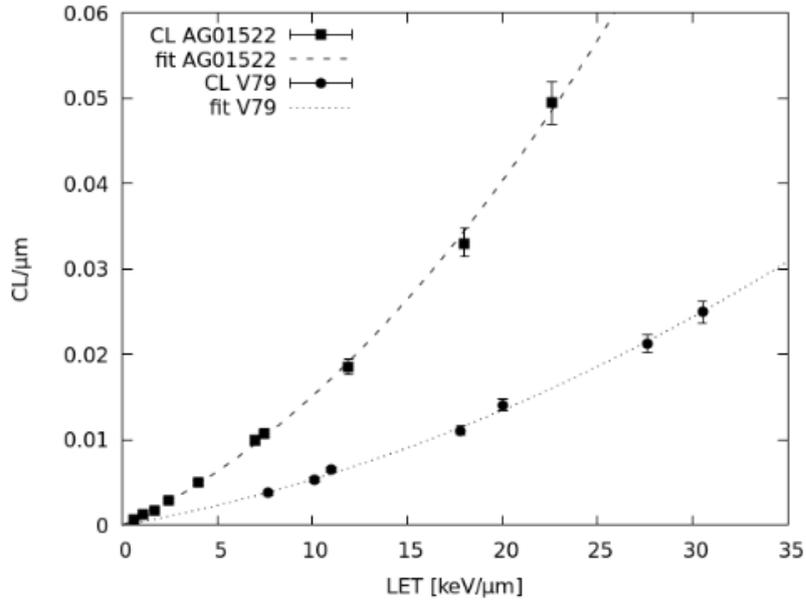


Figure 30 – CL yields used to simulate the survival of proton-irradiated V79 cells (lower line and points) and AG01522 cells (upper line and points). Each point represents the mean number of CLs per micrometer used as a code input, whereas the lines are linear-quadratic fits. A 5% relative error was assigned to each CL yield. Source: Carante *et al.* PMB. (2018) [24].

simulation outcomes in good agreement with the experimental results obtained by [25], who fitted each survival curve by a linear-quadratic function of the form $S(D) = \exp(-\alpha D - \beta D^2)$. The paper reports the α coefficient and the D_{10} coefficient (dose at 10% survival), from which we derived the β coefficient. Concerning the 16 curves that are not shown in Figure 31, for 11 of them the maximum percentage displacement between simulated curve and data fit was $\sim 30\%$, for four of them it was $\sim 50 - 60\%$, and only for the remaining one the agreement was worse. The LET-dependence of the CL yield will be discussed below.

For Carbon-irradiated normal human fibroblasts we considered the data reported by Hamada *et al.* [13], who irradiated AG01522 cells with Carbon beams of $76.3 \text{ keV}/\mu\text{m}$ and $108.0 \text{ keV}/\mu\text{m}$, and Kavanagh *et al.* [26], who irradiated AG01522 cells with Carbon beams of $48.8 \text{ keV}/\mu\text{m}$ and $147.6 \text{ keV}/\mu\text{m}$. Figure 32 reports simulated survival curves for Carbon beams with these four LET values. Like for proton-irradiated AG01522 cells, all the curves reported in Figure 32 were obtained with $f = 0.18$ based on [16]. Following adjustment of the CL yield for each LET value, CL yields in the range $0.052 - 0.354 \text{ CL}/\mu\text{m}$ provided simulation outcomes in good agreement with the considered data, since the simulations were within the experimental error bars with the only exception of the point at the highest dose for $147.6 \text{ keV}/\mu\text{m}$. However, this point is also much lower than the fitting function reported in the experimental paper. At $48.8, 108.0$ and $147.6 \text{ keV}/\mu\text{m}$ (excluding the point at the highest dose for $147.6 \text{ keV}/\mu\text{m}$) the reduced chi-square was between 1.1 and 1.7. At $76.3 \text{ keV}/\mu\text{m}$ the reduced chi-square was much higher (8.2), but it should be taken into account that only three experimental points were available, and the first one had a very small error bar.

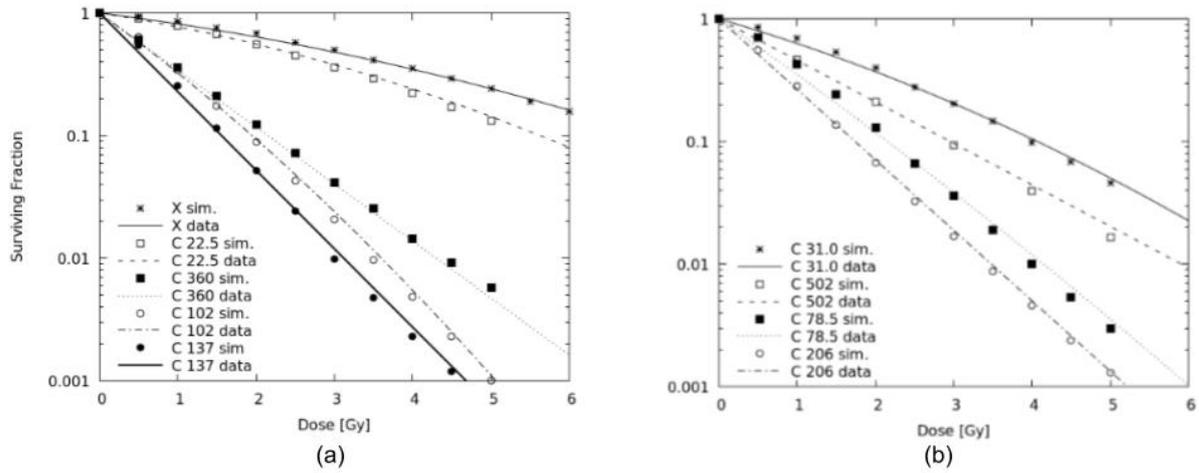


Figure 31 – Survival of V79 cells irradiated by Carbon beams of different (dose-averaged) LET, as well as X-rays as a reference. Panel (a): 22.5, 360.0, 102.0 and 137.0 $keV/\mu m$. Panel (b): 31.0, 502.0, 78.5 and 206.0 $keV/\mu m$. The points are simulation outcomes, whereas the lines are experimental data fits of the form $S(D) = \exp(-\alpha D - \beta D^2)$ taken from Furusawa *et al.* (2000) [25]. Source: Carante *et al.* PMB. (2018) [24].

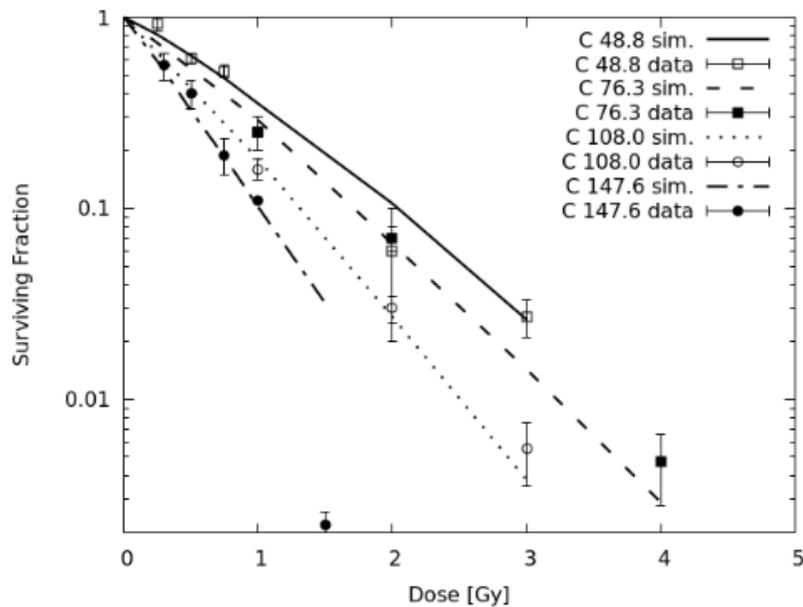


Figure 32 – Survival of AG01522 cells irradiated by Carbon beams of different LET (from top to bottom: 48.8, 76.3, 108.0 and 147.6 $keV/\mu m$). The lines are simulation outcomes, whereas the points are experimental data taken from Hamada *et al.* (2006) (76.3 and 108.8 $keV/\mu m$) [13] and Kavanagh *et al.* (48.8 and 147.6 $keV/\mu m$) (2013) [26]. Source: Carante *et al.* PMB. (2018) [24].

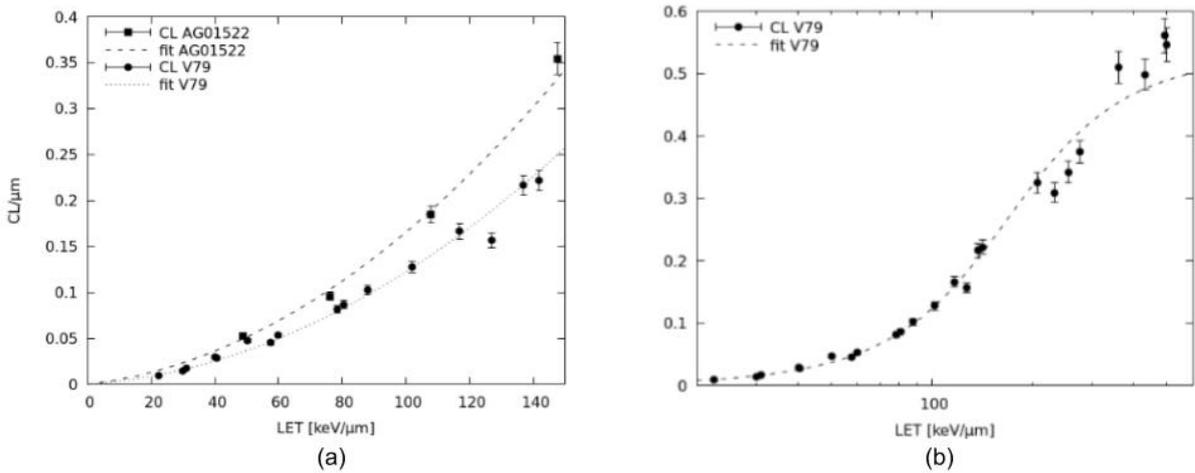


Figure 33 – (a) CL yields used to simulate the survival of carbon-irradiated AG01522 cells (upper line and points) and V79 cells (lower line and points) up to $\sim 150 \text{ keV}/\mu\text{m}$. Each point represents the mean number of CLs per micrometer used as a code input, whereas the lines are linear-quadratic fitting functions. A 5% relative error was assigned to each point. (b) CL yields used to simulate the survival of Carbon-irradiated V79 cells over the whole LET range ($22.5 - 502 \text{ keV}/\mu\text{m}$). Each point represents the mean number of CLs per micrometre used as a code input, whereas the line is a fit of the form $c \cdot \arctan(aL + bL^2)$. A 5% relative error was assigned to each point. Source: Carante *et al.* PMB. (2018) [24].

The LET-dependence of the CL yield (mean number of Cluster Lesions per μm) for Carbon ions is illustrated in Figure 33. Panel (a) shows the CL yield used for V79 and AG01522 cells irradiated by C-ions with LET below $150 \text{ keV}/\mu\text{m}$, thus excluding the over-killing region shown by V79 cells at higher LET. In this LET interval, like for protons, the LET-dependence of the CL yield was well described by a linear-quadratic function of the form $Y(L) = aL + bL^2$ for both cell lines. Moreover, the function obtained by multiplying the V79 fitting function by a fixed factor, fits reasonably well the CL yields of AG01522 cells. For each of the two cell lines, the CL fitting function used for Carbon was lower than that used for protons. This is consistent with the fact that heavier ions, which at a given LET have higher-energy secondary electrons, have a wider track, and thus are less effective in inducing DNA cluster damage. Panel (b) analyzes the whole LET interval considered for V79 cells, which showed the so-called over-killing phenomenon above $\sim 150 \text{ keV}/\mu\text{m}$. Interestingly, while above $\sim 150 \text{ keV}/\mu\text{m}$ the mean number of CLs per unit dose ($CL/(Gy \cdot cell)$) starts decreasing, the mean number of CLs per unit track length ($CL/\mu\text{m}$) continues to increase (although less rapidly), until it seems to reach a plateau. This can be explained by considering the relationship reported in Equation 3.3, according to which the mean number of CLs per unit track length is proportional to the product between the mean number of CLs per unit dose and the radiation LET. Due to the less rapid increase of the yield of CLs per micrometer in the over-killing region, the linear-quadratic fit was not applicable to the entire LET interval ($22.5 - 502 \text{ keV}/\mu\text{m}$). Although this issue deserves further investigation in the future, it is worth mentioning that, for instance, a function of the form $c \cdot \arctan(aL + bL^2)$ provided a much better fit, as shown in Figure 33(b).

