THE UNIVERSITY of ADELAIDE

FACULTY OF SCIENCES
SCHOOL OF PHYSICAL SCIENCES

A New Model for Dose-Response Relations in Hadron Therapy: A Statistical Analysis of Hadron Therapy Data

MELISSA ANNE MCINTYRE

Supervisors:
Dr. Ayşe KIZILERSÜ
Prof. Anthony W. THOMAS

A thesis presented for the degree of Master of Philosophy

May 2021

Declaration

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name, in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint-award of this degree.

I give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

I acknowledge the support I have received for my research through the provision of an Australian Government Research Training Program Scholarship.

Signature

Date

Abstract

In recent years, hadron therapy has become an increasingly popular cancer treatment alternative to conventional photon-based radiation. The distinct advantage of using proton or heavy ion radiation over other treatment modalities (x-rays) is the depositing of the desired dose directly onto a targeted tumour. This treatment avoids delivering lethal doses of radiation to the surrounding healthy and potentially radiation-sensitive tissues. The tissue sparing effect of hadron therapy significantly improves the quality of life and minimises long-term health side-effects in cancer patients from excessive radiation exposure.

Understanding the response of a eukaryotic cell to ionising radiation is of vital importance in the field. Many models have been developed to explain the response of a cell to ionising radiation, all of which are based on the Poisson count process. The most widely used model in the literature, the Linear-Quadratic model, is no exception. However, despite its wide use, the Linear-Quadratic model presents serious problems under statistical analysis and explaining the mid to high linear-energy-transfer (LET) region of experimental data.

In this study, we first make use of rigorous statistical analyses of experimental world hadron therapy dose-response data to test the validity range of the Linear-Quadratic model under different radiation exposure and biological conditions. Our statistical analysis showed that it has a limited range of applicability and is restricted to the low to mid-LET region. Moreover, we demonstrated that it exhibits discrepancies under the considered regression analysis.

To understand and explain these discrepancies, we make use of the TOPAS and Geant4 software toolkits to carry out a series of numerical simulations to study the dose-response relations by radiating V79 Chinese Hamster cells with a proton beam for a range of LET.

Our analysis of the simulated data shows that the distribution of lethal damages per cell is overdispersed in the mid to high-LET range, violating the equidispersion condition of the Poisson process. However as the LET decreases, an overdispersed distribution of lethal damages approach to an equidispersed distribution, satisfying

the Poisson condition.

To explain the experimental and simulated data better, we proposed a new stochastic model based on a fractional Poisson count process which converges to the Poisson count process in the low-LET region. We rigorously tested our newly proposed model against the experimental and our simulated dose-response data and found that they are in excellent agreement.

We showed that the distribution of lethal damages can be explained by a fractional Poisson process significantly better than the Poisson count process. The cell survival dose-response results exhibit a superior agreement with the Mittag-Leffler distribution which corresponds to zero count events of the fractional Poisson process in all LET ranges for different cell lines and radiation types. The Mittag-Leffler distribution predicts the DNA damage yield and therefore the relative biological effectiveness extremely accurately. Compared with the Linear-Quadratic model, we demonstrated that our proposed model is superior.

Acknowledgments

My Masters journey has been two of the most challenging but rewarding years of my life. I have found a field of physics that I am passionate about and I have learned so much. However, there are so many people to whom I owe my deepest gratitude for making this time enjoyable and, at times, tolerable.

First and foremost, I would like to thank my principal supervisor Dr. Ayşe Kızılersü for all of the constant support, guidance, understanding, patience and good company she has given me over the last two years. I can safely say that without her help I would not have had this opportunity. She was always happy to discuss anything from research and coursework, to careers, or even just a friendly chat. I am indebted to her for going above and beyond to ensure this thesis was completed to the highest quality. I know the months leading to the completion of this thesis were just as hard for her as they were for me and I cannot thank her sufficiently for her efforts. She has been a great mentor and friend throughout this journey and I look forward to working with her in the not so distant future.

I am extremely thankful to my secondary supervisor Professor Anthony Thomas for all of the continued support he as given me throughout my candidature. I would also not have been fortunate enough to undertake this project without his support from day one.

A special thank you to Dr. Markus Kreer for our fruitful and helpful collaboration. Also for his efforts in developing and editing this thesis. Your contributions were much appreciated and I believe your suggestions greatly improved the quality.

I wish to thank Dr. Fabien Voisin and the Phoenix HPC team for their dedicated support in helping me complete my simulations on the cluster. Without your assistance I would probably still be trying to collect my results. I cannot thank you enough for the lengths you have gone to. I would also like to thank Ramona Rogers for assisting me in my quest to run my simulations on just about every computer in the department.

Thank you to Professor Cynthia Keppel and Dr. Pawel Ambrozewicz of the JLab group for providing us with the experimental data and some of your results to

compare with.

Thank you to Dr. Michael Douglass for patiently answering my questions about Geant4 and radiation modelling/simulations. It was of great help and made the learning curve much easier to surpass. I would also like to thank Dr. Alex Santos for his support and for answering my questions about medical physics.

Thank you to Marium Khan and Intersect the assistance in storing my data. I appreciate the effort you went to in ensuring I had adequate space for my data and it was a load off my shoulders while I was collecting my results.

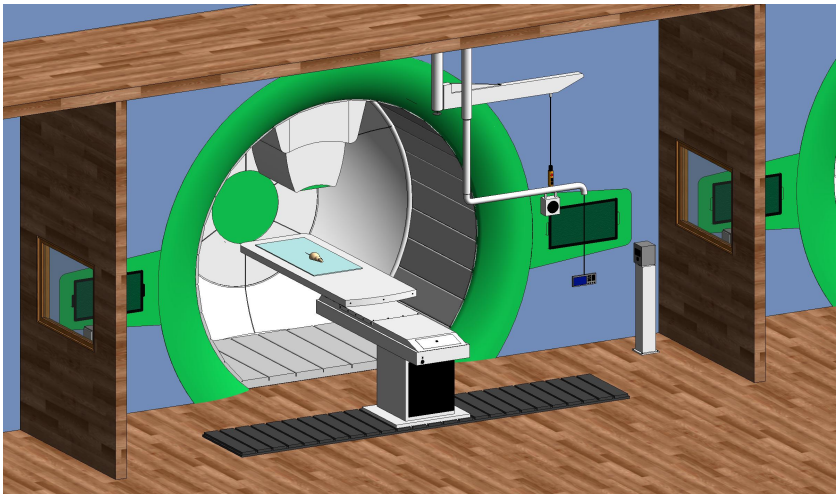
I am indebted to my parents and brother, Bryan, Maria and Ben for their endless support throughout all of my studies. You have always encouraged me to pursue my interests and you do everything in your power to ensure I am able to do so. Your patience and understanding in recent months means more than you think.

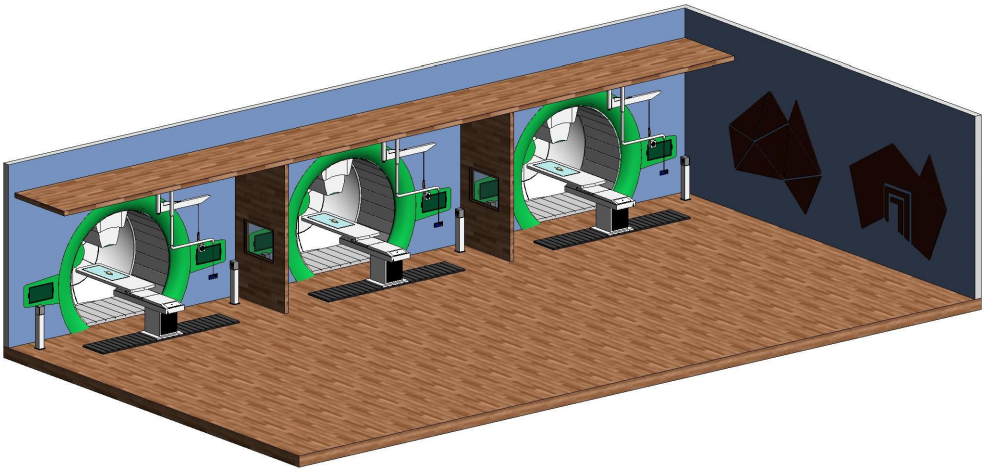
A heartfelt thank you my partner Dan who, despite being on the other side of the world, has remained extremely supportive and understanding - especially in the months leading up to the completion of this thesis. I know the last year has been extremely tough for you. I am thankful for you sticking by me and being a welcomed distraction from my work.

I am thankful to Lyn, Susan and my old co-workers in the kitchen at Estia Health Aberfoyle Park for the support and for listening to me talk about physics non-stop - with varying amounts of interest. Our shifts together were a nice distraction from my studies over the last several years and despite the ups and downs, I will look back fondly on my time working with you.

Finally, I would like to extend my thanks to all of my fellow CSSM students for making the department such a positive and supportive environment to be part of. Special thanks to the students of office 128 ("Siberia") for allowing me to run simulations on their desktop computers when they were not using them and tolerating several machines running concurrently over the last few months.

Special thanks to Adam Potter for producing the AutoCAD pictures and video. You went above and beyond and it made for a great addition to the thesis.





Contents

List of Figures

List of Tables

List of Abbreviations

BP	**B**ase **P**air
CTRW	**C**ontinuous **T**ime **R**andom **W**alk
CSV	**C**omma-**S**eparated **V**alue
CDF	**C**umulative **D**ensity **F**unction
DSKM	**D**ouble **S**tochastic Microdosimetric **K**inetic **M**odel
DNA	**D**eoxyribo**N**ucleic **A**cid
DSB	**D**ouble **S**trand **B**reak
EDF	**E**mpirical **D**istribution **F**unction
FPp	**F**ractional **P**oisson **P**rocess
GL	**G**enome **L**ength
GEANT	**GE**ometry **AN**d **T**racking
HPC	**H**igh **P**erformance **C**omputer
i.i.d	**i**ndependently **i**dentically **d**istributed
IRCU	**I**nternational **C**ommission on **R**adiation **U**nits and **M**easurements
JLab	Thomas **J**efferson National Accelerator Facility
LET	**L**inear **E**nergy **T**ransfer
LEM	**L**ocal **E**ffect **M**odel
LHC	**L**arge **H**adron **C**ollider
LPL	**L**ethal **P**otentially-**L**ethal
LQ	**L**inear-**Q**uadratic
MC	**M**onte **C**arlo

MKM	**M**icrodosimetric **K**inetic **M**odel
ML	**M**ittag-**L**effler
NLLS	**N**on-**L**inear **L**east **S**quares
NLR	**N**on-**L**inear **R**egression
NSCLC	**N**on-**S**mall **C**ell **L**ung **C**ancer
OER	**O**xygen **E**nhancement **R**atio
PIDE	**P**article **I**rradiation **D**ata **E**nsemble
PDD	**P**ercentage **D**epth **D**ose
PDF	**P**robability **D**ensity **F**unction
Pp	**P**oisson **P**rocess
RBE	**R**elative **B**iological **E**ffectiveness
RMR	**R**epair **M**is-**R**epair
RMSE	**R**oot **M**ean **S**quared **E**rror
SKM	**S**tochastic Microdosimetric **K**inetic **M**odel
SLR	**S**imple **L**inear **R**egression
SSB	**S**imple **S**trand **B**reak
SEM	**S**tandard **E**rror of the **M**ean
SSE	**S**um of **S**quared **E**rror
TLK	**T**wo **L**esion **K**inetic
TOPAS	**TO**ol for **PA**rticle **S**imulation

1

Introduction

Cancer is generally defined as a disease where abnormal cells divide at random and impact healthy tissues inside the body. It can occur in many forms and complexities, arising from numerous sources. In 2020 an average of 396 new cancer cases were diagnosed in Australia every day with 136 people losing their lives on average [1]. The survival rate of cancer patients has improved over time which can be attributed to advances in existing treatments and a wider range of alternative treatment options. Radiation treatments are among the most common to treat cancer.

The concept of using ionising radiation to treat cancer was first proposed in 1895 [2] using photons. Photons are massless and deposit energy primarily via the production of secondary electrons [3]. The advantage of this treatment modality is that the radiation targets a smaller section of the body than alternative treatments available at the time, such as chemotherapy where the whole body is affected [4]. The dose deposited by a photon is localised to its track and the small mass means that it is likely to traverse through the patient completely. This means that healthy tissues within the body experience negative effects due to radiation which can lead to further complications such as secondary cancer [2].

There have since been numerous advances in the field including fractionated radiotherapy and methods to minimise the delivery of unnecessary radiation doses to patients. In cases where the tumour is surrounded by sensitive organs or tissues (such as the brain or eye), external beam radiation can be delivered in small, low dose fractions (typically 20-30 fractions) over a long period of time [5]. This allows the healthy tissues surrounding the tumour to repair between fractions, avoiding health complications due to radiation effects. Recent research efforts have aimed to avoid these sensitive, non-cancerous tissues entirely. This led to the investigation of using ionising radiation from different particles to treat cancer. Alternatives investigated thus far include protons, heavy ions (such as helium and carbon) [6], alpha particles [7] and pions [8]. All of which have different characteristics, however the common thread is the sparing of healthy tissues during treatment. Our increasing

knowledge of radiobiology and engineering has seen a number of centres adapt these modalities to treat patients.

The use of proton radiation to treat deep-seated tumours in the body was first proposed by Robert Wilson in 1946 [9] and since the development of the first proton therapy centres in Berkeley (California, USA), Uppsala (Sweden) and Cambridge (Massachusetts, USA), it has grown in popularity to become a common alternative to photon-based radiotherapy treatments. This increase in popularity has seen numerous proton therapy centres built across Asia, North America and Europe. Australia is set to become the first country in Oceania to house such a facility in Adelaide (South Australia).

1.1 A Snapshot of Radiation Therapy

A patient's radiotherapy treatment plan is determined by the type and stage of the cancer [10]. Consideration is also given to the area in which the cancer is situated in the body, along with the tumour composition and size. The interaction processes that particles undergo when propagating through the body have been rigorously studied to improve treatments for many types of cancers. Modern radiotherapy research aims to minimise the radiation exposure to healthy tissue whilst maximising the dose absorbed by the cancer.

Radiation treatments come in many forms including external beam radiation treatment (EBRT), radiosurgery and brachytherapy [11]. Each modality has its own advantages when used to treat specific types of cancer. Brachytherapy has had great success in treating prostate cancer because of its ability to target the prostate without exposing the bladder and other surrounding radio-sensitive organs to radiation [12]. Methods for increasing the effectiveness of treatments have also been investigated, including the use of gold nanoparticles to enhance damage to the cells when applied to the tumour area [13]. Intensity modulated radiation therapy (IMRT) utilises multiple beams to irradiate the tumour from different angles reducing lethal exposure to healthy tissue without trading-off the tumour impact [14].

Since the first radiotherapy treatments over 100 years ago, mathematical modelling has been used to describe the radiation response of tissues and cells using different treatment modalities. Modern radiotherapy research has progressed from clinical trials to using Monte Carlo (MC) methods to optimise treatments and perform calculations in treatment planning. The ability to carry out MC simulations and apply mathematical models allows for a more tailored and accurate approach to a patient's treatment. MC methods use statistical sampling to simulate stochastic processes such as particle interactions and trajectories, becoming particularly useful

in the research of new radiation treatments.

Advances in the use of other modalities as an alternative to photon radiation have allowed for more selection in treating specific cancer types. In addition to photons, popular treatment alternatives include, but are not limited to, heavy ions, protons and alpha particles. The unique percentage depth dose (PDD) distributions of heavier particles allow for one region of tissue to receive a high percentage of dose while a lower percentage is delivered to surrounding tissues [2]. This is particularly important for deeply situated tumours where oncologists can avoid exposing healthy tissues to unnecessary doses of radiation which is a distinct disadvantage of photon-based therapies. This behavior is achieved through the Bragg Peak where a negligible dose is delivered at small depths before increasing substantially in the region where particles reach their range of propagation (green and yellow lines in Fig.(1.1)). In contrast, photons deposit a high dose in the entrance region with a gradual decrease (blue line in Fig.(1.1)). The healthy tissue is therefore exposed to high doses when photons are used whilst protons deposit a small, negligible dose. In the event that a single Bragg Peak cannot irradiate the entire tumour, the peak can be modulated to produce a Spread-Out Bragg Peak (SOBP, the red line in Fig.(1.1)).

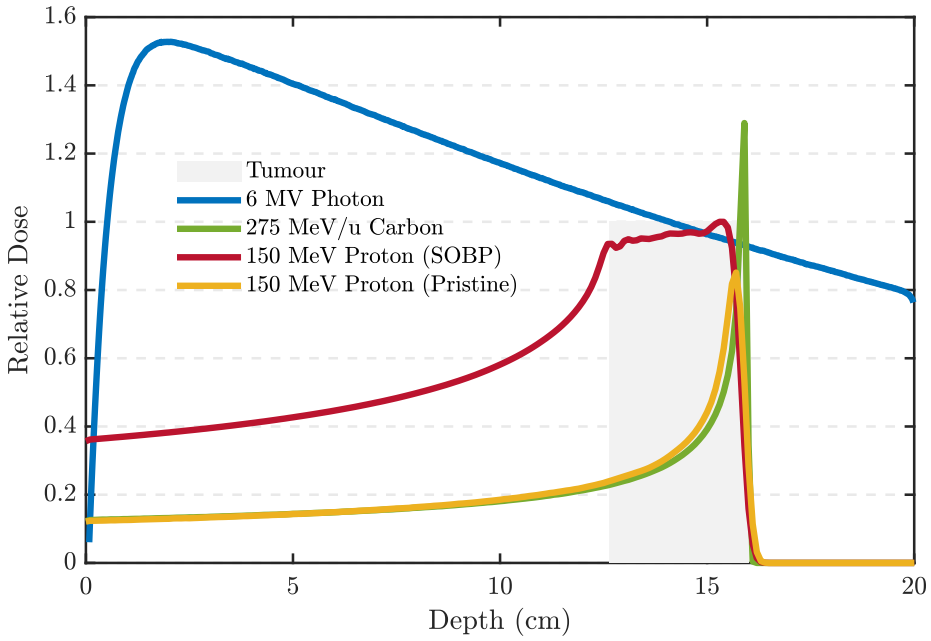


Figure 1.1: Comparison of percentage depth-dose (PDD) distributions for different types of ionising radiation in liquid water.

Whilst using proton and heavy ion radiation to treat cancer is not a recently discovered concept, it has gained popularity in recent years with a sharp increase in proton therapy centres being built in Europe, Japan and the United States. Initial research and clinical trials have shown that proton therapy has distinct advantages

over other therapies currently available to patients. However, there is a high cost involved to build, operate and maintain the equipment [15]. It has been called into question whether the benefits of proton therapy justify these high costs.

As an emerging form of cancer treatment, the amount of hadron therapy research published to-date is lesser than that of more established treatments. This makes it difficult to justify the high cost and infers the need for more research in the area. Nonetheless, the number of proton therapy centres in operation or currently under development has increased drastically in recent years and is expected to continue growing in the near future (see Fig.(1.2)).

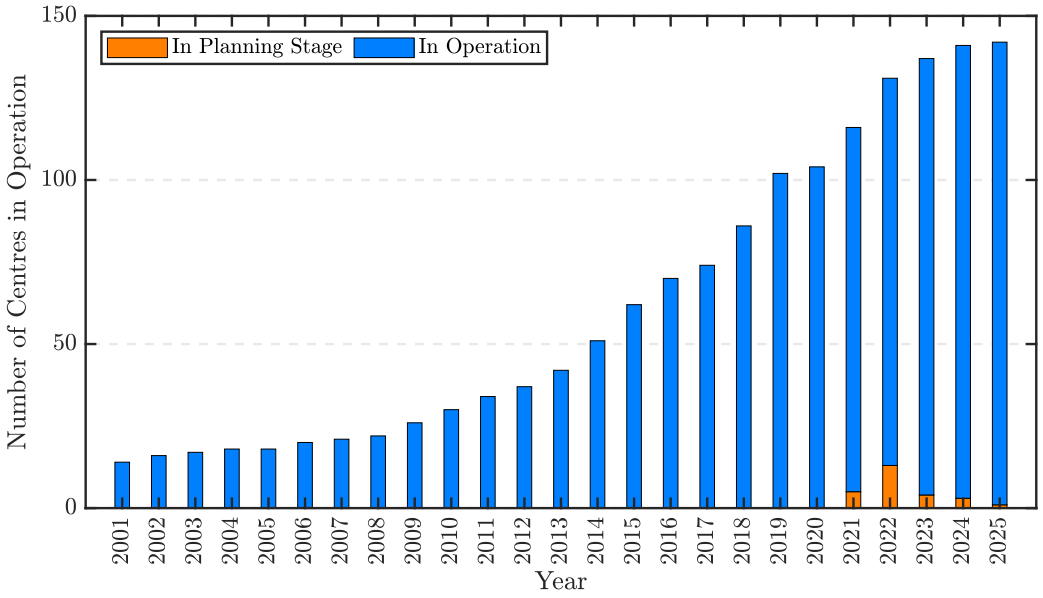


Figure 1.2: The number of hadron therapy centres in operation per year for the last 19 years and projections for the next five years as of July, 2020 (PTCOG, 2020) (*https://www.ptcog.ch/index.php/*).

The long-term advantages of hadron therapy are the success rate and overall improvement of quality of life for patients during and after treatment. For example, a study of proton therapy outcomes on non-small cell lung cancer (NSCLC) found a statistically significant improvement in the survival rate of patients treated with proton EBRT with respect to photon radiation 5 years after treatment (namely 60% survival using protons and 41% for photons) [16]. Long-term benefits of hadron therapy are still being investigated, however it is understood that the tissue-sparing effect of hadron therapy decreases health complications in patients, improving their quality of life and minimising the risk of secondary cancer formation.

1.1.1 Hadron Therapy

With the number of proton therapy centres growing worldwide (see Fig.(1.3)), more research is needed in the area to justify the high costs to build and maintain the required equipment. Accurate modelling and effective planning is of the utmost importance in successful cancer treatment. Whilst the behavior of protons and heavy ions in body tissue has been well documented, models describing the processes undergone by a cell after exposure are at times vague and less rigorous than one would prefer [17–20]. Many assumptions are made in these models about the complex system of a cell and how it responds to ionising radiation.

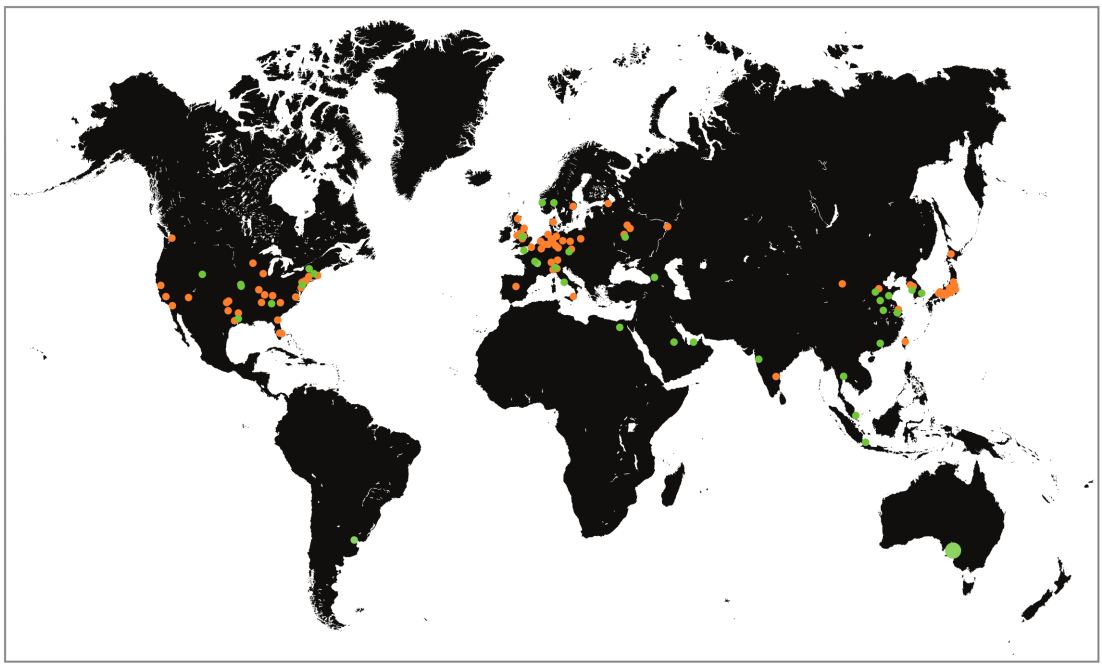


Figure 1.3: A map of proton and carbon ion therapy centres currently in operation (orange) and under construction/in the planning stages (green) (*https://www.ptcog.ch/index.php/*).