4.4.2.3 Helium ions

The model was then applied to Helium ions, for which there is a renewed interest in the hadrontherapy community (e.g. [251]). For V79 cells, we considered the paper by Furusawa *et al.* [25] used for Carbon ions, which also reports data on V79 cells exposed to ^3He -ions over a LET range from 18.6 to 90.8 $\text{keV}/\mu\text{m}$. Survival curves were simulated for all 14 LET values reported in the experimental paper. To avoid making the figure too difficult to read, only some of them are reported in Figure 34(a). Like for protons and Carbon ions, all the curves for He-irradiated V79 cells were obtained with $f = 0.08$. CL yields in the range $\sim 0.009 - 0.132 \text{ CL}/\mu\text{m}$ provided simulation outcomes in good agreement with the experimental data fits reported in [25]. Concerning the nine curves that are not shown in Figure 34(a), for six of them the maximum percentage displacement between simulated curve and data fit was $\sim 30\%$, for two of them it was $\sim 50\%$, and only for the remaining one the agreement was worse. Less data is available in the literature for He-ion-irradiated normal human fibroblasts.

In this work we considered those reported by Hamada *et al.* [13], who exposed AG01522 cells to ^4He ions of 16.2 $\text{keV}/\mu\text{m}$, and Neti et al (2004), who irradiated AG01522 cells with ^4He ions of 132 $\text{keV}/\mu\text{m}$. Figure 34(b) reports simulated survival curves at these two LET values, together with the experimental data for comparison. Like for proton- and Carbon-irradiated AG01522 cells, the curves reported in Figure 34(b) were obtained with $f = 0.18$. CL yields of 0.010 $\text{CL}/\mu\text{m}$ (at 16.2 $\text{keV}/\mu\text{m}$) and 0.465 $\text{CL}/\mu\text{m}$ (at 132 $\text{keV}/\mu\text{m}$) provided a good correspondence with the considered data, since the simulations were within the error bars or very close to them, except for the two highest doses of the two curves. The reduced chi-square was 1.5 at 16.2 $\text{keV}/\mu\text{m}$, and 3.9 at 132 $\text{keV}/\mu\text{m}$.

Figure 35 reports the mean number of Cluster Lesions per μm used to simulate the survival of He-irradiated V79 and AG01522 cells. For both cell lines the CL yield increased with LET in the considered LET interval, and it was well fitted by a linear-quadratic function of the form $Y(L) = aL + bL^2$. Also in this case, by multiplying the fit for V79 cells by a fixed factor, one obtains a function that fits well the CL yields of AG01522 cells. For each of the two cell lines, the CL fitting function used for He-ions was intermediate between the function used for protons and that used for C-ions. As discussed in subsection 4.4.2.2, this is consistent with the fact that different particles having the same LET have different track-structure properties, which make heavier particles less effective in the induction of DNA cluster damage.

4.4.2.4 Full predictions of ion-survival based on a reference cell line

As discussed above, in principle the CL yield for a given radiation quality (i.e., particle type and energy) should be adjusted separately for each considered cell line, since this parameter also depends on the target cell features: in general, for a given radiation quality, more radiosensitive cells require higher CL yields than less radiosensitive ones. However, as mentioned in the previous sections, the LET-dependence of the CL yield (mean number of CLs per unit traversal length) for the two considered cell lines showed a similar shape, since by multiplying the fitting

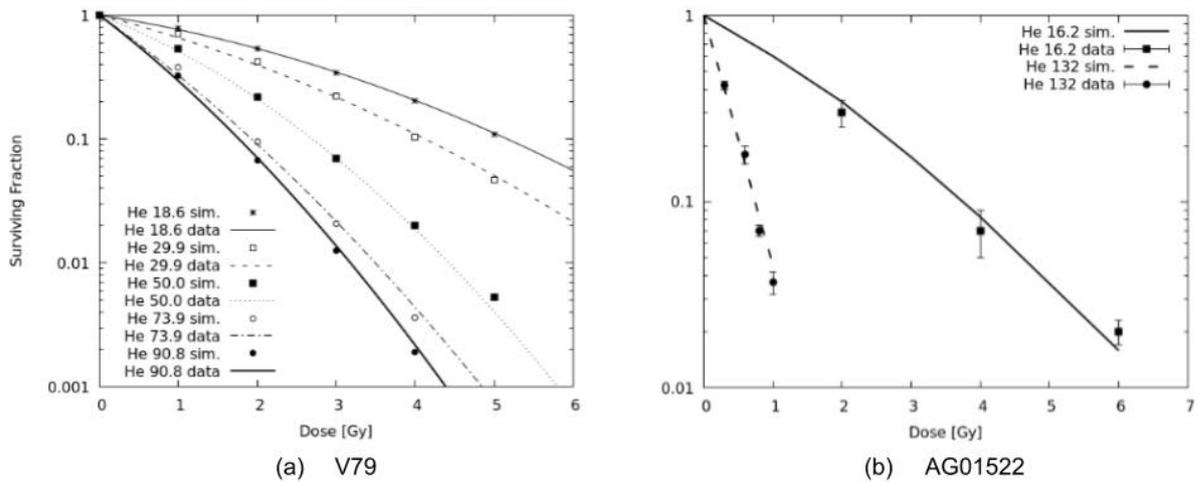


Figure 34 – (a) Survival of V79 cells irradiated by He-ion beams with different (dose-averaged) LET (from top to bottom: 18.6, 29.9, 50.0, 73.9 and 90.8 $keV/\mu m$). The points are simulation outcomes, whereas the lines are experimental data fits of the form $S(D) = \exp(-\alpha D - \beta D^2)$ taken from Furusawa *et al.* (2000) [25]. (b) Survival of AG01522 cells irradiated by He-ion beams of 16.2 $keV/\mu m$ (upper line and points) and 132 $keV/\mu m$ (lower line and points). The lines are simulation outcomes, whereas the points are experimental data taken from Hamada *et al.* (2006) (16.2 $keV/\mu m$) [13] and Neti *et al.* (2004) (132 $keV/\mu m$) [14]. Source: Carante *et al.* PMB. (2018) [24].

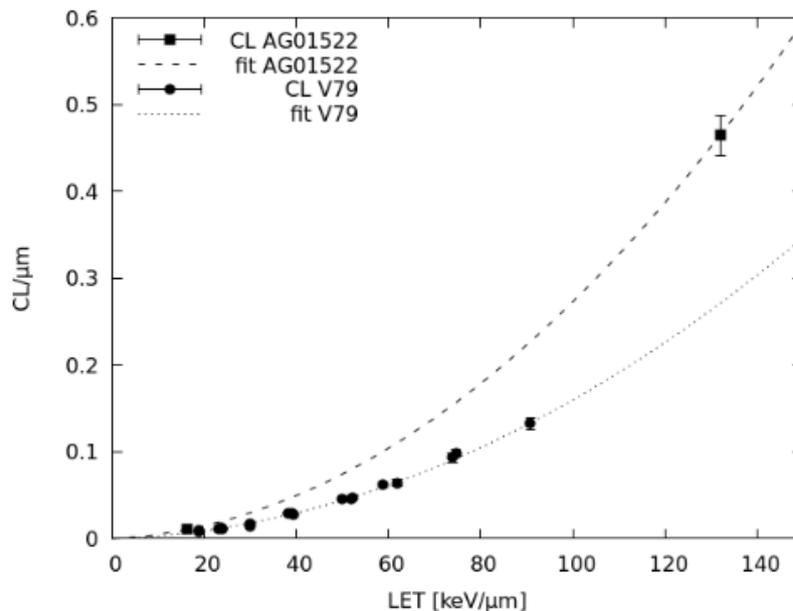


Figure 35 – CL yields used to simulate the survival of He-ion-irradiated V79 cells (lower line and points) and AG01522 cells (upper line and points). Each point represents the mean number of CLs per micrometre used as a code input, whereas the lines are linear-quadratic fits. A 5% relative error was assigned to each CL yield. Source: Carante *et al.* PMB. (2018) [24].

function of V79 cells by a given factor we obtained reasonable fits for the CL yields of AG01522 cells. If this factor can be obtained by comparing the photon responses of the two cell lines, this would represent a further improvement of the model, since it would allow to fully predict the ion-survival of the cell line of interest based on the ion-survival of a reference cell line, as well as the photon-response of each one.

To explore this possibility, we performed full predictions of proton survival for AG01522 cells based on the proton survival of V79 cells. More specifically, the mean number of CLs per μm to predict the survival of AG01522 cells at a given LET was derived by multiplying the value of the CL fitting function of V79 cells at that LET, by the following factor:

$$[(CL(Gy \cdot cell)^{-1})_{AG,X}/(CL(Gy \cdot cell)^{-1})_{V79,X}] \cdot V_{V79}/V_{AG} \quad (4.1)$$

In the expression above, $(CL(Gy \cdot cell)^{-1})_{AG,X}$ and $(CL(Gy \cdot cell)^{-1})_{V79,X}$ are the photon CL yields used for AG01522 and V79 cells, respectively, whereas V_{AG} and V_{V79} are their nucleus volumes. According to [Equation 3.3](#), this formula assumes that the ratio between the CL yields of the two cell lines is independent of LET.

For each of the proton LET values considered for AG01522 cells in [Figure 29\(b\)](#) (0.63, 1.68, 2.45 and 7.50 $\text{keV}/\mu\text{m}$), we took the proton CL yield of V79 cells from the fit reported in [Figure 30](#), and we inserted in [Equation 4.1](#) the x-ray CL yields of the two cell lines, as well as their nucleus volumes. Using the CL yields for AG01522 cells obtained in this way, we ran simulations at these four LET values, and compared the simulation outcomes -which are full predictions performed without any parameter adjustment -with the experimental data reported in [Figure 29\(b\)](#).

[Figure 36](#) reports the results obtained. In spite of a general tendency to overestimate the survival at 7.50 $\text{keV}/\mu\text{m}$, for the other three LET values the predictions were in very good agreement with the data: the simulation outcomes were within the experimental error bar for each considered experimental point with a single exception (the highest dose at 2.45 $\text{keV}/\mu\text{m}$), and the reduced chi-square was smaller than 0.7. To test if this method can be extended to other cell lines, the approach described above was also applied to proton-exposed U87 glioma cells. The experimental data for comparison was taken from the work by Chaudhary et al, who irradiated U87 cells (as well as AG01522 cells, as discussed above) at six depth positions (corresponding to LET values in the range 1.1 – 22.6 $\text{keV}/\mu\text{m}$) of a therapeutic proton beam available at INFN-LNS in Catania, Italy [27]. The nucleus of U87 cells was modelled as a cylinder with circular base, with a $98.6\mu\text{m}^2$ cross-sectional area and 7.7 μm thickness (Berardinelli, personal communication). Since in these cells the modal number of chromosomes is 44, very similar to the chromosome content of normal (human) cells, the number of chromosomes in the nucleus was left unchanged with respect to the AG01522 simulations. Analogous to AG01522 cells, the proton CL yields to perform full predictions of U87 cell survival were obtained by multiplying the V79

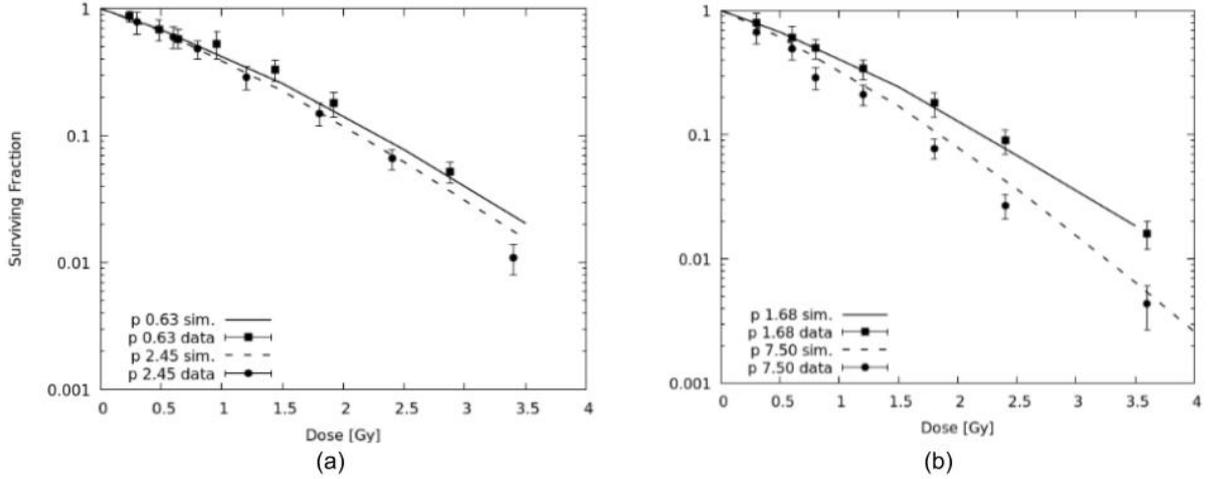


Figure 36 – (a) Survival of AG01522 cells irradiated with protons of $0.63 \text{ keV}/\mu\text{m}$ (upper line and points) and $2.45 \text{ keV}/\mu\text{m}$ (lower line and points). The lines are model full predictions, the points are experimental data taken from Marshall *et al.* (2016) [23]. (b) Survival of AG01522 cells irradiated with proton beams of $1.68 \text{ keV}/\mu\text{m}$ (upper line and points) and $7.50 \text{ keV}/\mu\text{m}$ (lower line and points). The lines are model full predictions (see the text for the details), the points are experimental data taken from Marshall *et al.* (2016) [23]. Source: Carante *et al.* PMB. (2018) [24].