An operational hadron therapy centre requires several pieces of equipment to treat patients (see Fig.(1.4)). The primary piece of equipment is a particle accelerator (commonly a cyclotron or a synchrotron) where the beam is generated (labelled "Accelerator" in Fig.(1.4)). To produce a proton beam, we use hydrogen as the starting product and the protons are separated from the electrons using an electric field [21]. The protons are accelerated to an energy sufficient to penetrate the medium, reaching the distal edge of the tumour. After the beam is generated, it travels through a beam line (see Fig.(1.4)) and is modulated to match the morphology of the tumour. The beam is delivered to the patients via a beam line. It is in this region where the beam is modulated appropriately to match the specifications of the tumour so as to avoid unnecessary dose delivery to healthy tissues in the patient (see diagram (a) Fig.(1.6)). Finally, the beam can be delivered to the patient

using a gantry (see Fig.(1.5)) which rotates 360° around the patient or is fixed (as per Fig.(1.4)) respectively.

A more modern approach of beam delivery is to use pencil beam scanning which generates a single Bragg Peak to uniformly cover the tumour volume in a scanning motion as in diagram (b) in Fig.(1.6). This is an extremely precise method but can result in uneven irradiation of the tumour if the scanning layers are far apart [2].

The tissue sparing effect of hadron therapy is owed to the large mass of protons or heavy ions with respect to photons. In the entrance region, the proton or heavy ion travels extremely fast before slowing down and coming to stop at larger depths. Briefly, the primary processes that cause this behavior are attributed to collisions with atomic electrons, scattering with atomic nuclei and direct collisions with a nucleus [2, 22]. The cross sections of these interactions vary with energy (inherently depth because a proton's energy will change at larger depths), suggesting that the dose deposition per particle track will also vary. This results in the Bragg Peak PDD in Fig.(1.1) (yellow and green). It is through careful manipulation of this behavior that radiologists are able to target deep-seated tumors with high doses of radiation whilst delivering lower doses to the surrounding healthy tissues [23].

Photons interact with the medium through which they are propagating via secondary electron interactions primarily [11]. However, the massless nature of the photon allows it to propagate deeper in tissue than protons or ions because the electron interaction cross sections are smaller than that of heavier particles. Therefore, the photon deposits more dose when it enters a medium than heavy particles and gradually decreases at larger depths. Additionally, more interactions occur between a photon and the medium in the entrance region before slowly decreasing as it propagates to larger depths (note that it does not come to a complete stop in Fig.(1.1) unlike the heavier particles). Hadron therapy is advantageous over photon radiation for this reason because their behavior can be tailored such that nearly all of its dose is deposited in a small region whilst photons cannot be controlled in this manner.

The primary cause of radiation-induced cell death is through lethal damage to the DNA caused by ionisations on the DNA segment [24]. Another advantage of hadron therapy is the high linear energy transfer (LET) of protons and heavy ions compared to x-rays. LET is defined as the energy deposited by a single particle per unit length, i.e. protons and heavy ions produce more ionisations per unit length than x-rays and the effect increases as the beam energy decreases. For this reason, hadron therapy is more effective on radio-resistant cells than conventional, photon-based radiation treatments. One can quantify this effect by calculating the *radiation biological effectiveness* (RBE) which is defined as the ratio of doses required for two different radiation types to achieve the same biological effect.

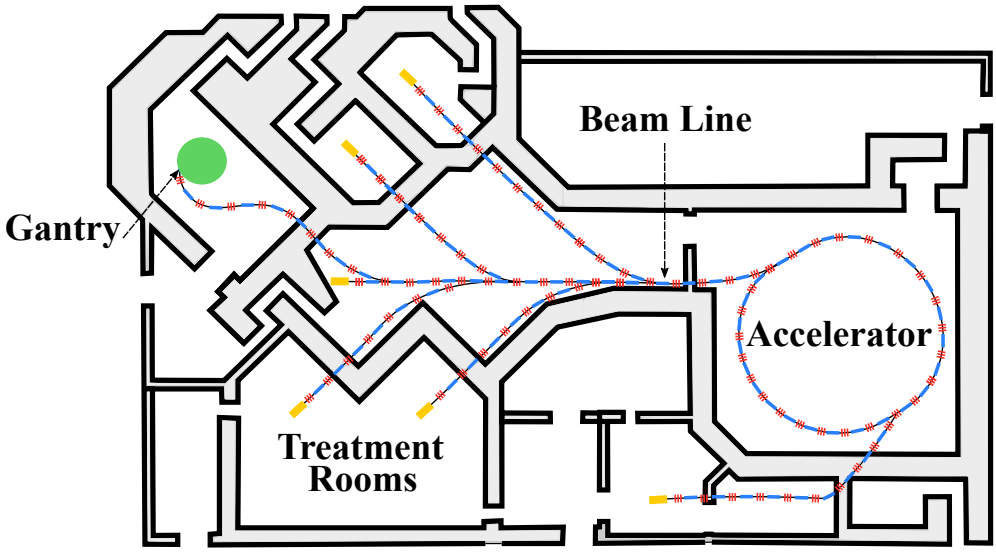


Figure 1.4: An overview of a hadron therapy facility in development in South-Eastern Europe with a particle accelerator, gantry and fixed beam line.

Figure 1.5: A gantry used to treat patients with hadron therapy. (Image courtesy of *https://www.philips.com.au/*).

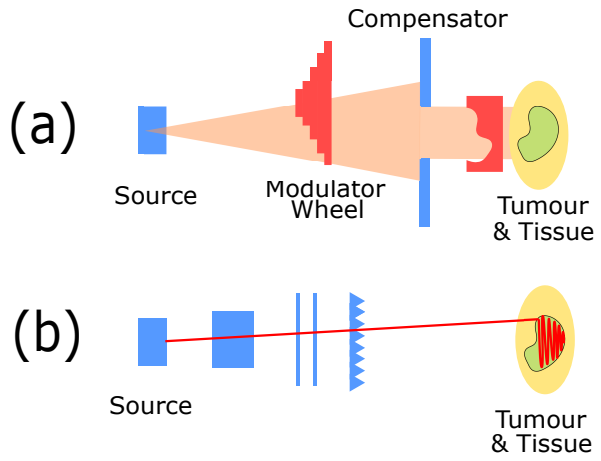


Figure 1.6: Methods of beam delivery (a) passive scattering method and (b) pencil beam method.

In clinical proton therapy treatments, an RBE of 1.1 is used clinically despite conflicting evidence [25–27]. As an example, a proton RBE of 1.1 would imply that 2.2Gy of photon radiation would be required to induce death in the same percentage of cells as 2.0Gy of proton radiation (RBE $= D_x/D_p = 2.2/2.0$). Numerous studies have seen the proton RBE rise to >1.5 for higher LET or larger depths on the Bragg Peak [28]. The RBE has also been shown to decrease as a function of absorbed dose, demonstrating less cell killing efficiency at high doses (see Fig.(1.7)). This evidence is vital to the success of hadron therapy because underestimating the RBE can result in the delivery of incorrect doses to patients, affecting the quality and outcome of their treatment. Accurate models to predict this behavior are therefore critical to treatment planning for cancer patients.

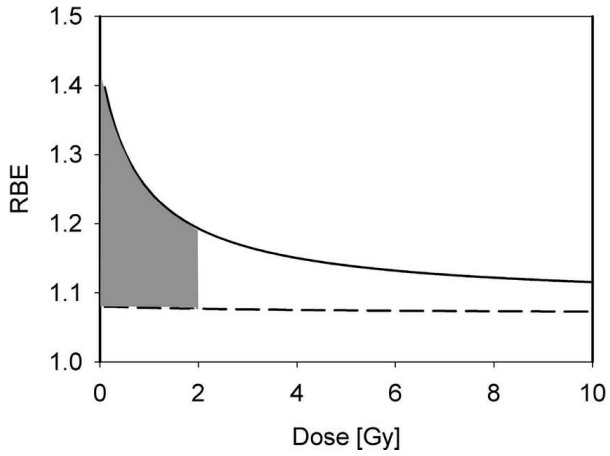


Figure 1.7: Proton RBE as a function of dose for low-LET radiation (Image adapted from [28]).

As is the case for all radiation therapies, the effectiveness of hadron therapy is governed by intrinsic biological parameters. Cellular response is often dependent

on the species from which the cell line originates and their genetic make-up. These parameters have been shown to affect the radio-sensitivity of the cell [29]. Conditions inside of the cellular environment are dependent on the "biological timeline" of the cell which is governed by the cell cycle [11]. Collections of cells are often asynchronous in nature[1] meaning that the radiation sensitivity varies between cells.

Cells are primarily damaged by ionising radiation via breakages in the DNA helix [24] and can occur with varying complexity (we call these breakages "lesions"). They can be induced via a series of *direct* and *indirect* processes. Direct effects are lesions induced by the deposition of energy on or within the vicinity of DNA and typically come from physical interactions (such as ionisations and excitations). Indirect effects induce lesions via secondary products originating from events that occur near the DNA but not close enough to cause a direct effect and often come from the production of chemical species (via the radiolysis of water) [11].

The cellular response to ionising radiation is a process undergone by all cells on a daily basis [24]. When exposed to fatal amounts of radiation, cells will undergo a series of processes which attempt to repair the DNA lesions induced. The cell's ability to repair is governed by the cell cycle phase in that repair enzymes exist with varying concentrations according to the cell cycle phase that it is in [11, 30].

When damages are induced in a cell, they are characterised as being *lethal*, *sub-lethal* or *potentially-lethal* [11]. By definition, a lethal lesion cannot be repaired and will always contribute to cell death. Sub-lethal damage is repairable unless additional lesions are induced before it repairs completely. This damage type is common in fractionated radiotherapy. Potentially-lethal damage is repairable but can also lead to a lethal lesion if not correctly repaired.

The number of lesions present inside a cell is correlated with its probability of survival. Often only a small number of lethal lesions are required to cause cell death. When a large number of lethal lesions are present, the cell can undergo apoptosis (programmed cell death) or continue through its cycle until it attempts mitosis (cell division) [11]. In the latter case, the damages are usually detected prior to division by the p53 gene - known as the "tumour suppressor" gene.

Overall, a cell is classified to have "*died*" when it has lost its "reproductive integrity". To elaborate, the cell can remain intact and be able to produce proteins or enzymes without being able to proliferate indefinitely and produce progeny. In this instance, the cell is classified as dead by definition [24]. In a multi-cellular experiment, the number of cells left surviving after exposure to ionising radiation is defined as the *survival fraction*.

[1] They exist in different cycles at once.

The process of cell irradiation and death is extremely complex and it is difficult to simulate all parameters within a single model. Currently, many radiobiological models exist and often aim to describe a select endpoint (such as lethal damage yield or survival fraction) under different conditions [23,31–33]. The more processes simulated in a single model, the more parameters are involved. The aspects of cellular response to ionising radiation are discussed in more detail in Chapter 2.

1.2 Modelling Radiobiological Phenomena

The modelling of radiobiological systems may be divided into two categories; *deterministic* and *stochastic*. A deterministic model consists of a series of equations in which specific information is substituted to gain additional knowledge about the system. Stochastic models simulate a system with a degree of randomness or noise, by definition. Deterministic models produce the same result with multiple repetitions, whereas a stochastic model can generate slightly different results.

In the 1940's the first target theory was proposed which quantitatively defined cell response to ionising radiation using simple cells, bacteria and yeast [34]. It achieved this by defining a series of biological targets capable of being affected by radiation - this can be DNA targets or the cell itself. It predicted the extent of radiation damage to a series of targets by measuring the number that are "hit" or "missed". The theory incorporated a degree of randomness by assuming that the response of a target after being hit is stochastic in nature [11]. The number of targets hit can be correlated to a particular endpoint describing the overall response of the system.