(proton) CL yield by the following factor, which is analogous to that reported in Equation 4.1:

$$[(CL(\text{Gy} \cdot \text{cell})^{-1})_{U87,X}/(CL(\text{Gy} \cdot \text{cell})^{-1})_{V79,X}] \cdot V_{V79}/V_{U87} \quad (4.2)$$

The photon CL yield for U87 cells, as well as their chromosome fragment un-rejoining probability f , was adjusted to the x-ray curve reported in [27]. $2.3 \text{ CL}/(\text{Gy} \cdot \text{cell})^{-1}$ and $f = 0.03$ provided a very good agreement with the experimental data, since the reduced chi-square was 0.9. The full predictions for the six considered LET values are reported in Figure 37. At the four lower LET values ($1.1, 4.0, 7.0$ and $11.9 \text{ keV}/\mu\text{m}$), good agreement was obtained between simulations and data, with a reduced chi-square between 0.3 and 1.7. At the two higher LET values (18.0 and $22.6 \text{ keV}/\mu\text{m}$), the simulations showed a tendency to underestimate the surviving fraction at the higher doses (as well as to overestimate it at low doses). As discussed in subsection 4.4.2.1, this seems to be a general behavior of the model at these intermediate LET values, which may be corrected in the future by applying a chromosome fragment end-joining probability that decreases monotonically with increasing the fragment distance. Furthermore, it is important to take into account that all simulations reported in Figure 37 were performed without any parameter adjustment.

Although the proposed method cannot be generalized at this stage, because it needs to be tested for other cell lines and other radiation qualities, these results are encouraging in view of a future development of a fully-predictive model.

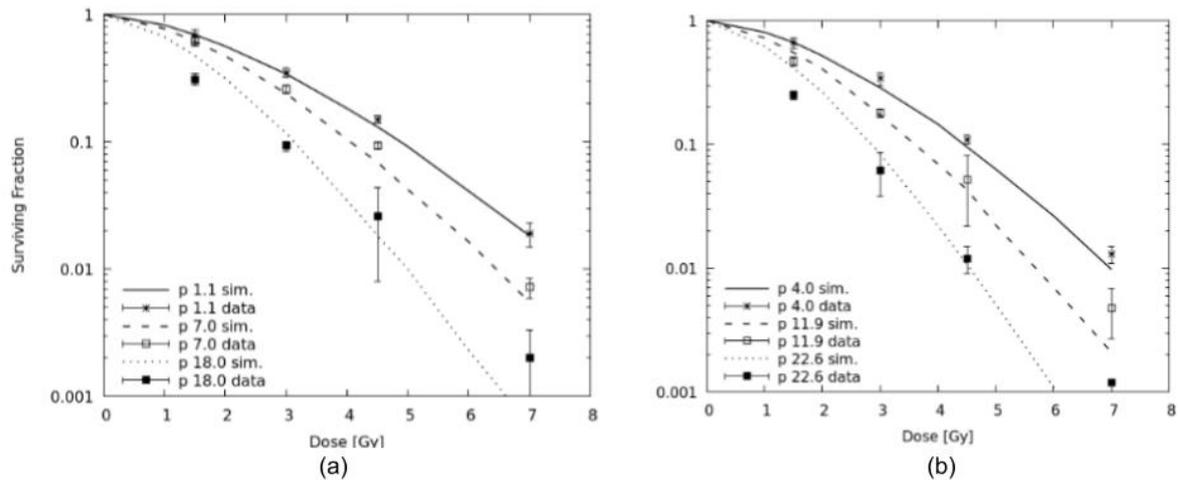


Figure 37 – (a) Survival of U87 cells irradiated by protons of different LET (from top to bottom: 1.1, 7.0 and 18.0 $keV/\mu m$). The lines are full predictions, the points are experimental data taken from Chaudhary *et al.* (2014) [27]. (b) Survival of U87 cells irradiated by protons of different LET (from top to bottom: 4.0, 11.9 and 22.6 $keV/\mu m$). The lines are full predictions, the points are experimental data taken from Chaudhary *et al.* (2014) [27]. Source: Carante *et al.* PMB. (2018) [24].

4.5 Including HiC-maps data in BIANCA

In order to further improve the geometrical modelling of the nucleus in BIANCA, it was decided to integrate the LorDG method into the code. The better the description of the target, the more accurate the biophysical models that can be developed to investigate the effects of ionizing radiation on the genome. Additionally, as the geometrical models become more detailed, the coordination with radiation transport codes becomes more feasible. Thus, the physico-chemical description of the radiation-DNA interaction could be incorporated in the biophysical modelling of chromosome aberration formation. Ideally, this could lead to the formulation of an integrated framework that spans from the 10^{-15} to the 10^4 s time scale, covering from physical events up to the observed biological effects.

4.5.1 FORTRAN implementation of LorDG

A fortran implementation of the LorDG method, which will be referred as fLorDG, is under development in order to make it an integral part of BIANCA. The initial version of the fLorDG code was implemented following the lines of the original java version described in [subsection 3.3.2.1](#). However, such implementation is not optimal in terms of memory requirements when dealing with large data sets like HiC maps at 50 $kbps$ and above. For instance, at 50 $kbps$ resolution, the number of constraints increases from 3×10^6 to 1.6×10^9 , which demands at least 33 GB of allocated memory. This number rises up to 40 GB if the allocation of wish distances is taken into account. Moreover, those estimations correspond to haploid HiC data; however, diploid models are necessary for comparison with experimental data on chromosome aberrations. A common strategy to construct diploid models out of haploid data is to simply associate each constraint to all possible combinations of the corresponding homologous chromo-

conserved. Nonetheless, individual scores of the Spearman coefficient for each chromosome are much higher and quite similar between implementations. Table 9 shows the ensemble-averaged values of the Spearman coefficient for each chromosome using fLorDG, and the corresponding LorDG values.

Table 9 – Ensemble-averaged Spearman coefficients per chromosome of models as reconstructed by the two implementations. LorDG values are taken from Table 1 in [28], which represent the ensemble average over 50 structures (no standard errors provided). fLorDG values correspond to 50 structures as well (all standard deviations were smaller than 2%).

	Chromosome											
	1	2	3	4	5	6	7	8	9	10	11	12
LorDG	0.78	0.77	0.78	0.86	0.84	0.81	0.78	0.84	0.84	0.80	0.79	0.80
fLorDG	0.80	0.77	0.73	0.89	0.85	0.85	0.79	0.86	0.86	0.84	0.82	0.82
	13	14	15	16	17	18	19	20	21	22	23	
LorDG	0.80	0.76	0.81	0.75	0.70	0.75	0.75	0.77	0.88	0.74	0.86	
fLorDG	0.82	0.78	0.81	0.75	0.73	0.75	0.76	0.69	0.90	0.70	0.89	

The results of the benchmark indicated that the implementation of fLorDG-v1 is correct. In other words, both codes are equivalent but, as mentioned, a fortran version is desirable for future integration into BIANCA.

4.5.1.1 Contributions in fLorDG

Once the method was implemented properly in fLorDG-v1, a series of versions that gradually incorporate new features into the code, started development. For instance, the second version of fLorDG included the creation of diploid models from haploid data. This was accomplished by applying the strategy outlined above, i.e., associating each constraint to the corresponding homologous pairs as well. Additionally, missing loci were inferred from the HiC data and added to the model along with adjacent interaction contact data (set equal to IF_{adj}). As a result, from this point, diploid models are available in fLorDG.

With regard to increasing the resolution of the models, two strategies were explored in a third version of the code. The first one consisted of 1) Creating 1 *Mbp* structures, 2) Loading higher resolution data (e.g. 500 *kbp*) and 3) Running the optimization method over the new data set with the previous model as the initial structure. In this strategy, it was tested both to keep the value of α from the coarser model or search it over the interval [1.0, 3.0]. A second strategy in LorDG-v3, involved an additional step before running the optimization with the second data set. As the 1 *Mbp* models were taken as initial configurations, it is reasonable to assume that a portion of the next inter-chromosomal contact data set will not be relevant in the optimization because the 1 *Mbp* would have already established inter-chromosome boundaries. Thus, the step introduced consisted of determining the set of restraints to be used during optimization, according to the initial configuration. To do so, it was decided to filter out the data related to

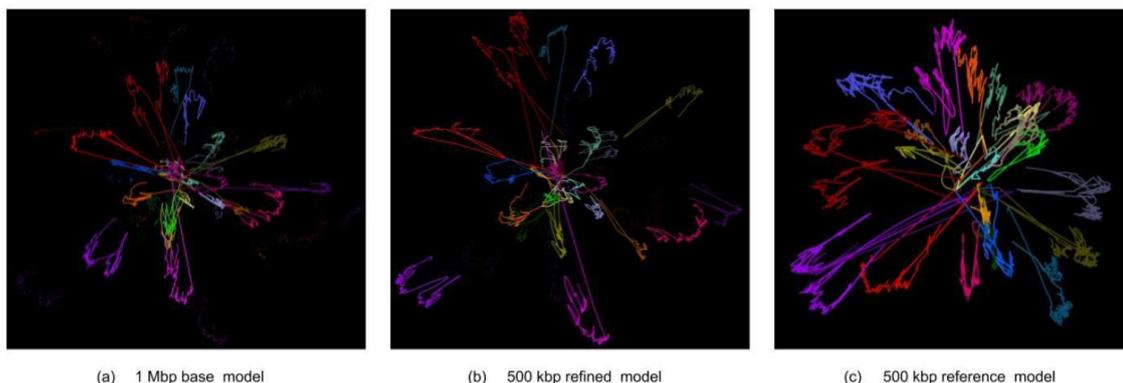


Figure 39 – (a) Base model of 1 *Mbp*. (b) 500 *kbp* Refined structured. The resolution of model (a) was doubled by using 500 *kbp* data in fLorDG-v3. The effect was an over-segregation of the chromosome territories (the topology was highly preserved) expressed as a radial shift of the chromosome with respect to the center of the nucleus. (c) 500 *kbp* Reference structure. A model at 500 *kbp* resolution was constructed directly from the corresponding HiC maps instead of applying the refinement to a 1 *Mbp* model. The structure is more compact but still, territories can be distinguished.

pairs of loci that were separated at a distance greater than d_{chro} in the coarse structure, where d_{chro} is the average distance between chromosomes as measured from their center of mass.

To test fLorDG-v3, it was ran using the 1 *Mbp* and 500 *kbp* data sets. This allowed the comparison of performance and quality of reconstructed models at 500 *kbp* resolution applying both strategies. Since it is straightforward to create 500 *kbp* structures by processing the HiC data in the usual way, these models were used as reference. The tests showed that both strategies tended to over-segregate the chromosome territories as they appeared radially shifted after refinement (see Figure 39). This behaviour was observed irrespective of the filtering distance in the second strategy. In fact, the first strategy is equivalent to the second one with $d_{chro} = \infty$. This implies that inter-chromosome contacts were considerably less satisfied because the initial configuration biases toward intra-chromosomal restraints. The reference models at 500 *kbp* resolution preserved the chromosome territory feature as expected, while keeping the nucleus more compact.

The idea behind the refinement strategy with filtered data, was to keep the size of the optimization space under control. That is, as the number of contact data increases quadratically with the resolution, a reduction in the interactions per locus to be considered represents a significant reduction in the optimization space for refinement. However, the results of the tests suggested that either the top-down or coarse-grain approach adopted in fLorDG-v3 is not appropriate for increasing resolution or it has to be applied to higher resolution initial configurations in order to have more inter-chromosome data that can influence the optimization more effectively. In any case, fLorDG-v3 highlighted the importance of being able to process larger data sets.

The latest version of fLorDG (fLorDG-v4), introduces a major series of modifications in the code to approach the large data set question. The most important one, is the implementation of an external library called *hicds* (HiC data set) that manages the input data as a database. The idea of this database is to provide tools for normalization, compression and queries of the HiC data in a faster way. Such changes aim at representing the data in binary form, reducing the size of the data files but more important, allowing faster access and queries to those files using the *hicds* library. The goal is to keep most of the information in disk and only load blocks of data to memory. Thus, many of the internal data structures of fLorDG have been redesigned accordingly. In this way, the optimization process has been adapted to this new data model. For the statistical tests, an external sort algorithm was implemented in order to calculate the rank values of the reconstructed distances for the Spearman coefficient.

All major changes that have led to a new version of fLorDG, have been benchmarked the same way as fLorDG-v1 in order to maintain consistency. The new data model incorporated in fLorDG-v4 also led to reconstructed models at 1 *Mbp* and 500 *kbp* consistent with the previous versions and LorDG. However, the true tests for fLorDG-v4 are with higher resolution maps like for example, 50 *kps* and above. Nevertheless, many adjustments to the code are still required in order to correctly process that amount of data and test the models. For instance, as the resolution increases, the HiC matrices become sparser making questionable the use of VC and KR normalization. This is being observed in the performance of the code as a dramatic increase in the running times since it takes much more steps before convergence. Also, the trade-off between lowering the memory usage and performing more reads to disk is yet to be assessed. Finally, the top-down refinement strategy have to be revisited as well.

4.5.2 Future prospects

The final goal of the project is to develop a fully atomistic model of the human genome that includes both HiC and PDB² data in order to carry out detailed simulations of the DNA-radiation interactions. High resolution models obtained with the aid of HiC maps can be further refined by filling the space between beads with the atomistic model of chromatin developed by Bernal *et al.* [252]. Such a model has been applied to the study of the mechanisms behind the formation of DSBs and their distribution in the cell nucleus [253]. This approach would help to gain insight into the role of DNA damage complexity and the formation of cluster lesions as these are key factors in the dynamics of chromosome aberration formation. Furthermore, portability and accessibility of the code is a major objective moving forward. To accomplish those goals, the data model used in the implementation is continuously revisited and cutting edge programming techniques are appraised. The challenge is to find the right balance between memory usage and computational power, required to achieve high quality models.

² The Protein Data Bank is a database for the three-dimensional structural data of large biological molecules, such as proteins and nucleic acids.

Conclusions

Along this PhD project, our in-house biophysical model/code of radiation-induced chromosome aberrations and cell death (BIANCA) was redesigned and upgraded seeking to improve its capabilities and performance. Specifically, an explicit description of interphase chromosome-arm domains and two new functions to describe proximity effects were implemented and tested. Such functions were a negative exponential- and a Gaussian- function, which served to model the interaction probabilities between chromatin free-ends, with increased sensitivity to their initial distance. Additionally, technical developments in the code were performed, in order to achieve better functionality, such as reduced execution times, micro-beam irradiation and extended scoring assessment capabilities.

To investigate the distance dependence of the rejoining probability between fragments, the model was then applied to gamma-irradiated normal human lymphocytes and fibroblasts. Such comparisons suggested that, although both functions performed better than a previously tested step-function, an exponential dependence upon the initial distance between free-ends can better describe proximity effects than a Gaussian one, at least for low-LET radiation. Afterwards, those studies were expanded to intermediate and high LET radiations.

The results suggested that, independently of radiation quality, in lymphocytes an exponential function of the form $\exp(-r/\sigma)$ can describe both the bias for inter-arm exchanges relative to inter-chromosome exchanges (*F-ratio*), and that for intra-arm to inter-arm exchanges (*G-ratio*). On the contrary, in fibroblasts this function can describe the former bias but underestimates the latter. This might be related to differences in chromatin mobility and/or in the 3D organization of the cell nucleus during interphase, as well as in the dimensions of the various chromosome territories. The fact that, with increasing LET, *F* decreased and *G* increased in both cell types, supports their possible role as "fingerprints" of high-LET exposure. A dose-dependence was also observed at high LET, where *F* increased with dose in both cell types, whereas *G* decreased in fibroblasts but not in lymphocytes. [Table 10](#) displays a schematic overview of the main findings related to the *F*- and *G*-ratios, including the agreement with experimental data, the dose-dependence and the LET-dependence.

In light of these findings, further studies, including chromosome aberration experiments with pan-telomeric probes and time-lapse observation of repair protein foci, are desirable to better characterize and quantify these effects. Furthermore, it is important that in the aberration studies, the specific conditions (e.g. dose-range, cell type and scoring criteria, including those adopted for higher-order multicentrics) are reported and taken into account when interpreting the results.

Table 10 – Agreement with data and dependence on LET and dose for the F-ratio and the G-ratio.

Cell type	Agreement with data		Dose dependence		LET dependence	
	F	G	F	G	F	G
lymph, Low-LET	OK	OK	–	–	↓	↑
lymph, High-LET	OK/under	OK	↑	–		
fibro, Low-LET	OK	under	–	–	↓	↑
lymph, Low LET	~OK	under	↑	↓		

lymph = lymphocytes; fibro = fibroblasts.

OK = good agreement; under = underestimation; OK/under = agreement for some data, underestimation for others.

– = no significance dependence; ↑ = increasing; ↓ = decreasing.

Also, two collaborations with regard to hadrontherapy applications were carried out, in which BIANCA served as a means to interpret experimental data. The partnership with the BioQuaRT project aimed at elucidating the biophysical mechanisms that govern chromosomal aberration induction by means of α -particles, with a focus on the correlation with particle track-structure features. Thus, experimental and theoretical analyses on chromosomal aberration induction in CHO-K1 cells that were individually targeted by an exact number of 5.5 or 17.8 MeV α -particles generated by a micro-beam, were performed. These analyses enabled quantification of the increased biological effectiveness of 5.5 vs. 17.8 MeV α -particles, both in terms of the percentage of aberrant cells and the aberration yields, which were found to increase by a factor of ~ 2 , consistent with data on RBE relationships available in the literature. The theoretical analysis performed with BIANCA, allowed the absorbed dose to be evaluated, and the experimental results to be interpreted in terms of DNA critical lesions, which in turn are related to radiation track structure. More specifically, the higher aberration yields observed for 5.5 MeV α -particles may be explained by the higher effectiveness of a single particle at inducing such critical lesions, which leads to an increase in the yield of cluster lesions per cell that is consistent with the increase in (total) aberrations observed in the experiments.

Later, the predictive power of the BIANCA model was tested against an experimental data set on different radiation qualities in the range of hadrontherapy applications. The good agreement between simulations and experimental data suggested that the upgraded model developed in this PhD, is suitable for calculating the biological effectiveness of hadrontherapy beams, especially considering that predictions can be performed also at LET values for which there is no experimental data. Furthermore, an approach was proposed to fully predict the ion-survival fraction of cell line(s) of interest based on the same quantity for a reference cell line, and the photon response of both cell lines.

Finally, an ongoing project seeking the integration of HiC data into the geometrical modelling in BIANCA, was initiated during this PhD project. Currently, the LorDG method

has been implemented and optimized in such a way that intermediate resolution (~ 100 *kbp*) diploid models of the entire genome can be constructed using HiC data. In order to increase the resolution of the models, a top-down/coarse-grain strategy has been tested with 500*kbp* and 1 *Mbp* HiC data. Preliminary results suggest that the strategy needs to be applied to higher initial resolution data sets before refinement. However, using high resolution HiC maps requires a careful examination of the data normalization. Currently, attention is being paid to the implementation of cutting edge programming techniques in order to process the vast amount of data comprising high resolution maps more efficiently.

List of Publications

Articles

- J. J. Tello, M. P. Carante, M. Bernal and F. Ballarini (2017). Proximity effects in chromosome aberration induction by low-LET ionizing radiation. *DNA Repair* 58, 38-46
- J. J. Tello, M. P. Carante, M. Bernal and F. Ballarini (2018). Proximity effects in chromosome aberration induction: dependence on radiation quality, cell type and dose. *DNA Repair* 64, 45-52
- A Testa, F. Ballarini, U. Giesen, O. Monteiro Gil, M. P. Carante, J. J. Tello, F. Langner, H. Rabus, V. Palma, M. Pinto, C. Patrono (2018). Analysis of radiation-induced chromosomal aberrations on cell-by-cell basis after 4He -ion microbeam irradiation: experimental data and simulations. *Radiation Research* 189(6), 597-604
- M. P. Carante, C. Aimè, J. J. Tello and F. Ballarini (2018). BIANCA, a biophysical model of cell survival and chromosome damage by protons, C-ions and He-ions at energies and doses used in hadrontherapy. *Physics in Medicine and Biology* 63(7), 075007
- F. Ballarini, M. P. Carante, J. J. Tello Cajiao, M. A. Bernal (2018). Biophysical modelling of proximity effects in chromosome aberration production. *Nuovo Cimento C* 41(6), 205
- M. P. Carante, J. J. Tello and F. Ballarini (2019). Predicting biological effects along hadrontherapy dose profiles by the BIANCA biophysical model. *Radiation Protection Dosimetry* 183, 111-115
- M. P. Carante, J. J. Tello Cajiao and F. Ballarini (2019). A radiobiological database produced by the BIANCA model to predict the biological effectiveness of hadrontherapy beams. *CERN-Proceedings-2019-001* (CERN, Geneva, 2019), pp. 313-320, ISSN 2518-315X (Online), ISSN 2078-8835 (Print). [dx.doi.org/10.23727/CERN-Proceedings-2019-001](https://doi.org/10.23727/CERN-Proceedings-2019-001)

Abstracts

- M. P. Carante, J. J. Tello, M. Bernal and F. Ballarini. Full predictions of cell death and chromosome damage along hadrontherapy dose profiles by the BIANCA biophysical model. ERRS Annual meeting, Essen, Germany, September 2017
- F. Ballarini, M.P. Carante, J. J. Tello. The BIANCA biophysical model/MC code: calculations of radiation-induced cell damage in view of hadrontherapy treatments. International Conference on Monte Carlo Techniques for Medical Applications (MCMA2017), Naples, Italy, October 2017. *Physica Medica* vol 42, Supplement 1, p. 7

-
- M. P. Carante, J. J. Tello and F. Ballarini, Predicting biological effects along hadrontherapy dose profiles by the BIANCA biophysical model, MICROS 2017, Venice, November 2017
 - J. Tello, M. Carante, M. Bernal, F. Ballarini. MODELING OF CHROMOSOME ABERRATION INDUCTION BY LOW LET IONIZING RADIATION. IRPA2018, Cuba, April 2018
 - M. P. Carante, J. J. Tello and F. Ballarini. A radiobiological database produced by the BIANCA model to predict the biological effectiveness of hadrontherapy beams. 15th Varenna Conference on Nuclear Reaction mechanisms, Varenna, Italy, 11-15 June 2018
 - F. Ballarini, M. P. Carante, J. J. Tello Cajiao, RUOLO DEI "PROXIMITY EFFECTS" NELL'INDUZIONE DI ABERRAZIONI CROMOSOMICHE. XVIII Convegno nazionale SIRR, Roma, 10-13 settembre 2018
 - M. P. Carante, C. Aimè, J. J. Tello Cajiao, F. Ballarini. UN'INTERFACCIA TRA FLUKA E IL MODELLO BIOFISICO BIANCA PER PREDIRE L'EFFICACIA BIOLOGICA DI FASCI ADROTERAPICI. XVIII Convegno nazionale SIRR, Roma, 10-13 settembre 2018

Bibliography

- [1] DESOUKY, O.; DING, N.; ZHOU, G. Targeted and Non-Targeted Effects of Ionizing Radiation. *J. Radiat. Res. and Applied Sci.*, v. 8, n. 2, p. 247–254, April 2015.
- [2] FRIEDRICH, T. et al. Systematic Analysis of RBE and Related Quantities Using a Database of Cell Survival Experiments with Ion Beam Irradiation. *J. Radiat. Res.*, v. 54, p. 494–514, 2013.
- [3] EA, V. et al. Contribution of Topological Domains and Loop Formation to 3D Chromatin Organization. *Genes*, v. 6, n. 3, p. 734–750, 2015.
- [4] SAVAGE, J. R. Classification and Relationships of Induced Chromosomal Structural Changes. *J. Med. Genet.*, v. 13, n. 2, p. 103–122, April 1976.
- [5] DURANTE, M. et al. From DNA Damage to Chromosome Aberrations: Joining the Break. *Mutat. Res.*, v. 756, p. 5–13, August 2013.
- [6] OBE, G. et al. Chromosomal Aberrations: Formation, Identification, and Distribution. *Mutat. Res.*, v. 504, p. 3–16, 2002.
- [7] DURANTE, M. et al. Rejoining and Misrejoining of Radiation-Induced Chromatin Breaks. Iii. Hypertonic Treatment. *Rad. Res.*, v. 149, n. 1, p. 68–74, 1998.
- [8] GOTOH, E.; KAWATA, T.; DURANTE, M. Chromatid Break Rejoining and Exchange Aberration Formation Following γ -Ray Exposure: Analysis in G2 Human Fibroblasts by Chemically Induced Premature Chromosome Condensation. *Int. J. Radiat. Biol.*, v. 75, p. 1129–1135, November 1999.
- [9] CORNFORTH, M.; BEDFORD, J. A Quantitative Comparison of Potentially Lethal Damage Repair and the Rejoining of Interphase Chromosome Breaks in Low Passage Normal Human Fibroblasts. *Rad. Res.*, v. 111, p. 385 – 405, 1987.
- [10] MCMAHON, S.; PRISE, K. Mechanistic Modelling of Radiation Responses. *Cancers*, v. 11, n. 2, p. 205, Feb 2019.
- [11] FRIEDLAND, W.; JACOB, P.; KUNDRÁT, P. Stochastic Simulation of DNA Double-Strand Break Repair by Non-Homologous End Joining Based on Track Structure Calculations. *Radiat. Res.*, v. 173, n. 5, p. 677–688, May 2010.
- [12] BALLARINI, F. et al. A Model of Radiation-Induced Cell Killing: Insights into Mechanisms and Applications for Hadrontherapy. *Radiat. Res.*, v. 180, n. 3, p. 307–315, 2013.
- [13] HAMADA, N. et al. LET-Dependent Survival of Irradiated Normal Human Fibroblasts and Their Descendants. *Radiat. Res.*, v. 166, n. 1, p. 24–30, 2006.