Many initial radiobiological models are formulated from target theory, including the single hit, single target and multiple hit, single target theories [23, 34, 35]. The former assumes that a single target is contained in each cell which is constantly exposed to radiation for time t where the target can be affected after a single hit. The theory then assumes that in a small fraction of time Δt, the target with probability k of being hit per unit time has the probability $k\Delta t$ of being hit in the interval Δt. This forms the basis for Poisson statistics which are discussed in more detail in Chapter 5. The total number of targets present is denoted by N_0 and as this number increases, the probability of a target being hit also increases. We therefore write the probability k as κN_0 where κ is the likelihood of a target being hit [11]. If we rewrite the probability of a target being hit in time interval Δt, we get $\kappa N_0 \Delta t$. Therefore, the probability of a target not being hit is $1 - \kappa N_0 \Delta t$ and based on the assumption that a single hit will affect the target. After n time intervals of length $\Delta t/t$ this expression becomes $(1 - \kappa N_0 \Delta t)^n$, reflecting the survival probability (see Fig.(1.8)). If we denote $\kappa N_0 \Delta t$ as λ and take the limit

$$\lim_{n\to\infty} \left(1 - \frac{\lambda}{n}\right)^n = e^{-\lambda} \quad .$$

The definition of λ has since been denoted as D/D_0 where D is the absorbed dose and D_0 is the dose required to affect 1/e cells. This allows us to rewrite the survival probability of N_0 cells in terms of its exposure to dose D.

$$\boxed{\textbf{Single hit, single target: } S = \frac{N}{N_0} = e^{-\frac{D}{D_0}}}$$

The single hit, multiple target theory [34] extends the previous theory to DNA of which there are many targets inside a single cell. It assumes that a single hit to a single target will affect the cell, however multiple hits are required to induce cell death. This theory is more reminiscent of the true system because multiple DNA lesions are required to affect a cell. The basic premise is that we take the survival probability derived from the single hit, single target theory and raise it by power m which represents the total number of targets that need to be hit inside of the cell to induce death. The result is a "shouldered" survival curve with low cell killing efficiency for small doses and higher for high doses (orange line in Fig.(1.8)).

$$\boxed{\textbf{Single hit, multiple target: } S = 1 - \left(1 - e^{-\frac{D}{D_0}}\right)^m}$$

Currently, the most well-known and utilised radiobiological model in the literature is the Linear-Quadratic (LQ) model designed in the 1970's [23, 35, 36]. It combines the single hit single target and double hit single target theories to form the survival expression below where a single target can be affected by a single lesion event or two independent events to form single cluster of lesions. This again improves upon the previous two target theory models which do not account for the interaction of individual hits to form lethal lesions.

$$\boxed{\textbf{Linear-Quadratic Model: } S = e^{-\alpha D - \beta D^2}}$$

In this expression S is the fraction of cells that survived after exposure to dose D of a given radiation type. The parameters α and β typically describe the contributions of single and multiple hits contributing to cell death, respectively [11]. Although, the interpretation of α and β can differ slightly between models [37]. However, this results in a steeper decline in the "shoulder" region and a defined curve at high doses compared with previous target theories (see Fig.(1.8)). The statistical basis of the LQ model is discussed in more detail in Chapter 5, however it is worth noting here that the underlying stochastic process of DNA lesion formation is based on Poisson statistics. It has had good success in describing dose response data under specific conditions such as low-LET radiation. Also when the data is collected *a priori* and the survival equation is fitted [26, 27, 38]. However, the literature has uncovered evidence that the LQ model may not be the most appropriate model to use in some

situations such as in the high-LET or high dose regions [18–20, 25, 39, 40] due to the small parameter space and the complex biological, chemical and physical parameters to account for in the system. In the literature, improvements have been made upon the LQ model [31, 41–43] by accounting for some biological processes or adding additional parameters for accurate modelling in the high dose region. However the small parameter space still appears to deplete the model's validity, an aspect that is revisited throughout this thesis.

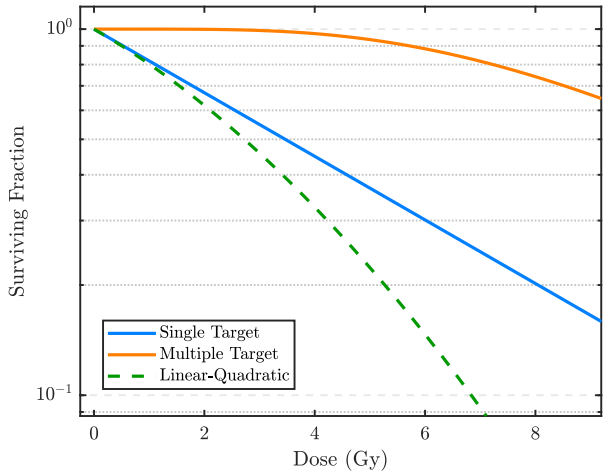


Figure 1.8: A comparison of the single hit single target, multiple hit single target and Linear-Quadratic models.

Models *beyond the LQ*, which will be discussed in more detail in Chapter 2, include the Local Effect Model (LEM) [44] and the Microdosimetric Kinetic Model (MKM) [45]. Both models are currently used clinically in hadron therapy centres across Germany and Japan [30]. They employ a mechanistic approach to predict endpoints such as cell survival by counting the number of lethal events in a group of cells. Other models, such as the Two Lesion Kinetic (TLK) [33] and Lethal-Potentially Lethal (LPL) [32] models, predict the number of lethal lesions alike to the LEM and MKM whilst modelling the DNA repair processes that occur inside a cell post-irradiation. These models involve larger parameter spaces than the LQ model and are generally considered to be more accurate [40].

1.3 Objectives for the Current Work

The objectives of the current work are

- identify discrepancies and verify/falsify the validity of the LQ model.

- perform a systematic, rigorous analysis of the LQ model on experimental cell survival data and perform an in depth statistical analysis on the fits with different cell lines and radiation exposure conditions.
- propose a new stochastic model for cellular dose response.
- develop a reliable Monte Carlo simulation workflow to generate a complete set of data for different endpoints including lesion yield and cell survival.
- using Monte Carlo techniques, simulate the irradiation of the V79 (Chinese Hamster) cell line with known biological parameters and different exposure conditions - by varying the beam energy and the depth of the cells in water.
- verify/falsify the proposed stochastic model by testing on experimental data from the literature and our MC simulated data for different endpoints (DNA damage yield and survival).

These objectives were carried out using rigorous statistical analysis techniques, numerous software programs and tool-kits and Machine Learning algorithms. The initial work was separated into two sections, with the first being a statistical study of experimental proton cell survival data and the second being the generation and statistical study of MC simulated proton cell survival data. The flowchart provided in Fig.(1.9) summarises this work. The two sections were then combined to develop the overall conclusions of the study.

The survival data from the literature was supplied to us by Professor Cynthia Keppel and Dr. Pawel Ambrozewicz at Jefferson Lab in Newport News, Virginia [17] and additional data obtained from the Particle Irradiation Data Ensemble (PIDE) [46]. The data described cell survival for numerous cell lines, exposure conditions and biological parameters. The focus of this study is irradiation of cells with protons.

All of our analysis codes are written in MATLAB™(R2019a) to undertake all of the statistical and regression analyses that were performed on the experimental and simulated data. Our findings showed that there are cases for which the LQ model does not describe the data sufficiently well. This formed the basis of our idea to develop a model that describes radiation-induced cell death with more accuracy under different biological and exposure conditions.

The newly proposed model presented in this thesis aims to re-assess the underlying statistical assumptions used by the LQ model that the formation of lethal lesions obey the Poisson process. The model achieves this using the Fractional Poisson Process (fPp) as an alternative to the standard Poisson process which is the foundation of the LQ model and other target theories already discussed. The proposed theory was developed in close collaboration with Dr. Markus Kreer of Johann Wolfgang

Goethe-Universität, Frankfurt, Germany.

Initial testing for the proposed model utilised the experimental proton cell survival data provided to us, for which curve-fitting was performed to each data set. The experimental data was only available for select parameters, suggesting that this alone would not be enough to test the model developed in this thesis effectively. This warranted the need to generate cell survival data with parameters not currently available in the literature, allowing the model to be tested effectively.

We test the proposed model using experimental proton cell survival data provided to us by fitting the model's survival function and assessing the goodness-of-fit. The literature from which the data was derived does not always report all the relevant parameters. Therefore, to test the our proposed model effectively, we require a set of dose response data where the relevant physical, chemical and biological parameters are known.

The simulated proton cell survival data was generated using the MC toolkit *Tool for Particle Simulation* (TOPAS) [47], in which cells were randomly placed in a monolayer formation[2] and a selected number of primary particles were generated to interact with the cells. The software output from these simulations gives a detailed account of the track structure for each particle that traverses a cell. This information includes the coordinates of each ionisation event, the energy absorbed from the interaction by the cell, the cell number and organelle[3] effected and the interaction process associated with the event.

The simulation process was rigorously tested and continuously improved to derive the most accurate results whilst optimising computation time and memory requirements. The importance of simulating specific processes inside the cells was investigated. Due to time constraints we were unable to perform simulations for other cell lines or under different biological conditions. We performed 30 repetitions on every dose point in the high and intermediate LET range. 5 repetitions were performed for the lowest LET values considered.

Our simulations were carried out on the Phoenix and HPC1 supercomputing clusters (University of Adelaide, Adelaide, Australia). The memory requirements and processing time for larger doses, varied according to particle LET. High-LET beams required an average of 15 CPU cores with a processing time ranging from 24 to 38 hours. Low-LET beams required >45 hours on average with the same number of CPU cores. This is due to more particle tracks required per Gy of low-LET radiation which was in the range of 10^6Gy^{-1}. High-LET radiation simulations required approximately $10^5 \text{particles} \text{Gy}^{-1}$. The output files from each simulation containing

[2] A monolayer formation means to lay the cells on a flat surface.
[3] An organelle is a given component of the cell - for example the nucleus and cytoplasm.

interaction information and coordinates were approximately 100GB in size per repetition and 30TB of storage space was required to store all of our results.

When the damage sustained by each cell is known, a post-simulation analysis was performed to estimate the probability of survival for each cell using the TLK model. The model used in this work was based on previously developed models [33, 48, 49]. The importance of ionisation efficiency, repair kinetic of cell lines and effects of radiation exposure to different cell organelles were investigated.

Our newly proposed model was tested together with the LQ model on our simulated data for multiple exposure conditions. The data simulated were identically and independently distributed (i.i.d.) from each other and enough repetitions of each data point were generated, such that a statistical analysis could be undertaken.

The two sections of this work were then brought together and a conclusion was developed. The final objective of this study is to highlight the importance of statistical significance in radiobiological modelling. Time constraints restricted the number of simulations and parameters that could be simulated and analysed, therefore our results describe these select conditions and future efforts will focus on expanding this.

1.4 Outline of the Current Thesis

This thesis was produced in three components including an introduction and literature review; testing and methodology and results and discussions. **Chapter 2** consists of a literature review of the known physical, chemical and biological processes involved in cellular radiation response and an overview of the radiobiological models currently available. It concludes with a description of what entails an effective radiobiological model for use in the current thesis. The importance of making direct comparisons with existing experimental results at various biological endpoints is highlighted.

Chapter 3 explains how the Monte Carlo method is used for particle tracking inside of a material and a brief description of the framework of software toolkits Geometry and Tracking DNA (Geant4-DNA) and TOPAS (together with the TOPAS-nBio radiobiology extension) used to carry out the simulations in this work. This chapter concludes with an overview of what can be achieved when performing simulations in the current work using Geant4-DNA and TOPAS.

Chapter 4 provides an in depth description of the development of our MC simulation workflow used to generate simulated dose response data for model testing. All stages of the model in this study are described in detail and any decisions made that impact the accuracy-optimisation balance described above are tested. All of

our simulation parameters are given in this chapter.

Chapter 5 consists of a rigorous statistical analysis of LQ model fits to cell survival data using the world experimental data (an outline of the experiments analysed throughout this work is presented in Appendix D). The fits are tested against the assumptions of regression using error (residual) analysis and hypothesis testing. We aim to verify the limits of the LQ model under a range of radiation exposure and biological conditions including beam energy/LET and cell line.

In **Chapter 6** we propose a new stochastic model to describe radiation-induced cell death using a fractional Poisson count process in place of a standard Poisson process. We aim to increase the flexibility of the LQ model to better explain that data under conditions where the Poisson process breaks down.

Chapter 7 provides a comparison of the results generated using the simulation framework developed in Chapter 4. The results from various biological endpoints are derived from our simulations at multiple stages for direct comparison with experimental results present in the literature. These results include interaction yield and cross section comparisons, DNA damage yields, estimation of cellular survival and calculation of RBE.