- [14] NETI, P. et al. A Multi-Port Low-Fluence alpha-Particle Irradiator: Fabrication, Testing and Benchmark Radiobiology Studies. *Radiat. Res.*, v. 161, n. 6, p. 732–738, 2006.
- [15] BALLARINI, F. et al. The Bianca Model/Code of Radiation-Induced Cell Death: Application to Human Cells Exposed to Different Radiation Types. *Radiat. Environ. Biophys.*, v. 53, n. 3, p. 525–533, 2014.
- [16] CARANTE, M.; F.BALLARINI. Calculating Variations in Biological Effectiveness for a 62 MeV Proton Beam. *Front. Oncol.*, v. 6, p. 76, 2016.
- [17] DEKKER, J.; MARTI-RENO, M. A.; MIRNY, L. A. Exploring the Three-Dimensional Organization of Genomes: Interpreting Chromatin Interaction Data. *Nat. Rev. Gen.*, v. 14, p. 390–403, 2013.
- [18] BAUCHINGER, M.; SCHMID, E. Let Dependence of Yield Ratios of Radiation-Induced Intra- and Interchromosomal Aberrations in Human Lymphocytes. *Int. J. Radiat. Biol.*, v. 74, p. 17–25, 1998.
- [19] TELLO, J. J. et al. Proximity Effects in Chromosome Aberration Induction by Low-LET Ionizing Radiation. *DNA Repair*, v. 58, p. 38–46, 2017.
- [20] CORNFORTH, M.; BAILEY, S.; GOODWIN, E. Dose Responses for Chromosome Aberrations Produced in Noncycling Primary Human Fibroblasts by alpha Particles, and by gamma Rays Delivered at Sublimiting Low Dose Rates. *Radiat. Res.*, v. 158, p. 43–53, 2002.
- [21] TELLO, J. J. et al. Proximity Effects in Chromosome Aberration Induction: Dependence on Radiation Quality, Cell Type and Dose. *DNA Repair*, v. 64, p. 45–52, 2018.
- [22] BELLI, M. 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. Rad. Biol.*, v. 74, n. 4, p. 501–509, 1998.
- [23] MARSHALL, T. I. et al. Investigating the Implications of a Variable RBE on Proton Dose Fractionation Across a Clinical Pencil Beam Scanned Spread-Out Bragg Peak. *Int. J. Radiat. Oncol. Biol. Phys.*, v. 95, p. 70–75, 2016.
- [24] CARANTE, M. P. et al. Bianca, a Biophysical Model of Cell Survival and Chromosome Damage by Protons, C-Ions and He-Ions at Energies and Doses Used in Hadrontherapy. *Phys. Med. Biol.*, v. 63, n. 7, p. 075007, March 2018.
- [25] FURUSAWA, Y. et al. Inactivation of Aerobic and Hypoxic Cells from Three Different Cell Lines by Accelerated ^3He -, ^{12}C - and ^{20}Ne -Ion Beams. *Radiat. Res.*, v. 154, p. 485–496, 2000.
- [26] KAVANAGH, J. N. et al. Antiproton Induced DNA Damage: Proton Like in Flight, Carbon-Ion Like Near Rest. *Scientific Reports*, v. 3, 2013.
- [27] CHAUDHARY, P. et al. Relative Biological Effectiveness Variation Along Monoenergetic and Modulated Bragg Peaks of a 62-MeV Therapeutic Proton Beam: a Preclinical Assessment. *Int. J. Radiat. Oncol. Biol. Phys.*, v. 90, p. 27–35, 2014.

- [28] TRIEU, T.; CHENG, J. 3D Genome Structure Modeling by Lorentzian Objective Function. *Nuc. Acids Res.*, v. 45, n. 3, p. 1049–1058, 2017.
- [29] KUZIN, A. M. On the Role of DNA in the Radiation Damage of the Cell. *Int. J. Radiat. Biol.*, v. 6, p. 201–209, 1963.
- [30] KADHIM, M. et al. Transmission of Chromosomal Instability After Plutonium alpha-Particle Irradiation. *Nature*, v. 355, p. 738–740, 1992.
- [31] NAGASAWA, H.; LITTLE, J. Induction of Sister Chromatid Exchanges by Extremely Low Doses of alpha-Particles. *Cancer Res.*, v. 52, p. 6394–6396, 1992.
- [32] JACOB, P. et al. Possible Expressions of Radiation-Induced Genomic Instability, Bystander Effects Or Low-Dose Hyper-Sensitivity in Cancer Epidemiology. *Mutat. Res.*, v. 687, p. 34–39, 2010.
- [33] BAVERSTOCK, K. Why Do We Need a New Paradigm in Radiobiology?. *Mutat. Res.*, v. 687, p. 3–6, 2010.
- [34] UNSCEAR2008. *Sources and Effects of Ionizing Radiation*. New York, 2010.
- [35] HALL, E. J. *Radiobiology for the Radiologist*. 7th. ed. Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins, 2012.
- [36] IAEA. *Diagnostic Radiology Physics: a Handbook for Teachers and Students*. Vienna: IAEA, 2014.
- [37] WHITE, D. R. et al. Tissue Substitutes in Radiation Dosimetry and Measurement. *J. of the ICRU*, v. 23, Jan 1989.
- [38] BOHR, N. On the Decrease of Velocity of Swiftly Moving Electrified Particles in Passing Through Matter. *Phil. Mag and J. Sci.*, v. 30, p. 581–612, 1915.
- [39] BETHE, H. Zur Theorie Des Durchgangs Schneller Korpuskularstrahlen Durch Materie. *Ann. der Physik*, v. 397, p. 325–400, 1930.
- [40] BLOCH, F. Zur Bremsung Rasch Bewegter Teilchen Beim Durchgang Durch Materie. *Ann. der Physik*, v. 408, p. 285–320, 1930.
- [41] LEO, W. R. *Techniques for Nuclear and Particle Physics Experiments*. 2nd. ed. Heidelberg: Springer-Verlag, 2012.
- [42] MALMER, C. J. ICRU Report 63. Nuclear Data for Neutron and Proton Radiotherapy and for Radiation Protection. *AAPM*, v. 28, p. 816, May 2001.
- [43] ICRP-2007. *The 2007 Recommendations of the International Commission on Radiological Protection*. Essen: Elsevier, 2007. (ICRP Publication 103, 2-4).
- [44] IAEA. *Radiation Biology: a Handbook for Teachers and Students*. Austria: IAEA Publishing Section, 2010. (42, I).

- [45] IAEA. *Radiation Oncology Physics: a Handbook for Teachers and Students*. Vienna: IAEA Publishing Section, 2005.
- [46] GRIFFITHS, A. J. et al. *An Introduction to Genetic Analysis*. 7th. ed. New York: W. H. Freeman, 2000.
- [47] WATSON, J. D.; CRICK, F. H. C. Molecular Structure of Nucleic Acids: a Structure of the Deoxyribose Nucleic Acid. *Nature*, v. 171, p. 737–738, 1953.
- [48] VISCHER, E.; CHARGAFF, E. The Separation and Quantitative Estimation of Purines and Pyrimidines in Minute Amounts. *J. Biol. Chem*, v. 176, p. 703–714, 1948.
- [49] CHARGAFF, E. Chemical Specificity of Nucleic Acids and Mechanism of Their Enzymatic Degradation. *Experientia*, v. 6, p. 201–209, 1950.
- [50] WILKINS, A. R. S. M. F. H.; WILSON, H. R. Molecular Structure of Nucleic Acids: Molecular Structure of Deoxypentose Nucleic Acids. *Nature*, v. 171, p. 738–740, 1953.
- [51] ROSALIND, E. F.; GOSLING, R. G. Molecular Configuration in Sodium Thymonuclate. *Nature*, v. 171, p. 740–741, 1953.
- [52] ZALLEN, D. T. Despite Franklin'S Work, Wilkins Earned His Nobel. *Nature*, v. 425, p. 15, 2003.
- [53] DICKERSON, R. E. et al. The Anatomy of A-, B- and Z-DNA. *Science*, v. 216, p. 475–485, 1982.
- [54] KNEE, K. M. et al. Spectroscopic and Molecular Dynamics Evidence for a Sequential Mechanism for the A-To-B Transition in DNA. *Biophys. J.*, v. 95, n. 7, p. 257–272, 2008.
- [55] WANG, A. H. et al. Molecular Structure of a Left-Handed Double Helical DNA Fragment at Atomic Resolution. *Nature*, v. 282, p. 680–686, 1979.
- [56] THAMANN, T. J. et al. High Salt Form of Poly(Dg-Dc)-Poly(Dg-Dc) Is Left-Handed Z-DNA: Raman Spectra of Crystals and Solutions. *Nucleic Acids Res.*, v. 9, p. 5443–5457, 1981.
- [57] HA, S. C. et al. Crystal Structure of a Junction Between B-DNA and Z-DNA Reveals Two Extruded Bases. *Nature*, v. 479, n. 10, p. 1183–1186, 2005.
- [58] RICH, A.; ZHANG, S. Z-DNA: the Long Road to Biological Function. *Nat. Rev. Genet.*, v. 4, p. 566–572, July 2003.
- [59] ARNEODO, A. et al. Multi-Scale Coding of Genomic Information: from DNA Sequence to Genome Structure and Function. *Physics Reports*, v. 498, n. 2-3, p. 45–188, 2011.
- [60] HOLDE, K. E. van. *Chromatin*. New York: Springer-Verlag, 1988. (Springer series in molecular and cell biology).
- [61] WOLFFE, A. P. *Chromatin Structure and Function*. 3rd. ed. London: Academic Press London, 1998.

- [62] LUGER, K. et al. Crystal Structure of the Nucleosome Core Particle at 2.8 Å Resolution. *Nature*, v. 389, p. 251–260, 1997.
- [63] WIDOM, J. Structure Dynamics and Function of Chromatin in Vitro. *Annu. Rev. Biophys. Biomol. Struct.*, v. 27, p. 285–327, 1998.
- [64] KORNBERG, R.; LORCH, Y. Twenty-Five Years of the Nucleosome, Fundamental Particle of the Eukaryote Chromosomes. *Cell*, v. 98, p. 285–294, 1999.
- [65] RICHMOND, T. J.; DAVEY, C. A. The Structure of DNA in the Nucleosome Core. *Nature*, v. 423, p. 145–150, 2003.
- [66] CHANTALAT, L. et al. Structure of the Histone-Core Octamer in KCl/Phosphate Crystals at 2.15 Å Resolution. *Acta Crystallogr. D Biol. Crystallogr.*, v. 59, p. 1395–1407, 2003.
- [67] ALBERTS, B. et al. *Molecular Biology of the Cell*. 4th. ed. New York: Garland Publishing, 2002.
- [68] SATCHWELL, S.; DREW, H.; TRAVERS, A. Sequence Periodicities in CHiCken Nucleosome Core DNA. *J. Mol. Biol.*, v. 191, p. 659–675, 1986.
- [69] IOSHIKHES, I. et al. Nucleosome DNA Sequence Pattern Revealed by Multiple Alignment of Experimentally Mapped Sequences. *J. Mol. Biol.*, v. 262, p. 129–139, 1996.
- [70] WIDOM, J. Role of DNA Sequence in Nucleosome Stability and Dynamics. *Q. Rev. Biophys.*, v. 34, p. 269–324, 2001.
- [71] AUDIT, B. et al. Long-Range Correlations in Genomic DNA: a Signature of the Nucleosomal Structure. *Phys. Rev. Lett.*, v. 86, p. 2471–2474, 2001.
- [72] AUDIT, B. et al. Long-Range Correlations Between DNA Bending Sites: Relation to the Structure and Dynamics of Nucleosomes. *J. Mol. Biol.*, v. 316, p. 903–918, 2002.
- [73] AUDIT, B. et al. Wavelet Analysis of DNA Bending Profiles Reveals Structural Constraints on the Evolution of Genomic Sequences. *J. Biol. Phys.*, v. 30, p. 33–81, 2004.
- [74] LAEMMLI, U. et al. Scaffold-Associated Regions: Cis-Acting Determinants of Chromatin Structural Loops and Functional Domains. *Current Opin. Genetics Dev.*, v. 2, p. 275–285, 1992.
- [75] SAITOH, Y.; LAEMMLI, U. From the Chromosomal Loops and the Scaffold to the Classic Bands of Metaphase Chromosomes. *Cold Spring Harb. Symp. Quant. Biol.*, v. 58, p. 755–765, 1993.
- [76] COOK, P. A Chromomeric Model for Nuclear and Chromosome Structure. *J. Cell Sci.*, v. 108, p. 2927–2935, 1995.
- [77] COOK, P. Predicting Three-Dimensional Genome Structure from Transcriptional Activity. *Nat. Genet.*, v. 32, p. 347–352, 2002.