Chapter 8 presents the results of testing the stochastic model proposed in Chapter 6 against experimental and simulated hadron therapy data. The model's performance under different physical, chemical and biological conditions is compared with the LQ model and the standard Poisson process.

In **Chapter 9**, an overview of the work performed in the this thesis is provided and accompanied by the conclusions made from the testing of our model in comparison with the LQ model. The chapter concludes with a discussion of the significance of our results and any future work required.

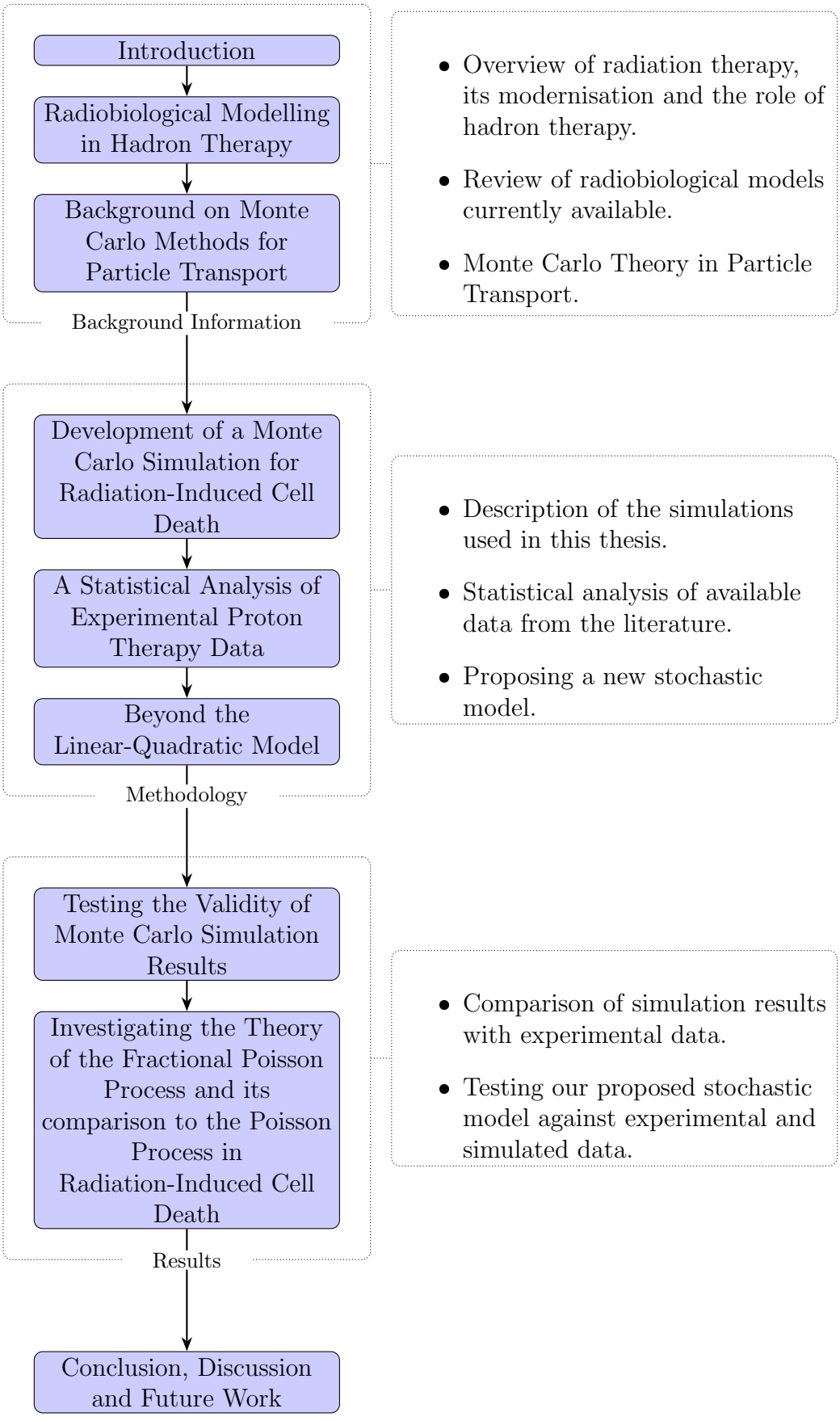


Figure 1.9: Structure of this thesis.

2

Radiobiological Modelling in Hadron Therapy

Radiobiological models are used widely in hadron therapy centres and research to optimise patient treatment and test new modalities. Developing accurate radiobiological models is at the forefront of radiation therapy research and plays an important role in the successful treatment of cancer patients. In this chapter, a review of current radiobiological models applicable to hadron therapy is conducted with an assessment of what encompasses an effective model.

Gaining a complete and correct understanding of cellular dose-response is at times a tedious and difficult task. In addition, solving empirical models is challenging computationally with the large number of parameters involved. Cellular response is governed by a series of physical, chemical and biological processes that occur inside a cell when it is exposed to radiation. The stochastic nature of these interactions means that these models often involve a trade-off between model accuracy and simplicity/computational optimisation.

This chapter consists of background information and a discussion on the relevant processes and stages that a cell undergoes when exposed to ionising radiation, with a focus on protons. The cell response consists of several processes that are physical, chemical and biological in nature. We describe how these processes contribute to the radiation-induced damage inflicted on a cell and how the cell repairs this damage. First the physical and chemical processes and how they induce damage are described. Following this is a description of the various damage repair pathways undergone by a cell and under which conditions cell death/survival is invoked. To conclude the chapter, a review of radiobiological models currently available in the literature is performed with a description of how the processes and stages discussed prior are implemented.

2.1 The Physics, Chemistry and Biology of Treating Cancer

The process of cell irradiation interactions and response can be separated into several stages [24] (see Fig.(2.1)). These are :

- **the physical stage** ($t < 10^{-16}$ s) - a very short stage in which energy depositions inside the cell cause ionisations and excitations,
- **the physico-chemical stage** ($10^{-16} < t < 10^{-12}$ s) - the products from interactions in the physical stage form a series of charged and uncharged products,
- **the chemical stage** ($10^{-12} < t < 10^{-6}$ s) - the products of physico-chemical reactions produce chemical species,
- **the cell response stage** ($10^{-6} < t <$ hrs) - where cell division (mitosis) is hindered or cell death is induced in response to the presence of damage,
- and **long-term effects** (years) - cells that contain lethal or mis-repaired lesions can continue to proliferate causing long-term health complications or secondary cancers.

Whilst the behavior of different radiation types can change the response of cells, the timeline of irradiation and cell response remains the same. Of all assumptions made in radiobiological modelling, the most common is that biological effects due to radiation are primarily caused by the formation of DNA damage by *direct* and *indirect* effects [24]. Direct effects are caused by interactions that occur on the DNA segment itself and are typically caused by physical processes such as ionisations or excitations. Indirect effects occur from species produced in the vicinity of a DNA segment and are typically caused by the production of free radicals via a series of chemical processes known as the radiolysis of water [50]. These processes are familiar to researchers but some aspects remain uncertain and are still to be explained.

After the commencement of irradiation and lesions are induced inside the cells, they will immediately begin to go through the repair process via several pathways (described later in this chapter). After the induction of damage on the DNA segment, they can either be repaired, mis-repaired or un-repaired, the latter contributes to cell death. A number of them will be too complex to be successfully repaired (lethal lesions) whilst some will partially repair. However, a majority of this damage will repair correctly, leaving only a small fraction un-repaired [30]. However, even a single complex DNA lesion is capable of inducing cell death [51] and therefore the presence of one lethal lesion can be detrimental to its reproduction ability. By definition, we characterise a cell as dead when it has lost its ability to reproduce indefinitely.

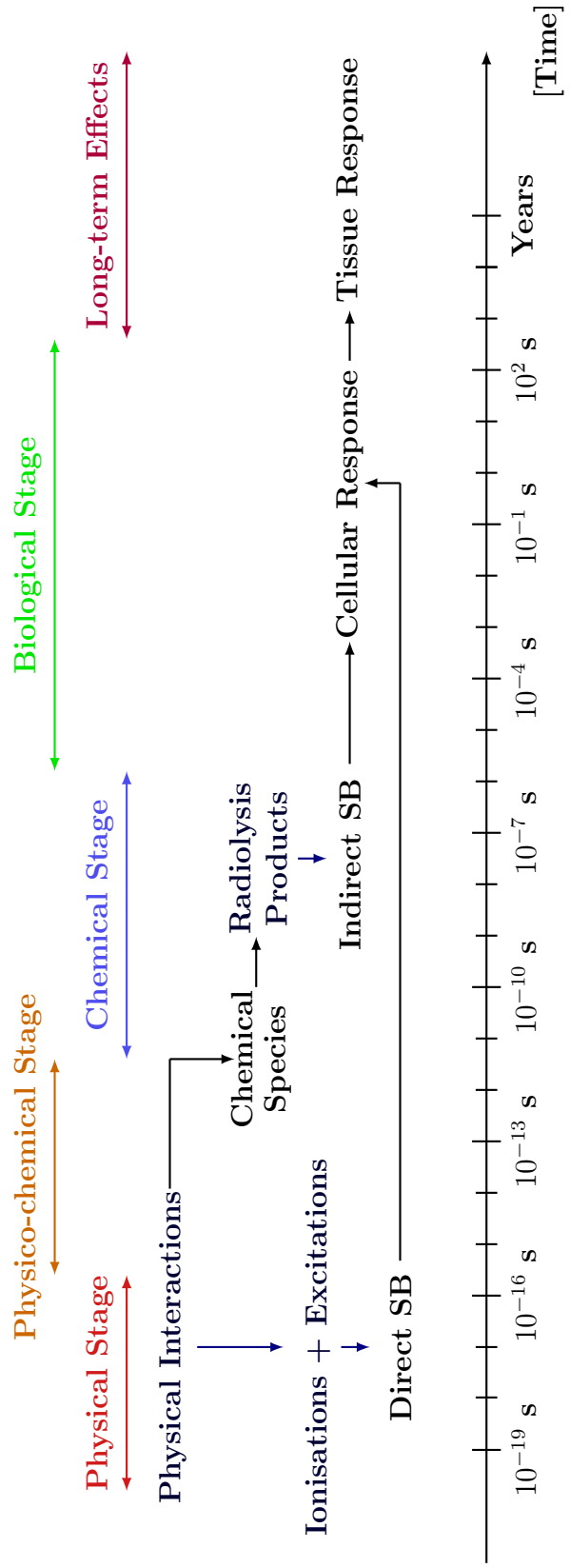


Figure 2.1: The time scale of cell response to ionising radiation.

2.1.1 Physical Interactions of Ionising Proton Radiation

The timeline of cell irradiation begins with the physical stage ($t < 10^{-16}$ s). It is this stage where direct DNA damage is caused by the products of a series of physical interactions which occur inside the medium as the particles propagate.

The primary mechanisms for energy loss by a proton propagating through a medium are via elastic and inelastic Coulomb interactions and nuclear scattering interactions [22], Fig.(2.2). Whilst Bremsstrahlung processes can occur, their interaction cross-section is so small at clinical proton energies that their dose contribution is negligible. Inelastic Coulomb interactions occur when a propagating proton ejects an atomic electron from an atom, Fig.(2.3(a)). As the proton slows down, the cross-section of this interaction increases, leading to the formation of the Bragg Peak. They also have the largest contribution to the proton range inside the medium (the depth to which the proton will penetrate the medium). Elastic Coulombic interactions occur when a proton travels in the vicinity of an atomic nucleus where it experiences a repulsive force, changing its trajectory, Fig.(2.3(b)). The final common interaction type is the production of secondary particles via nuclear scattering interactions between the proton and an atomic nucleus, producing secondary protons, neutrons, photons or a recoil nucleus, Fig.(2.3(c)), through which dose deposition continues.

These physical interactions can produce a series of chemical species which fuel the physico-chemical and chemical stages that follow. If we consider water (a common substitute for biological materials) to be the medium through which the proton is propagating, these physical processes lead to ionisation ($H_2O \longrightarrow H_2O^+ + e^-$) and electronic excitation of the water molecules ($H_2O \longrightarrow H_2O^*$). If the processes occur in a DNA segment, it will likely induce a DNA strand break (SB).