- [78] MAHY, N.; PERRY, P.; BICKMORE, W. Gene Density and Transcription Influence the Localization of Chromatin Outside of Chromosome Territories Detectable by Fish. *J. Cell Biol.*, v. 159, p. 753–763, 2002.
- [79] LIEBERMAN-AIDEN, E. et al. Comprehensive Mapping of Long Range Interactions Reveals Folding Principles of the Human Genome. *Science*, v. 326, n. 5950, p. 289–293, 2009.
- [80] MÜNDEL, C.; LANGOWSKI, J. Chromosome Structure Predicted by a Polymer Model. *Phys. Rev. E*, v. 57, n. 5888, 1998.
- [81] MATEOS-LANGERAK, J. et al. Spatially Confined Folding of Chromatin in the Interphase Nucleus. *PNAS*, v. 106, p. 3812–3817, March 2009.
- [82] GROSBERG, A. Y.; NECHAEV, S.; SHAKHNOVICH, E. The Role of Topological Constraints in the Kinetics of Collapse of Macromolecules. *J. Phys. France*, v. 49, n. 12, p. 2095–2100, 1988.
- [83] GROSBERG, A. et al. Crumpled Globule Model of the Three-Dimensional Structure of DNA. *Europhysics Letters (EPL)*, v. 23, n. 5, p. 373–378, 1993.
- [84] VISSER, A. et al. High Resolution Analysis of Interphase Chromosome Domains. *J. Cell Sci.*, v. 113, p. 2585–2593, 2000.
- [85] CREMER, T.; CREMER, C. Chromosome Territories, Nuclear Architecture and Gene Regulation in Mammalian Cells. *Nature Reviews Genetics*, v. 2, p. 292–301, 2001.
- [86] PERIC-HUPKES, D. et al. Molecular Maps of the Reorganization of Genome-Nuclear Lamina Interactions During Differentiation. *Mol. Cell*, v. 38, n. 4, p. 603–613, May 2010.
- [87] CHAUMEIL, J. et al. A Novel Role for Xist Rna in the Formation of a Repressive Nuclear Compartment into WHiCh Genes Are Recruited When Silenced. *Genes Dev.*, v. 20, n. 16, p. 2223–2237, Aug 2006.
- [88] SEXTON, T.; CAVALLI, G. The Role of Chromosome Domains in Shaping the Functional Genome. *Cell*, v. 160, p. 1049–1059, March 2015.
- [89] CHADWICK, K.; LEENHOUTS, H. *The Molecular Theory of Radiation Biology*. 1. ed. Berlin: Springer-Verlag Berlin Heidelberg, 1981. (Monographs on Theoretical and Applied Genetics, v. 5).
- [90] MORENO, S.; HAYLES, J.; NURSE, P. Regulation of P34Cdc2 Protein Kinase During Mitosis. *Cell*, v. 58, n. 2, p. 361–372, July 1989.
- [91] GOTOH, E.; ASAKAWA, Y.; KOSAKA, H. Inhibition of Protein Serine/Threonine Phosphatases Directly Induces Premature Chromosome Condensation in Mammalian Somatic Cells. *Biom. Res.*, v. 16, n. 1, p. 63–68, Jan 1995.

- [92] DURANTE, M.; FURUSAWA, Y.; GOTOH, E. A Simple Method for Simultaneous Interphase-Metaphase Chromosome Analysis in Biodosimetry. *Int. J. Radiat. Bio.*, v. 74, n. 4, p. 457–462, Oct 1998.
- [93] SIMPSON, J. P.; SAVAGE, J. R. Detecting 'Hidden' Exchange Events Within X-ray-Induced Aberrations Using Multicolour Chromosome Paints. *Chromosome Res.*, v. 3, n. 1, p. 69–72, Jan 1995.
- [94] SAVAGE, J. R.; SIMPSON, J. P. Fish 'Painting' Patterns Resulting from Complex Exchanges. *Mutat. Res.*, v. 312, n. 1, p. 51–60, Feb 1994.
- [95] TUCKER, J. D. et al. A Proposed System for Scoring Structural Aberrations Detected by Chromosome Painting. *Cytogenet. Cell. Genet.*, v. 68, n. 3-4, p. 211–221, 1995.
- [96] ROTHKAMM, K.; LÖBRICH, M. Evidence for a Lack of DNA Double-Strand Break Repair in Human Cells Exposed to Very Low X-ray Doses. *Proc. Natl. Acad. Sci. U. S. A.*, v. 100, p. 5057–5062, 2003.
- [97] KARRAN, P. DNA Double Strand Break Repair in Mammalian Cells. *Curr. Opin. Genet. Dev.*, v. 10, p. 144–150, 2000.
- [98] WETERINGS, E.; GENT, D. van. The Mechanism of Non-Homologous End-Joining: a Synopsis of Synapsis. *DNA repair*, v. 3, p. 1425–1435, 2004.
- [99] LIANG, F. et al. Homology-Directed Repair Is a Major Double-Strand Break Repair Pathway in Mammalian Cells. *Proc. Natl. Acad. Sci. U.S.A.*, v. 95, p. 5172–5177, 1998.
- [100] JOHNSON, R.; LIU, N.; JASIN, M. Mammalian Xrcc2 Promotes the Repair of DNA Double-Strand Breaks by Homologous Recombination. *Nature*, v. 401, n. 6751, p. 397–399, september 1999.
- [101] PIERCE, A. et al. Xrcc3 Promotes Homology-Directed Repair of DNA Damage in Mammalian Cells. *Genes Dev.*, v. 13, p. 2633–2638, 1999.
- [102] JONES, N.; STEWART, S.; THOMPSON, L. Biochemical and Genetic Analysis of the Chinese Hamster Mutants Irs1 and Irs2 and Their Comparison to Cultured Ataxia Telangiectasia Cells. *Mutagenesis*, v. 5, p. 15–23, 1990.
- [103] BLUNT, T. et al. Defective DNA-Dependent Protein Kinase Activity Is Linked to V(D)J Recombination and DNA Repair Defects Associated with the Murine Scid Mutation. *Cell*, v. 80, p. 813–823, 1995.
- [104] HINZ, J. et al. Repression of Mutagenesis by Rad51D-Mediated Homologous Recombination. *Nucleic Acids Res.*, v. 34, p. 1358–1368, 2006.
- [105] NAGASAWA, H. et al. Some Unsolved Problems and Unresolved Issues in Radiation Cytogenetics: a Review and New Data on Roles of Homologous Recombination and Non-Homologous End Joining. *Mutat. Res. Gen. Toxicology and Env. Mutagenesis*, v. 701, n. 1, p. 12–22, August 2010.

- [106] HABER, J. E. Partners and Pathways: Repairing a Double-Strand Break. *Trends in Genetics*, v. 16, p. 259–264, june 2000.
- [107] ASAITHAMBY, A. et al. Repair of Hze-Particle-Induced DNA Double-Strand Breaks in Normal Human Fibroblasts. *Rad. Res.*, v. 169, n. 4, p. 437–446, april 2008.
- [108] COSTES, S. et al. Spatiotemporal Characterization of Ionizing Radiation Induced DNA Damage Foci and Their Relation to Chromatin Organization. *Mutat. Res.*, v. 704, p. 78–87, 2010.
- [109] BROWN, J. M.; EVANS, J. W.; KOVACS, M. S. Mechanism of Chromosome Exchange Formation in Human Fibroblasts: Insights from 'Chromosome Painting'. *Environmental and Molecular Mutagenesis*, v. 22, p. 218–224, 1993.
- [110] DARROUDI, F. et al. Kinetics of the Formation of Chromosome Aberrations in X-Irradiated Human Lymphocytes, Using Pcc and Fish. *Mutat. Res.*, v. 404, p. 55–65, 1998.
- [111] SHIBATA, A. et al. Factors Determining DNA Double-Strand Break Repair Pathway Choice in G2 Phase. *The EMBO Journal*, v. 30, p. 1079–1092, November 2011.
- [112] VERKAIK, N. S. et al. Different Types of V(D)J Recombination and End-Joining Defects in DNA Double-Strand Break Repair Mutant Mammalian Cells. *EJI*, v. 32, n. 3, p. 701–709, 2002.
- [113] ILIAKIS, G. et al. Mechanisms of DNA Double Strand Break Repair and Chromosome Aberration Formation. *Cytogenet. Genome Res.*, v. 104, n. 1-4, p. 14–20, 2004.
- [114] MLADENOV, E.; ILIAKIS, G. Induction and Repair of DNA Double Strand Breaks: the Increasing Spectrum of Non-Homologous End Joining Pathways. *Mutat. Res.*, v. 711, p. 61–72, 2011.
- [115] YAN, C. T. et al. Igh Class Switching and Translocations Use a Robust Non-Classical End-Joining Pathway. *Nature*, v. 449, p. 478–482, 2007.
- [116] ZHANG, Y.; JASIN, M. An Essential Role for Ctip in Chromosomal Translocation Formation Through an Alternative End-Joining Pathway. *Nat. Struct. Mol. Biol.*, v. 18, n. 1, p. 80–84, 2010.
- [117] SARTORI, A. A. et al. Human Ctip Promotes DNA End Resection. *Nature*, v. 450, p. 509–514, 2007.
- [118] QUENNET, V. et al. Ctip and Mrn Promote Non-Homologous End-Joining of Etoposide-Induced DNA Double-Strand Breaks in G1. *Nuc. Acids Res.*, v. 39, p. 2144–2152, 2011.
- [119] DIBIASE, S. J. et al. DNA-Dependent Protein Kinase Stimulates an Independently Active, Nonhomologous, End-Joining Apparatus. *Cancer Res.*, v. 60, p. 1245–1253, 2000.
- [120] SAVAGE, J. R. The Transmission of Fish-Painted Patterns Derived from Complex Chromosome Exchanges. *Mutat. Res.*, v. 347, n. 2, p. 87–95, July 1995.

- [121] BOEI, J. J.; NATARAJAN, A. T. Combined Use of Chromosome Painting and Telomere Detection to Analyse Radiation-Induced Chromosomal Aberrations in Mouse Splenocytes. *Int. J. Rad. Biol.*, v. 73, n. 2, p. 125–133, Feb 1998.
- [122] BOVERI, T. *Zur Frage Der Entstehung Maligner Tumoren*. Germany: Jena, Fischer, 1914.
- [123] RABBITS, T. H. Chromosomal Translocations in Human Cancer. *Nature*, v. 371, n. 6502, p. 143–149, Nov 1994.
- [124] HAGMAR, L. et al. Chromosomal Aberrations in Lymphocytes Predict Human Cancer: a Report from the European Study Group on Cytogenetic Biomarkers and Health (Esch).. *Cancer Res.*, v. 58, p. 4117–4121, Sept. 1998.
- [125] CURIE, M. Sur L'Étude Des Courbes De Probabilité Relatives À L'Action Des Rayons X Sur Les Bacilles. *Comptes Rendus l'Académie des Sci.*, v. 188, 1929.
- [126] CROWTHER, J. Some Considerations Relative to the Action of X-rays on Tissue Cells. *Royal Society B*, v. 96, Apr 1924.
- [127] PUCK, T. T.; MARCUS, P. I. Action of X-rays on Mammalian Cells. *J. Exper. Med.*, v. 103, n. 5, p. 653–666, May 1956.
- [128] PUCK, T. T. et al. Action of X-rays on Mammalian Cells. *J. Exper. Med.*, v. 106, n. 4, p. 485–500, Oct 1957.
- [129] BARENDSSEN, G. W. et al. Effects of Different Ionizing Radiations on Human Cells in Tissue Culture: II. Biological Experiments. *Radiat. Res.*, v. 13, n. 6, p. 841–849, Dec 1960.
- [130] SINCLAIR, W. *The Shape of Radiation Survival Curves of Mammalian Cells Cultured in Vitro*. Vienna, Austria, 1966. (Biophysical Aspects of Radiation Quality: 58).
- [131] STEEL, G. G. *Basic Clinical Radiobiology*. 3rd. ed. London: Arnold Publishers, 1993.
- [132] MARPLES, B.; COLLIS, S. J. Low-Dose Hyper-Radiosensitivity: Past, Present, and Future. *Int. J. Radiat. Oncol. Biol. Phys.*, v. 70, n. 5, p. 1310–1318, 2008.
- [133] KELLERER, A.; ROSSI, H. A Generalized Formulation of Dual Radiation Action a Generalized Formulation of Dual Radiation Action. *Radiat. Res.*, v. 75, p. 471–488, 1978.
- [134] CHADWICK, K. H.; LEENHOUTS, H. P. A Molecular Theory of Cell Survival. *Phys. Med. Biol.*, v. 18, n. 1, p. 78–87, 1973.
- [135] TOBIAS, C. A. The Repair-Misrepair Model in Radiobiology: Comparison to Other Models. *Radiat. Res. Supp.*, v. 8, p. S77–S95, Nov 1985.
- [136] CURTIS, S. B. Lethal and Potentially Lethal Lesions Induced by Radiation –A Unified Repair Model. *Radiat. Res.*, v. 106, n. 2, p. 252–270, May 1986.
- [137] GOODHEAD, D. T. Saturable Repair Models of Radiation Action in Mammalian Cells. *Radiat Res. Supp.*, v. 8, p. S58–S67, Nov 1985.