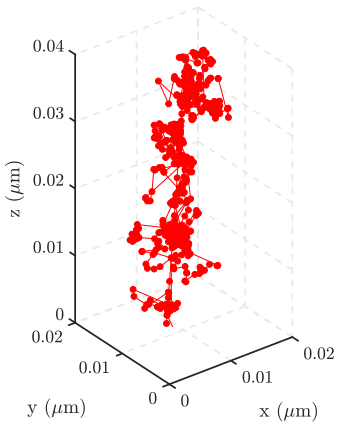
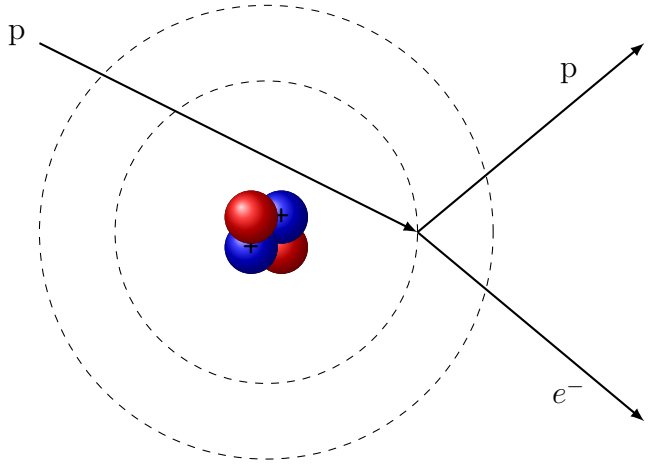
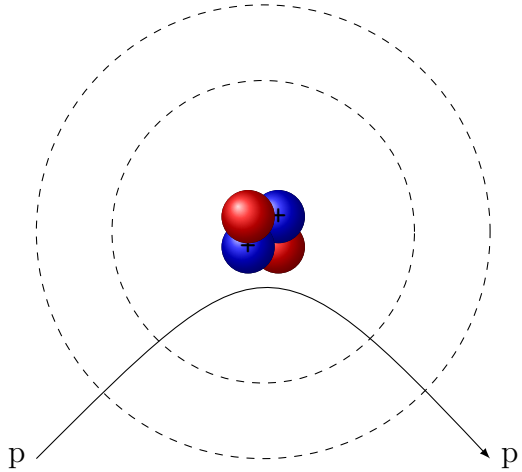


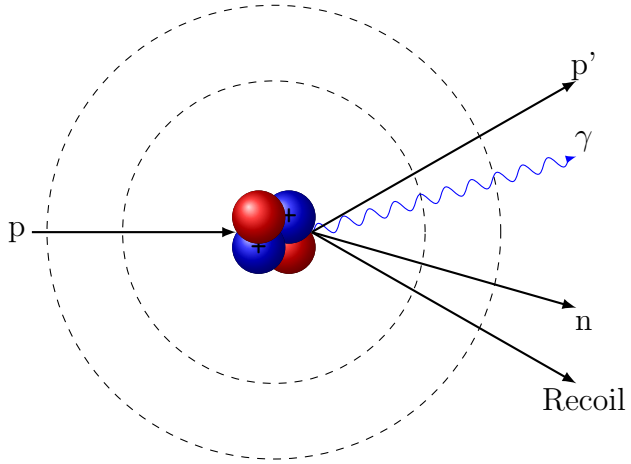
Figure 2.2: Visualisation of a 1MeV proton propagating through liquid water generated using the Monte Carlo toolkit TOPAS.



(a) Inelastic Coulomb Scattering



(b) Elastic Coulomb Scattering



(c) Inelastic Nuclear Scattering

Figure 2.3: Physical interactions of protons propagating through a medium.

2.1.2 Indirect Effects from Radiolysis of Water

Following the immediate physical interactions are the production of radical species formed from the radiolysis of water molecules. The process is initialised in the physico-chemical stage via various reaction pathways [50] and includes the production of charged and uncharged ($\bullet$) radical species via ion-molecule reactions, Eq.(2.1), dissociative relaxation, Eq.(2.2) and the production of a solvated/aqueous electron, Eq.(2.3).

$$H_2O^+ + H_2O \longrightarrow H_3O^+ + {}^\bullet OH \tag{2.1}$$

$$H_2O^* \longrightarrow H^\bullet + {}^\bullet OH \tag{2.2}$$

$$e^- + H_2O \longrightarrow e^-_{aq} \tag{2.3}$$

Following production of the above species, the chemical stage commences where the radicals react with each other and the surrounding water molecules to form a range of radical species ($e^-_{aq}, H^\bullet, {}^\bullet OH, HO_2{}^\bullet, OH^-, H_3O^+, H_2$ and H_2O_2), some of which are charged. The reactions proceed via a number of pathways, some examples of which are shown below in Eqs.(2.4)-(2.7).

$$H_2 + O \xrightarrow{H_2O} H_2 + 2\,{}^\bullet OH \tag{2.4}$$

$$H_2O^+ + e^- \xrightarrow{H_2O} {}^\bullet OH + H_2 \xrightarrow{H_2O} {}^\bullet OH + H_2 + OH^- \tag{2.5}$$

$${}^\bullet OH + e^-_{aq} \longrightarrow OH^- \tag{2.6}$$

$$H_3O^+ + e^-_{aq} \longrightarrow H^\bullet \tag{2.7}$$

As the radicals are produced, they tend to propagate away from the ionisation track thus expanding the volume over which DNA damage can be produced. This stage becomes particularly important when using low-LET radiation due to its sparse ionisation track [52]. Alternatively for high-LET radiation, highly reactive radicals (e.g. ${}^\bullet OH$ and $H^\bullet$) react quicker to produce less reactive species due to the larger density of radicals produced. This means that there are less species available to induce DNA damage. Conclusively, physical interactions appear to dominate the DNA damage yield for high-LET radiation whilst indirect damage dominates at low-LET.

The resultant yield of each species is known to deplete towards the end of the chemical stage due to a decrease in the reaction cross section. This means that the yield of highly reactive radicals is high initially before decreasing with time. On the

contrary, the yield of secondary products such as H_2O_2 and H_2 will increase with time.

2.1.3 The Target of Ionising Radiation

The primary target of ionising radiation to induce cell death is the DNA (Deoxyribonucleic acid) helix [11]. It consists of four different molecules (Adenine (A), Guanine (G), Cytosine (C) and Thymine (T)) and are arranged in a double helix formation (see Fig.(2.4)). The bases are paired together to form base pairs (BP) in form A-T and C-G. Each base pair is held in a double helix formation by a sugar-phosphate backbone known as a strand. The base and strand are joined via a pentose sugar, a segment known as a nucleotide. Ionising radiation can damage the DNA by breaking segments via the direct and indirect effects discussed above. Such breakages can occur on the strand or base. Cell death is most correlated with breakages to the strands in multiple places within close proximity because an isolated strand break is easily repaired [2,11,53]. Possible breakage types are discussed with more detail in the following section.

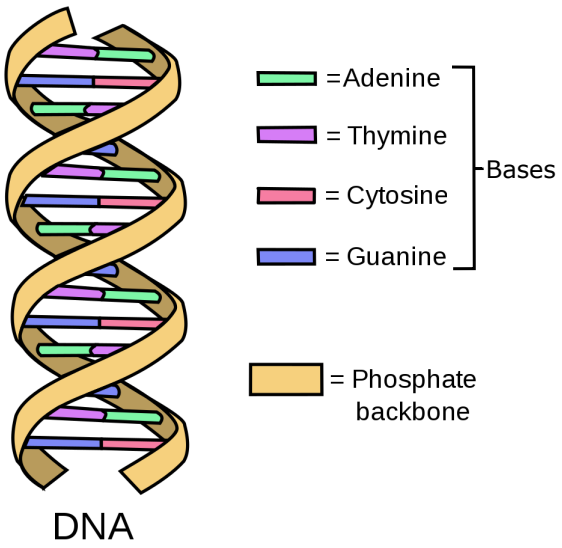


Figure 2.4: The DNA double helix (Credit: wikicommons).

2.1.4 Biological Mechanisms of Repair/Mis-repair

Following the production of DNA damage, a eukaryotic cell will attempt to repair via a series of pathways. Such a response is not only unique to the cell after sustaining radiation damage but is a continuous process undergone by every cell as it progresses through the cycle. As discussed above, damage sustained by DNA can

come in two forms, a base damage or a strand break (SB) - both resulting in disruptions to the DNA sequence structure [2].

The probability of correct repair of a DNA lesion is dependent on its complexity. Typically SB damage is placed in three categories, the single strand break (SSB), the double strand break (DSB) and the complex double strand break (cDSB). These are defined as clusters of damage that contain one, two or greater than two SBs within a distance of 10-20 base pairs (bp, the distance between nucleotides that join the DNA sugar-phosphate strands together) on a given DNA strand respectively (as per Fig.(2.5)). Radiation will often produce a DSB yield of 30-50Gy^{-1} [33], however most of these damages will be repaired via several pathways.

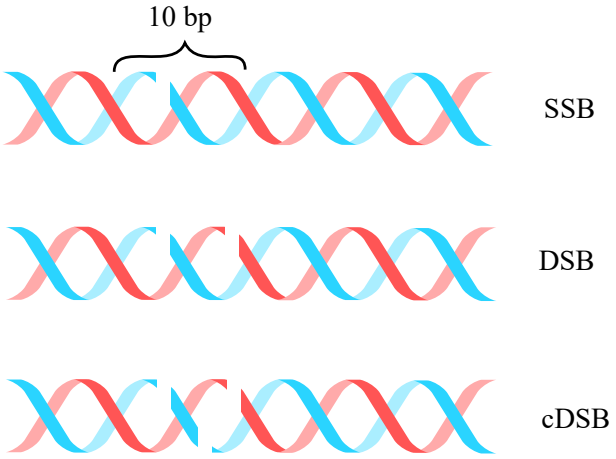


Figure 2.5: DNA damage complexities.

Eukaryotic cells are able to identify when these damages are present and respond by activating a range of repair pathways. The pathways that have the highest contribution to the repair of base and strand damages are :

- base excision repair (BER),
- nucleotide excision repair (NER),
- mismatch repair (MMR),
- non-homologous end joining (NHEJ),
- and homologous recombination (HR).

These pathways are most dominant throughout various stages of the cell cycle [54]. Briefly, the cell cycle is progressed by four stages [11] :

- Mitosis (M) - where the cell divides into two,
- The "first" gap (G_1) - the preparation phase for DNA synthesis,

- The synthetic phase (S) - where DNA synthesis is undertaken,
- The "second" gap (G_2) - the preparation phase for cell division (i.e. mitosis).

The first three processes listed above contribute to the repair of base and strand breaks. Base excision repair (BER) is the process where a damaged base which could lead to DNA mutations via misrepair are removed from the helix. This occurs via DNA glycosylase enzymes which flip the damaged base outside of the helix and remove it entirely. This leaves the helix with an abasic site (a segment of DNA without a base) which can provide a pathway to SSB repair. Nucleotide excision repair (NER) is a process that corrects misrepaired bases. It is able to identify DNA lesions, cut the bases on either side of the lesion and rejoin the damaged side via DNA synthesis and ligation (i.e. the rejoining of broken ends of the DNA structure). The final pathway for base repair is MMR. It identifies errors in the DNA base sequence which may have formed during other repair processes and corrects them. This is performed by removing the DNA segment containing the mismatch, copying it using a template with the same sequence as the cut away segment and replacing it via DNA polymerase (an enzyme that can synthesize DNA) [55].

The repair of DNA DSBs have the largest contribution to cellular radiation response because their mis-repair can lead to lethal damages and ultimately lead to cell death. Homogeneous-recombination (HR) is the most accurate method of DSB repair [54] but is slower than NHEJ due to its complexity. It proceeds via a series of steps and is activated when opposing strands on the DNA helix are damaged (see Fig.(2.6)). It utilises synthesized DNA strands with the same base sequence as the damaged region allowing for a higher probability of accurate repair than NHEJ. This process is only active in the S and G_2 phases of the cell cycle.

Non-homologous end joining (NHEJ) is the simplest and most rapid pathway to DSB repair, owed to the fact it does not require a DNA duplicate to proceed - unlike HR. It is also the most dominant of the two processes and is available to the cell regardless of where it is in the cell cycle. In this process, the chromosome ends that have been broken are immediately reattached via repair proteins on the "sticky ends" of the DNA segment (see Fig.(2.7)). There are no steps in between the beginning of the process and the end result. NHEJ has a small probability of mis-repairing a lesion as it does not always correctly identify a lesion [56].