- [138] LOEFFLER, J. S.; DURANTE, M. Charged Particle Therapy-Optimization, Challenges and Future Directions. *Nature Rev. Clin. Oncol.*, v. 10, p. 411–424, May 2013.
- [139] SHARMA, R. A. et al. Clinical Development of New Drug-Radiotherapy Combinations. *Nature Rev. Clin. Oncol.*, v. 13, p. 627–642, Jun 2016.
- [140] BARKER, H. E. et al. The Tumour Microenvironment After Radiotherapy: Mechanisms of Resistance and Recurrence. *Nature Rev. Cancer*, v. 15, p. 409–425, Jun 2015.
- [141] NIKJOO, H. et al. Perspectives in Radiation Biophysics: from Radiation Track Structure Simulation to Mechanistic Models of DNA Damage and Repair. *Radiat. Phys. Chem.*, v. 128, p. 3–10, Jun 2016.
- [142] METZGER, L.; ILIAKIS, G. Kinetics of DNA Double-Strand Break Repair Throughout the Cell Cycle As Assayed by Pulsed Field Gel Electrophoresis in Cho Cells. *Int. J. Radiat. Biol.*, v. 59, p. 1325–1339, 1991.
- [143] FOWLER, J. F. Is Repair of DNA Strand Break Damage from Ionizing Radiation Second-Order Rather Than First-Order? a Simpler Explanation of Apparently Multiexponential Repair. *Radiat. Res.*, v. 152, n. 2, p. 124–136, Aug 1999.
- [144] CUCINOTTA, F. A. et al. Biochemical Kinetics Model of Dsb Repair and Induction of γ -H2Ax Foci by Non-Homologous End Joining. *Radiat. Res.*, v. 169, n. 2, p. 214–222, Feb 2008.
- [145] TALEEI, R.; NIKJOO, H. The Non-Homologous End-Joining (Nhej) Pathway for the Repair of DNA Double-Strand Breaks: I. a Mathematical Model. *Radiat. Res.*, v. 179, n. 5, p. 530–539, April 2013.
- [146] FRIEDLAND, W.; KUNDRÁT, P. Chromosome Aberration Model Combining Radiation Tracks, Chromatin Structure, Dsb Repair and Chromatin Mobility. *Radiat. Prot. Dosim.*, v. 166, p. 71–74, 2015.
- [147] FORAY, N. et al. The Repair Rate of Radiation-Induced DNA Damage: a Stochastic Interpretation Based on the gamma Function. *J. Theor. Biol.*, v. 236, p. 448–458, Oct 2005.
- [148] WOODS, M.; BARNES, C. Mechanistic Modelling and Bayesian Inference Elucidates the Variable Dynamics of Double-Strand Break Repair. *bioRxiv*, p. 026070, 2016.
- [149] BODGI, L. et al. Mathematical Models of Radiation Action on Living Cells: from the Target Theory to the Modern Approaches. a Historical and Critical Review. *J. Theor. Biol.*, v. 394, p. 93–101, April 2016.
- [150] NIKJOO, H. et al. Can Monte Carlo Track Structure Codes Reveal Reaction Mechanism in DNA Damage and Improve Radiation Therapy?. *Rad. Phys. and Chem.*, v. 77, p. 1270–1279, Oct 2008.

- [151] ZAIDER, M.; BRENNER, D. J.; WILSON, W. E. The Applications of Track Calculations to Radiobiology I. Monte Carlo Simulation of Proton Tracks. *Radiat. Res.*, v. 95, n. 2, p. 231–247, Aug 1983.
- [152] NIKJOO, H. et al. Energy Deposition in Small Cylindrical Targets by Ultrasoft X-rays. *Phys. Med. Biol.*, v. 34, n. 6, p. 691–705, Jul 1989.
- [153] NIKJOO, H. et al. Computational Modelling of Low-Energy Electron-Induced DNA Damage by Early Physical and Chemical Events. *Int. J. Radiat. Biol.*, v. 71, n. 5, p. 467–483, May 1997.
- [154] NIKJOO, H. et al. Computational Approach for Determining the Spectrum of DNA Damage Induced by Ionizing Radiation. *Radiat. Res.*, v. 156, n. 5-2, p. 577–583, Nov. 2001.
- [155] GOODHEAD, D. T. Initial Events in the Cellular Effects of Ionizing Radiations: Clustered Damage in DNA. *Int. J. Radiat. Biol.*, v. 65, p. 7–17, 1994.
- [156] KAWRAKOW, I.; ROGERS, D. *The Egsnrc Code System: Monte Carlo Simulation of Electron and Photon Transport*. Ottawa, Canada, 2000.
- [157] BATTISTONI, G. et al. The Fluka Code: an Accurate Simulation Tool for Particle Therapy. *Front. Oncol.*, v. 6, p. 116, 2016.
- [158] AGOSTINELLI, S. et al. Geant4-A Simulation Toolkit. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, v. 506, n. 3, p. 250–303, 2003.
- [159] NIKJOO, H. et al. Radiation Track, DNA Damage and Response - a Review. *Rep. Prog. Phys.*, v. 79, p. 116601, Nov. 2016.
- [160] GOORLEY, T. et al. Initial Mcnp6 Release Overview. *Nuclear Tech.*, v. 180, p. 298–315, 2012.
- [161] FRIEDLAND, W. et al. Track Structures, DNA Targets and Radiation Effects in the Biophysical Monte Carlo Simulation Code Partrac. *Mutat. Res.*, v. 711, n. 1-2, p. 28–40, Jun 2011.
- [162] KYRIAKOU, I. et al. Microdosimetry of Electrons in Liquid Water Using the Low-Energy Models of Geant4. *J. Applied Phys.*, v. 122, p. 024303, 2017.
- [163] LAZARAKIS, P. et al. Investigation of Track Structure and Condensed History Physics Models for Applications in Radiation Dosimetry on a Micro and Nano Scale in Geant4. *Biomed. Phys. Eng. Express*, v. 4, n. 2, p. 024001, Feb 2018.
- [164] INCERTI, S. et al. The Geant4-DNA Project. *Int. J. Modeling, Simulation, and Scientific Computing*, v. 1, n. 2, p. 157–178, 2010.

- [165] BERNAL, M. A. et al. Track Structure Modeling in Liquid Water: a Review of the Geant4-DNA Very Low Energy Extension of the Geant4 Monte Carlo Simulation Toolkit. *Physica Medica: EJMP*, v. 31, p. 861–874, Jul 2015.
- [166] PERL, J. et al. Topas: an Innovative Proton Monte Carlo Platform for Research and Clinical Applications. *Medical Phys.*, v. 39, p. 6818–6837, Oct 2012.
- [167] MCNAMARA, A. et al. Validation of the Radiobiology Toolkit Topas-Nbio in Simple DNA Geometries. *Physica Medica: EJMP*, v. 33, p. 207–215, Jan 2017.
- [168] BUG, M. U. et al. An Electron-Impact Cross Section Data Set ($10\text{ eV} - 1\text{ keV}$) of DNA Constituents Based on Consistent Experimental Data: a Requisite for Monte Carlo Simulations. *Radiat. Phys. and Chem.*, v. 130, p. 459–479, Jan 2017.
- [169] FRANCIS, Z. et al. Calculation of Lineal Energies for Water and DNA Bases Using the Rudd Model Cross Sections Integrated Within the Geant4-DNA Processes. *J. Applied Phys.*, v. 122, p. 014701, 2017.
- [170] BEGUSOVA, M. et al. Radiolysis of Nucleosome Core DNA: a Modelling Approach. *Int. J. Radiat. Biol.*, v. 76, n. 8, p. 1063–1073, Aug 2000.
- [171] KARAMITROS, M. et al. Diffusion-Controlled Reactions Modeling in Geant4-DNA. *J. Comput. Phys.*, v. 274, p. 841–882, Oct 2014.
- [172] RAMOS-MÉNDEZ, J. et al. Monte Carlo Simulation of Chemistry Following Radiolysis with Topas-Nbio. *Phys. Med. Biol.*, v. 63, n. 10, p. 105014, May 2018.
- [173] BOSCOLO, D. et al. Trax-Chem: a Pre-Chemical and Chemical Stage Extension of the Particle Track Structure Code Trax in Water Targets. *Chem. Phys. Letters*, v. 698, p. 11–18, Apr 2018.
- [174] MCNAMARA, A. et al. Geometrical Structures for Radiation Biology Research As Implemented in the Topas-Nbio Toolkit. *Phys. in Med. and Biol.*, v. 63, n. 17, p. 175018, Sep 2018.
- [175] HENTHORN, N. T. et al. Nanodosimetric Simulation of Direct Ion-Induced DNA Damage Using Different Chromatin Geometry Models. *Radiat. Res.*, v. 188, n. 6, p. 770–783, Aug 2017.
- [176] DELAGE, E. et al. Pdb4DNA: Implementation of DNA Geometry from the Protein Data Bank (Pdb) Description for Geant4-DNA Monte-Carlo Simulations. *Comp. Phys. Communications*, v. 192, p. 282–288, Jul 2015.
- [177] SCHUEMANN, J. et al. A New Standard DNA Damage (Sdd) Data Format. *Radiat. Res.*, v. 191, n. 1, p. 76–92, Jan 2019.
- [178] HENTHORN, N. T. et al. In Silico Non-Homologous End Joining Following Ion Induced DNA Double Strand Breaks Predicts That Repair Fidelity Depends on Break Density. *Scientific Reports*, v. 8, p. 2654, Feb 2018.

- [179] CHEN, A. M. et al. Computer Simulation of Data on Chromosome Aberrations Produced by X Rays Or alpha Particles and Detected by Fluorescence in Situ Hybridization. *Radiat. Res.*, v. 148, n. 5, p. 93–101, Nov 1997.
- [180] SACHS, R. et al. Random Breakage and Reunion Chromosome Aberration Formation Model; an Interaction-Distance Version Based on Chromatin Geometry. *Int. J. Radiat. Biol.*, v. 76, p. 1579–1588, 2000.
- [181] PONOMAREV, A.; GEORGE, K.; CUCINOTTA, F. Computational Model of Chromosome Aberration Yield Induced by High- and Low-LET Radiation Exposures. *Radiat. Res.*, v. 177, p. 727–737, 2012.
- [182] BALLARINI, F.; CARANTE, M. Chromosome Aberrations and Cell Death by Ionizing Radiation: Evolution of a Biophysical Model. *Radiat. Phys. Chem.*, v. 128, n. C, p. 18–25, 2016.
- [183] CARLSON, D. et al. Combined Use of Monte Carlo DNA Damage Simulations and Deterministic Repair Models to Examine Putative Mechanisms of Cell Killing. *Radiat. Res.*, v. 169, n. 4, p. 447–479, Apr 2008.
- [184] ELSÄSSER, T.; KRÄMER, M.; SCHOLZ, M. Accuracy of the Local Effect Model for the Prediction of Biologic Effects of Carbon Ion Beams in Vitro and in Vivo. *Int. J. Radiat. Oncol.*, v. 71, p. 866–872, Jul 2008.
- [185] INANIWA, T. et al. Treatment Planning for a Scanned Carbon Beam with a Modified Microdosimetric Kinetic Model. *Phys. Med. Biol.*, v. 55, n. 22, p. 6721–6737, Oct 2010.
- [186] FRIEDRICH, T. et al. Calculation of the Biological Effects of Ion Beams Based on the Microscopic Spatial Damage Distribution Pattern. *Int. J. Radiat. Biol.*, v. 88, p. 103–107, 2012.
- [187] HAWKINS, R.; INANIWA, T. A Microdosimetric-Kinetic Model for Cell Killing by Protracted Continuous Irradiation Ii: Brachytherapy and Biologic Effective Dose. *Radiat. Res.*, v. 182, n. 1, p. 72–82, Jun 2014.
- [188] SATO, T. et al. Microdosimetric Modeling of Biological Effectiveness for Boron Neutron Capture Therapy Considering Intra- and Intercellular Heterogeneity in ^{10}B Distribution. *Scientific Reports*, v. 8, p. 988, Jan 2018.
- [189] MCMAHON, S. J. et al. Biological Consequences of Nanoscale Energy Deposition Near Irradiated Heavy Atom Nanoparticles. *Scientific Reports*, v. 1, p. 3319–3330, 2011.
- [190] KRÄMER, M.; SCHOLZ, M. Treatment Planning for Heavy-Ion Radiotherapy: Calculation and Optimization of Biologically Effective Dose. *Phys. Med. Biol.*, v. 45, n. 11, p. 3319–3330, 2000.

- [191] SCHIPLER, A.; ILIAKIS, G. DNA Double-Strand-Break Complexity Levels and Their Possible Contributions to the Probability for Error-Prone Processing and Repair Pathway Choice. *Nucleic Acids Res.*, v. 41, p. 7589–7605, 2013.
- [192] BALLARINI, F. et al. Chromosome Aberrations Induced by Light Ions: Monte Carlo Simulations Based on a Mechanistic Model. *Int. J. Radiat. Biol.*, v. 75, n. 1, p. 35–46, Jan 1999.
- [193] OTTOLENGHI, A. et al. The Quality of DNA Double-Strand Breaks: a Monte Carlo Simulation of the End-Structure of Strand Breaks Produced by Protons and alpha Particles. *Radiat. Environ. Biophys.*, v. 34, n. 4, p. 239–244, 1995.
- [194] OTTOLENGHI, A.; BALLARINI, F.; BIAGGI, M. Modelling Chromosomal Aberration Induction by Ionising Radiation: the Influence of Interphase Chromosome Architecture. *Adv. Space Res.*, v. 27, p. 369–382, 2001.
- [195] BALLARINI, F.; BIAGGI, M.; OTTOLENGHI, A. Nuclear Architecture and Radiation-Induced Chromosome Aberrations: Models and Simulations. *Rad. Prot. Dosim.*, v. 99, p. 175–182, 2002.
- [196] BALLARINI, F.; OTTOLENGHI, A. Chromosome Aberrations As Biomarkers of Radiation Exposure: Modelling Basic Mechanisms. *Adv. Space Res.*, v. 31, p. 1557–1568, 2003.
- [197] EDWARDS, A. A. et al. Chromosome Aberrations Induced in Human Lymphocytes by 8.7 MeV Protons and 23.5 MeV Helium-3 Ions. *Int. J. Radiat. Biol.*, v. 50, n. 1, p. 137–145, 1986.
- [198] PURROTT, R. J. et al. The Induction of Chromosome Aberrations in Human Lymphocytes by in Vitro Irradiation with α -Particles from Plutonium-239. *Int. J. Radiat. Biol.*, v. 38, n. 3, p. 277–284, 1980.
- [199] BALLARINI, F.; OTTOLENGHI, A. A Model of Chromosome Aberration Induction: Applications to Space Research. *Rad. Res.*, v. 164, p. 567–570, 2005.
- [200] BALLARINI, F. et al. Radiation Risk Estimation: Modelling Approaches for "Targeted" and "Non-Targeted" Effects. *Adv. Space Res.*, v. 40, p. 1392–1400, 2007.
- [201] BALLARINI, F. et al. Heavy-Ion Effects: from Track Structure to DNA and Chromosome Damage. *New J. Phys.*, v. 10, p. 075008, 2008.
- [202] GEORGE, K. et al. Biological Effectiveness of Accelerated Particles for the Induction of Chromosome Damage Measured in Metaphase and Interphase Human Lymphocytes. *Radiat. Res.*, v. 160, p. 425–435, 2003.
- [203] BALLARINI, F. From DNA Radiation Damage to Cell Death: Theoretical Approaches. *J. Nucleic Acids*, v. 2010, 2010.

- [204] NEUMAIER, T. et al. Evidence for Formation of DNA Repair Centres and Dose Response Nonlinearity in Human Cells. *PNAS*, v. 109, p. 443–448, 2012.
- [205] BRIDGER, J. M. et al. Identification of an Interchromosomal Compartment by Polymerisation of Nuclear-Targeted Vimentin. *J. Cell. Sci.*, v. 111, p. 1241–1253, 1998.
- [206] WU, H.; GEORGE, K.; YANG, T. Estimate of the Frequency of True Incomplete Exchanges in Human Lymphocytes Exposed to 1 GeV/U Fe Ions in Vitro. *Int. J. Radiat. Biol.*, v. 75(5), p. 593–599, 1999.
- [207] WU, H. et al. Analysis of Unrejoined Chromosomal Breakage in Human Fibroblast Cells Exposed to Low- and High-LET Radiation. *J. Radiat. Res.*, v. 53, p. S181–S185, 2002.
- [208] BOEI, J. et al. Impact of Radiation Quality on the Spectrum of Induced Chromosome Exchange Aberrations. *Int. J. Radiat. Biol.*, v. 77, n. 8, p. 847–857, 2001.
- [209] CORNFORTH, M. N.; BEDFORD, J. S. On the Nature of a Defect in Cells from Individuals with Ataxia-Telangiectasia. *Science*, v. 227, n. 4694, p. 1589–1591, 1985.
- [210] TEAM gfortran. *Using Gnu Fortran*. pre-release. Boston, MA 02110-1301, USA: Free Software Foundation, 2016.
- [211] CHATTERJEE, A.; SCHAEFER, H. Microdosimetric Structure of Heavy Ion Tracks in Tissue. *Radiat. Environ. Biophys.*, v. 13, p. 215 – 227, 1976.
- [212] KIEFER, J.; STRAATEN, H. A Model of Ion Track Structure Based on Classical Collision Dynamics. *Phys. Med. Biol.*, v. 31, p. 1201 – 1209, 1986.
- [213] ZIRBEL, R. et al. Evidence for a Nuclear Compartment of Transcription and Splicing Located at Chromosome Domain Boundaries. *Chromosome Res.*, v. 1, p. 93 – 106, 1993.
- [214] CARANTE, M. et al. Modeling Radiation-Induced Cell Death: Role of Different Levels of DNA Damage Clustering. *Radiat. Environ. Biophys.*, v. 54, p. 305–316, 2015.
- [215] MUHLMANN-DIAZ, M.; BEDFORD, J. Comparison of gamma-ray-Induced Chromosome Ring and Inversion Frequencies. *Radiat. Res.*, v. 143, p. 175–180, 1995.
- [216] BENKHALED, L. et al. Analysis of gamma-rays Induced Chromosome Aberrations: a Fingerprint Evaluation with a Combination of Pan-Centromeric and Pan-Telomeric Probes. *Int. J. Radiat. Biol.*, v. 82, p. 869–875, 2006.
- [217] SACHS, R. et al. Intra-Arm and Interarm Chromosome Intrachanges: Tools for Probing the Geometry and Dynamics of Chromatin. *Radiat. Res.*, v. 148, p. 330–340, 1997.
- [218] HOLLEY, W. et al. A Model for Interphase Chromosomes and Evaluation of Radiation-Induced Aberrations. *Radiat. Res.*, v. 158, p. 568–580, 2002.
- [219] MESTRES, M. et al. Analysis of α -Particle Induced Chromosome Aberrations in Human Lymphocytes, Using Pan-Centromeric and Pan-Telomeric Probes. *Int. J. Radiat. Biol.*, v. 80, p. 737–744, 2004.

- [220] BERKUM, N. L. van et al. Hi-C: a Method to Study the Three-Dimensional Architecture of Genomes. *J. Vis. Exp.*, v. 39, p. 1869, 2010.
- [221] MIRNY, L. A. The Fractal Globule As a Model of Chromatin Architecture in the Cell. *Chrom. Res.*, v. 19, p. 37–51, 2011.
- [222] BARBIERI, M. et al. A Model of the Large-Scale Organization of Chromatin. *Biochem. Soc. Trans.*, v. 41, n. 2, p. 508–512, 2013.
- [223] SANBORN, A. L. et al. Chromatin Extrusion Explains Key Features of Loop and Domain Formation in Wild-Type and Engineered Genomes. *PNAS*, v. 112, n. 47, p. E6456–E6465, 2015.
- [224] ZHANG, Z. et al. 3D Chromosome Modeling with Semi-Definite Programming and Hi-C Data. *J. Comput. Biol.*, v. 20, n. 11, p. 831–846, 2013.
- [225] VAROQUAUX, N. et al. A Statistical Approach for Inferring the 3D Structure of the Genome. *Bioinformatics*, v. 30, p. i26 – 33, 2014.
- [226] LESNE, A. et al. 3D Genome Reconstruction from Chromosomal Contacts. *Nat. Methods*, v. 11, p. 1141–1143, 2014.
- [227] MIELE, A.; BYSTRICKY, K.; DEKKER, J. Yeast Silent Mating Type Loci Form Heterochromatic Clusters Through Silencer Protein-Dependent Long-Range Interactions. *PLoS Genet.*, v. 5, n. 5, p. e1000478, 2009.
- [228] DIXON, J. R. et al. Topological Domains in Mammalian Genomes Identified by Analysis of Chromatin Interactions. *Nature*, v. 485, p. 376–380, 2012.
- [229] NORA, E. P. et al. Spatial Partitioning of the Regulatory Landscape of the X-Inactivation Centre. *Nature*, v. 485, n. 7398, p. 381–385, 2012.
- [230] ROUSSEAU, M. et al. Three-Dimensional Modeling of Chromatin Structure from Interaction Frequency Data Using Markov Chain Monte Carlo Sampling. *BMC Bioinformatics*, v. 12, p. 414, 2011.
- [231] CREMER, T.; CREMER, C. Rise, Fall and Resurrection of Chromosome Territories: a Historical Perspective. Part I. the Rise of Chromosome Territories. *Eur. J. Histochem.*, v. 50, n. 3, p. 161–176, 2006.
- [232] RAO, S. S. et al. A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell*, v. 159, n. 7, p. 1668–1680, 2014.
- [233] SINKHORN, R.; KOPP, P. Concerning Nonnegative Matrices and Doubly Stochastic Matrices. *Pacific J. Math.*, v. 21, n. 2, p. 343–348, 1967.
- [234] KNIGHT, P. A.; RUIZ, D. A Fast Algorithm for Matrix Balancing. *IMA J. Num. Anal.*, v. 33, p. 1029–1047, 1967.

- [235] TESTA, A. et al. Analysis of Radiation-Induced Chromosomal Aberrations on a Cell-By-Cell Basis After alpha-Particle Microbeam Irradiation: Experimental Data and Simulations. *Rad. Res.*, v. 189, n. 6, p. 597–604, June 2018.
- [236] RYDBERG, B. et al. Spatial Distribution and Yield of DNA Double-Strand Breaks Induced by 3-7 MeV Helium Ions in Human Fibroblasts. *Radiat. Res.*, v. 158, p. 32–42, 2002.
- [237] WU, H. et al. Centric Rings: Acentric Rings and Excess Acentric Fragments Based on a Random-Walk Interphase Chromosome Model. *Int. J. Radiat. Biol.*, v. 71, p. 487–496, 1997.
- [238] JAKOB, B.; SPLINTER, J.; TAUCHER-SCHOLZ, G. Positional Stability of Damaged Chromatin Domains Along Radiation Tracks in Mammalian Cells. *Radiat. Res.*, v. 171, p. 405–418, 2009.
- [239] FRIEDLAND, W.; KUNDRÁT, P. Track Structure Based Modelling of Chromosome Aberrations After Photon and alpha-Particle Irradiation. *Mutat. Res. - Genet. Toxicol. Environ. Mutagen.*, v. 756, p. 213–223, 2013.
- [240] COSTES, S. et al. Image-Based Modeling Reveals Dynamic Redistribution of DNA Damage into Nuclear Sub-Domains. *PLoS Comput. Biol.*, v. 3, 2007.
- [241] DENG, W. et al. A Comparative Study on Potential Cytogenetic Fingerprints for Radiation LET in Human Lymphocytes. *Int. J. Radiat. Biol.*, v. 76, p. 1589–1598, 2000.
- [242] SASAKI, M. et al. Chromosome Aberration Frequency and Radiation Dose to Lymphocytes by alpha-Particles from Internal Deposit of Thorotrast. *Radiat. Environ. Biophys.*, v. 26, p. 227–236, 1987.
- [243] TAWN, E.; HALL, J.; SCHOFIELD, G. Chromosome Studies in Plutonium Workers. *Int. J. Radiat. Biol.*, v. 47, p. 599–610, 1985.
- [244] RAZIN, S. et al. Chromatin Domains and Regulation of Transcription. *J. Mol. Biol.*, v. 369, p. 597–607, 2007.
- [245] NASONOVA, E. et al. Induction of Chromosomal Damage in Cho-K1 Cells and Their Repair-Deficient Mutant Xrs5 by X-ray and Particle Irradiation. *Adv. Space Res.*, v. 22, p. 569–578, 1998.
- [246] CARPENTER, S. et al. Radiobiology of Ultrasoft X Rays: Iv. Flat and Round-Shaped Hamster Cells (Cho-10B, Hs-23). *Radiat. Res.*, v. 119, n. 3, p. 523–533, Sep 1989.
- [247] NAGASAWA, H.; LITTLE, J. B. Bystander Effect for Chromosomal Aberrations Induced in Wild-Type and Repair Deficient Cho Cells by Low Fluences of alpha Particles. *Mutat. Res.*, v. 508, p. 121–129, Oct 2002.
- [248] VILLAGRASA, C. et al. Geant4-DNA Simulation of DNA Damage Caused by Direct and Indirect Radiation Effects and Comparison with Biological Data. *EPJ web of conferences*, v. 153, n. 3, p. 04019, Oct 3-6 2017.

-
- [249] FOLKARD, M. et al. Inactivation of V79 Cells by Low-Energy Protons, Deuterons and Helium-3 Ions. *Int. J. Rad. Biol.*, v. 66, n. 6, p. 729–738, 1996.
- [250] CUTTONE, G. et al. Catania Protontherapy Facility: the State of Art of Clinical and Dosimetric Experience. *Eur. Phys. J. Plus*, v. 126, n. 65, 2011.
- [251] MAIRANI, A. et al. Biologically Optimized Helium Ion Plans: Calculation Approach and Its *in vitro* Validation. *Phys. Med. Biol.*, v. 61, p. 4283–4299, 2016.
- [252] BERNAL, M. A. et al. An Atomistic Geometrical Model of the B-DNA Configuration for DNA-Radiation Interaction Simulations. *Comp. Phys. Comm.*, v. 184, p. 2840–2847, December 2013.
- [253] TELLO, J. J. et al. Numerical Insight into the Dual Radiation Action Theory. *Physica Medica*, v. 43, p. 120–126, 2017.