These pathways often repair most of the DNA lesions induced by radiation, however high-LET radiation makes the repair process more difficult due to the increased interaction of DSB damages. That is to say high-LET radiation induces more complex DSBs and is therefore highly efficient when it comes to cell killing. Each pathway has a significant impact on the final result, thus strongly advocating their inclusion in radiobiological modelling.

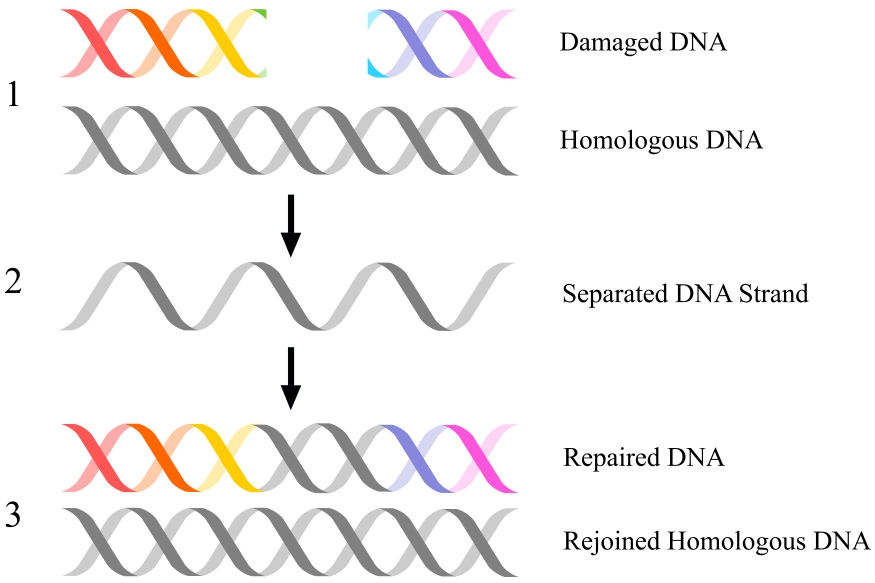


Figure 2.6: Homologous DNA rejoining (HR).

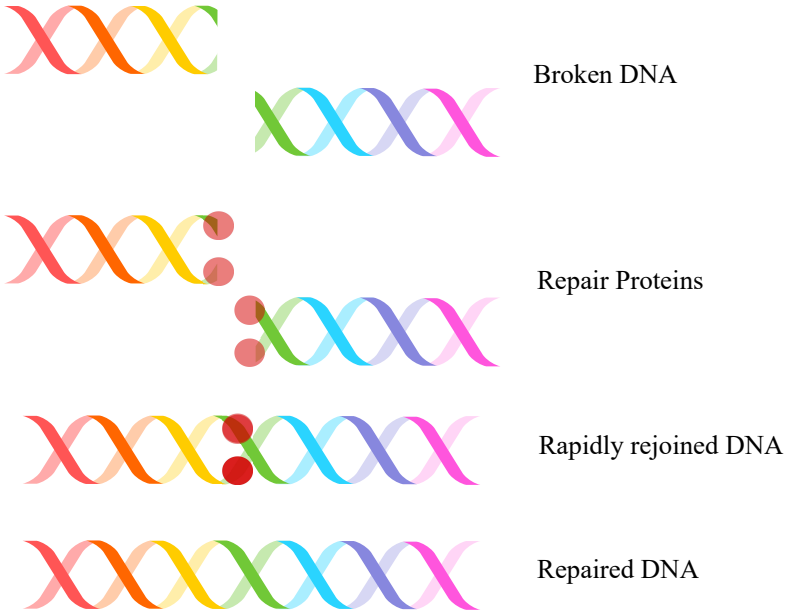


Figure 2.7: Non-homologous end rejoining of DNA (NHEJ).

2.1.5 Biological Influences of Radiation-Induced Cell Death

In addition to the stages of cellular irradiation, there are factors within the cell environment that can influence radio-sensitivity and the resulting shape of the dose-response curve.

2.1.5.1 Oxygen content in cells ($[pO_2]$)

The role of pO_2[1] content inside a cell and its resultant radio-sensitivity is a well documented concept [29, 57]. The effect suggests that a cell's radio-sensitivity is at its highest with an oxygen content of 100% (oxic) and gradually decreases with pO_2 concentration, reaching its lowest value at 0% (hypoxic). The reason for this is the lack of molecules available for charged radiolysis species to react within the vicinity of DNA, causing a decrease in the DNA damage yield.

This effect can be quantified by taking the ratio of the doses required to achieve the same biological effect under oxic and hypoxic conditions, giving rise to the oxygen enhancement ratio (OER), Fig.(2.8). Experiment shows that this value is approximately 3 on average for most cell lines [30] but can also vary with LET.

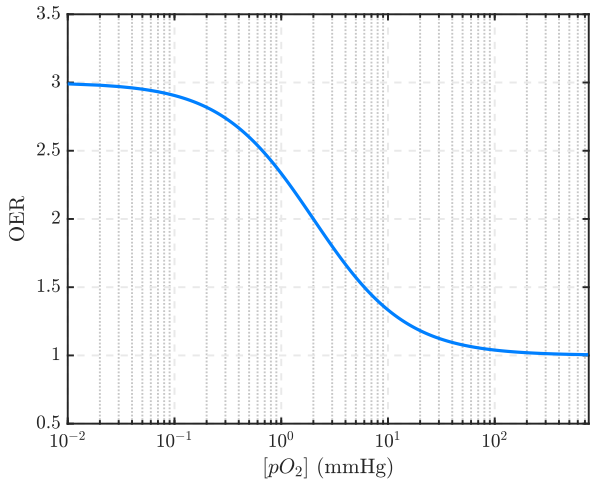


Figure 2.8: Oxygen Enhancement Ratio as a function of $[pO_2]$.

What makes this effect particularly important in cancer research, is that regions of cells within a tumour become hypoxic over time [11,24]. It has been shown that as cancerous cells proliferate to form a tumour, the cells enclosed in the central region of the tumour are the most hypoxic whilst the outer cells are the most oxic. This variance in oxygen conditions can make the process of treatment planning difficult

[1]Intracellular oxygen tension.

as one would need to know the oxygen content of each cell inside of the tumour to determine the OER and calculate the dose required to achieve optimal cell killing.

2.1.5.2 Cell Cycle Phases and Cell Death

The phase of the eukaryotic cell cycle is known to impact a cell's radiation sensitivity [58]. The cell is at its most radio-sensitive in the G_2 - M phase, less sensitive in the G_1 phase and the least sensitive in the S phase. Part of the reason for this is the presence or lack of repair pathways available to the cell in different phases already discussed in the beginning of this section.

Cell death can take place via two pathways, necrosis and apoptosis. Necrosis is defined as the premature death of a cell due to the presence of damage whilst apoptosis is known as programmed cell death and is triggered by the cell in response to DNA damage. The cell cycle phase has also been shown to influence apoptosis due to the relative presence of the "p53 protein" across cell cycle phases [59]. Its role is to regulate the cell as it changes phase whilst checking for mutations that may lead to the proliferation of unhealthy cells. The concentration of p53 inside the cell is positively correlated with apoptosis, hence the dependence of apoptosis on the cell cycle.

Conclusively, cell cycle and oxygenation distributions across a collection of cells have a significant impact on the resulting dose-response curves. This does not necessarily introduce uncertainties into the measurement but is important to be mindful of because these parameters can easily turn a radio-sensitive cell into a radio-resistant cell. The source of such effects and their impact on the result should be known for accurate dose calculations for treatment.

2.2 Modelling Radiobiological Hadron Therapy Data

There are numerous models used to explain the response of cells (such as DNA damage yield or survival) when they are exposed to radiation. It is typical of these models to account for as many of the processes described in the previous sections as possible whilst balancing the trade-off between accuracy and solvability of the model. An ideal model would account for all of the damage and repair processes described in Section (2.1) for numerous biological conditions inside the cells. It should explain cell response to ionising radiation in stages leading to cell death or survival. However, the complexity of the system makes this task extremely difficult. Some of the most accepted models that incorporate the processes from Section (2.1) are described in more detail in the following section.

2.2.1 Proton Radiation Characteristics and Their Role in Radiobiological Modelling

One of the distinct advantages hadron therapy has over conventional, photon-based radiotherapies is that it densely ionises the tumour as the beam propagates through it whilst sparing the surrounding tissue. Whereas a photon beam will sparsely ionise its surroundings as it travels through a patient, Fig.(1.1). The reason for this behavior in protons and heavy ions is their large rest mass in comparison with a photon which suggests that it has a high probability of interacting with a material as it propagates. The photon on the other hand is massless and therefore has a smaller probability of interacting with the material through which it is propagating with respect to protons and heavy ions. Consequently, the probability of a particle interacting with a material as it propagates through is correlated with its mass.

In the context of a single proton propagating through a tumour, it will inflict a far greater amount of damage in the cell than a photon will. It is also worth noting that these ionisations tend to be localised to the track of a single particle and if it traverses a DNA helix, the resulting DNA damage is typically clustered.

This ionisation density of a particle along its track (the energy lost after propagating through a material of infinitesimal thickness dx) is defined quantitatively as the linear energy transfer (LET).

Definition 2.2.1. *LET is the average energy absorbed by a medium per unit length of a single particle track,*

$$LET = \frac{dE}{dx}$$

The LET is an averaged quantity because the dose deposited in a single track will vary.

LET can be computed using dose-averaging (LET_d) or track-averaging (LET_t). Figure (2.9) visualises how the LET of the same pattern of dose depositions is computed using each method. Let us assume each circle represents an equal energy deposition E. The track-averaged LET (top of Fig.(2.9)) is determined by splitting the track length into equal partitions, dx. The five segments of length dx contain an average deposition of $4E$ (four depositions of energy E). To determine the dose-averaged LET we compute the average distance over which an average energy deposition dE occurs. In the Fig.(2.9) (bottom) we observe five energy depositions totalling $5E$ which occur over four different lengths dx_1, dx_2, dx_3 and dx_4. By determining the mean of the four lengths dx we have retrieved the dose-averaged LET.

The dose-averaged LET is the most commonly reported quantity in experiment (in fact, it is recommended for RBE studies) [25, 46] and will be used throughout the entirety of this thesis. The LET is also dependent on the speed at which the particle is moving. The slower the particle moves, the more energy it will deposit

per unit length. Therefore, it is logical to think that the LET varies with depth and particle energy, Fig.(2.10).

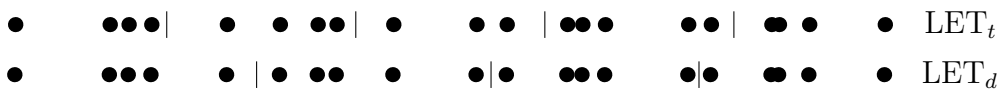


Figure 2.9: A visualisation of track-averaged LET and dose-averaged LET.

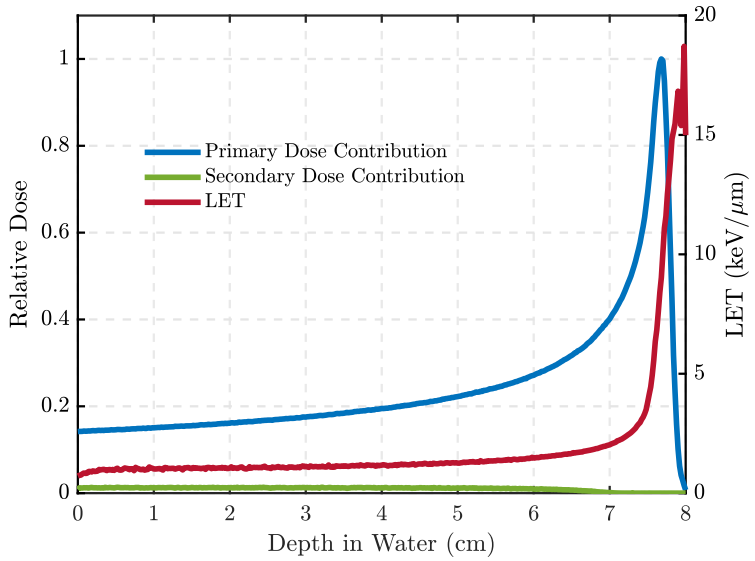


Figure 2.10: The percentage depth dose (PDD) distribution of a 100MeV proton beam propagating through water with dose contributions from primary/secondary particles and LET. The dose is normalised to the maximum absorbed primary dose.

Among other parameters, the LET of a particle is also known to influence the RBE of the beam.

Definition 2.2.2. *If we recall that the spatial distribution of events along a single track varies with particle type, it is reasonable to assume that the absorbed dose of radiation from one particle type will effect a biological target differently to the same absorbed dose delivered by a different particle. In the context of hadron therapy, 1Gy of absorbed dose delivered by a proton will have a far greater biological effect than the same dose delivered by a photon. This is how we define RBE,*

$$RBE[Endpoint] = \frac{Endpoint_{Reference}}{Endpoint_{Radiation}}, \tag{2.8}$$

where the reference radiation is almost always a photon or gamma-ray beam and the endpoint can be any effect or quantity dependent on the reference beam and the radiation type in question [25] (for example dose, survival and DNA damage characteristics can be used). A visual example of this calculation is shown in Fig.(2.11).

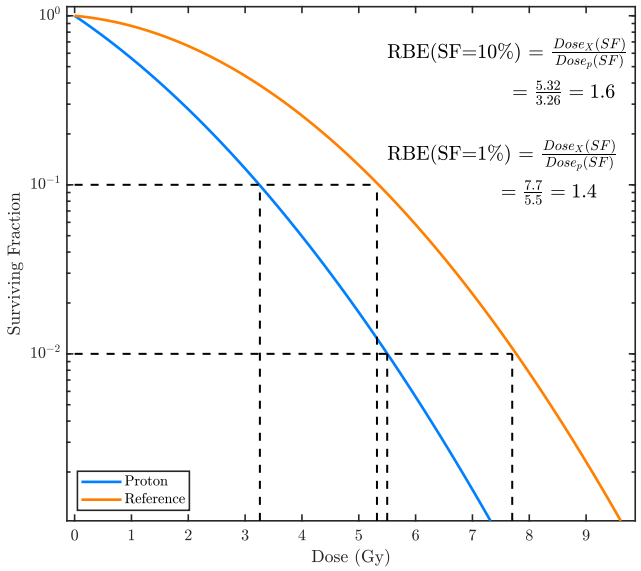


Figure 2.11: Visualisation of a proton RBE calculation.

In addition to LET, there are many other important parameters that also have an influence on the slope of the dose-response curve and hence the RBE [24].

- **Absorbed dose :** Shown in Fig.(2.11) is an example of how the RBE can change if it is measured at lower doses (a higher survival fraction) compared with higher doses (a lower survival fraction). In the high dose region, the reference beam loses cell killing efficiency faster than a proton beam. Therefore the difference between the doses required to achieve the same survival fraction is larger in this region than for low doses. The RBE is therefore higher in the low dose region when more efficient cell killing is achieved than for high doses.

- **Fractionated dose treatment :** In fractionated radiotherapy a given dose is delivered to the patient in multiple fractions over a given period of time where the cells are allowed to repair in between, resulting in a shallow survival curve and lower RBE.

- **Dose rate :** when the beam is delivered with a higher rate, the survival fraction will decrease at a faster rate resulting in a larger RBE.

- **Cellular conditions :** The biological stages and conditions in which the cells are kept can have a substantial effect on how they respond to radiation. Parameters such as cell cycle phase and oxygen content appear to play an important role [29,49] where the yield of indirect ionisation events from charged water radicals is dependent on the oxygen available for radiolysis reactions to take place. Therefore, when cells have a low oxygen content the RBE is expected to decrease because higher doses on radiation are required to achieve a given survival.

2.2.2 The Dose-Response Curve

Physical, chemical and biological processes discussed in the previous sections all contribute to the resultant shape of the dose-response curve. The dose-response curve is used by oncologists and researchers to describe the survival of a group of cells that are exposed to increasing doses of radiation - alternatively the curve describes the cell's sensitivity to different radiation types. Obtaining dose-response curves is extremely important, however generating them accurately is a challenging task.

The most common method to generate a dose-response curve experimentally is through cell culture techniques. Since the earliest radiobiological studies over 100 years ago, cell culture techniques have progressed from using yeast or bacteria to real mammalian and human cell lines which have allowed researchers to investigate the impact of a cell's biological conditions on their response to radiation [34, 60, 61].

Briefly, a typical cell culture experiment would see a group of cells suspended in a medium. After the cell concentration is counted, known numbers of cells are dispersed into petri dishes and exposed to a known dose of radiation. The cells are left for a period of time and allowed to grow into colonies. The surviving fraction is determined by counting the number of colonies produced (i.e. the number of cells that survived irradiation) and comparing with the number of cells plated. When this process is repeated with increasing dose, a dose-response curve is formed.

The curve itself is commonly displayed on a logarithmic scale where the change in behavior in the low dose region is "shouldered" before becoming more gradual in the high dose region, Fig.(2.12). For low-LET radiation, the ionisations occur sparsely, meaning the cell killing is more gradual with a shallow shoulder. As the LET increases, the curve decreases more rapidly, appearing linear on a logarithmic scale for all doses, Fig.(2.12).

The method with which the beam is delivered to the cells can influence the dose-response curve. Fractionated radiotherapy is the most common method of dose delivery over the single dose approach. When larger fractions of small doses are delivered with a given period of time in between, the cells are able to repair some of the damage induced. If we plot the resultant survival curve after several doses of fractionated radiotherapy, we would observe the initial "shoulder" region of the curve in Fig.(2.12) repeated with a slightly lower survival fraction after each fraction compared to the previous one. The resultant survival expression is presented in the following section, Eq.(2.14).

Cellular response is not only dependent on the radiation used but can also differ between cell lines. For example, some may be more radio-resistant than others, meaning that higher doses of a given radiation type are required to achieve a given survival fraction than a less radio-resistant cell line. By comparing the response

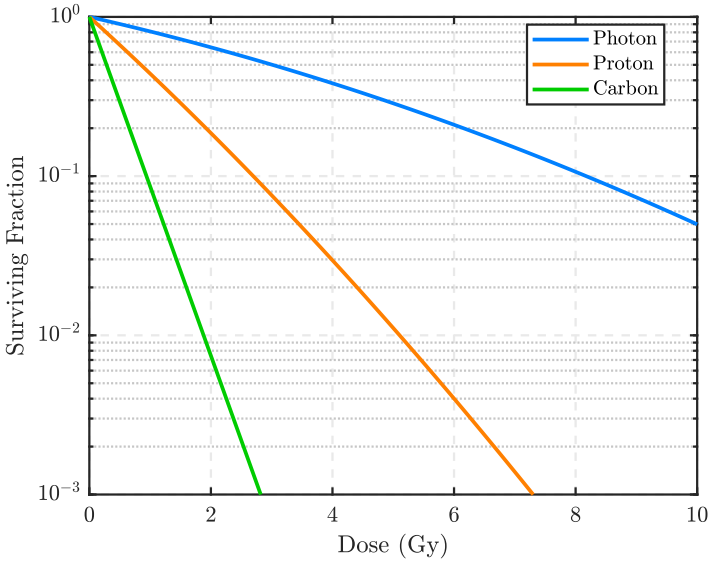


Figure 2.12: Cell survival curves for different radiation types.

of different cells under the same exposure conditions, inferences can also be made about conditions within the cell. For instance, we have already discussed the dependence of cell response on the cell cycle. A steeper decline in dose response may infer that more of the irradiated cells are in the M or G_2 phases of the cell cycle where they are most radio-sensitive. Conclusively, it is justified to suggest that there is no such thing as a single dose-response curve for a given set of parameters. Existing models that explain this behavior are discussed further below. The response should represent what is observed when the cells of a full tumour are irradiated. To achieve this, the beam should be generated such that it has reached the same energy after travelling through the healthy tissue and reaching the tumour.

2.2.3 Existing Cell Response Models

In Chapter 1 we briefly introduced some of the existing models in radiobiology. In this section we describe the evolution of radiobiological models introduced previously. We describe how early models were purely based on physics, before the introduction of biological repair models in the 1980's.

2.2.3.1 Target Models (1924)

Early radiobiological models were based on target theory and primarily proposed by physicists [34]. Target models are fundamentally based on three concepts [11, 62] :

1. We consider radiation to be a series of projectiles capable of hitting or missing a target.

2. The targets of radiation are the cells considered to be a single point in space.

3. When a target is hit, its response is stochastic in nature.

The first target theories were proposed by Crowther in 1924 [63] based on experimental results by Strangeways and Oakley in 1923 [64] with soft x-rays incident on chick embryo cells. Following the generation of the first survival curves in 1929 [64, 65] via exposure of bacteria to UV and photon radiation, Curie made the statement *"to destroy a bacillus it is necessary that its sensitive zone absorbs a minimal number of quanta"* [66]. Such findings allowed the single hit single target [34] and multiple hit single target models [67] to be derived.

These models have been introduced in the previous chapter but will be briefly reiterated here for consistency. The single hit single target model was proposed in 1946 [34] and the multiple hit single target model in 1967 [67]. Both models assume a Poisson probabilities that the cell targets experience k hits at time t,

$$P(k,t) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}. \tag{2.9}$$

We derive the survival probability according the number of targets that are not hit (we call this N) out of the total number of targets considered (call this N_0). The resultant survival probabilities are written as

$$P(0,t) = e^{-\lambda t} = \frac{N}{N_0} \tag{2.10}$$

where $\lambda = \frac{D}{D_0}$. The resulting survival probability in terms of dose is given by

$$S(D) = e^{-\frac{D}{D_0}} \tag{2.11}$$

in the single hit single target case and

$$S(D) = 1 - \left(1 - e^{-\frac{D}{D_0}}\right)^n \tag{2.12}$$

in the multiple hit single target case. However, such models failed to effectively explain the behavior in the case of radio-sensitive cells. In such cases, exponential behavior was observed which led to the proposal of the most utilised model in radiobiology to this day.

2.2.3.2 The Linear-Quadratic Model (1972)

The Linear-Quadratic (LQ) model which will be examined extensively in coming chapters, is a highly accepted, widely used model in the literature. It predicts the fraction of a group of cells that are likely to survive after being exposed to dose D of radiation based on the assumption that there are two components to cell killing. Single hits, described by parameter α [Gy^{-1}], are reminiscent of the single hit single

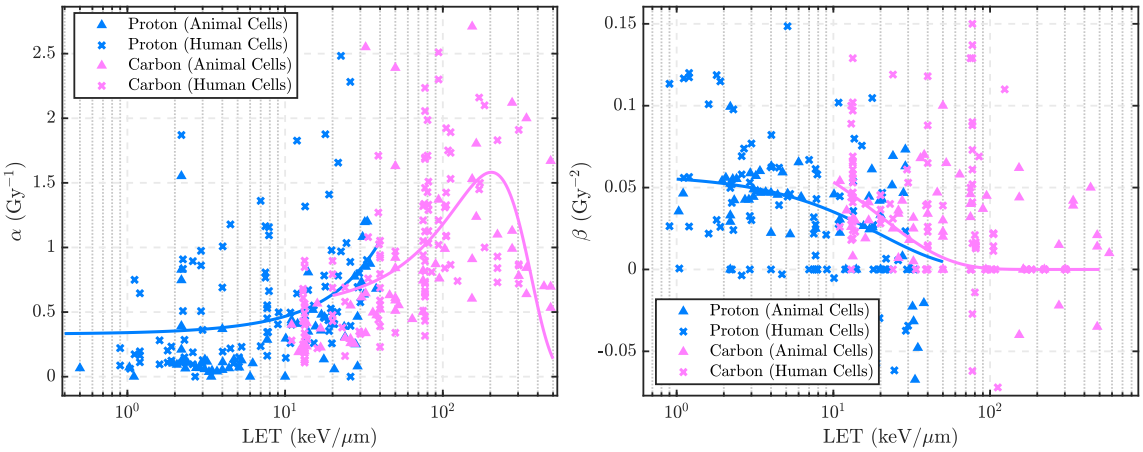


Figure 2.13: α and β parameters as a function of LET from fits to experimental data [46].

target model and multiple hits, described by parameter β [Gy^{-2}], reminiscent of multiple hit single target model. It is defined as

$$S = e^{-\alpha D - \beta D^2}. \tag{2.13}$$

Despite its simplicity it has maintained its status as the most used cell survival model in clinics and research until today. The model does not originate from a single source but is a amalgamation of different ideas developed within the same time frame [23, 35]. The foundations of the model were developed from radiation target theories described above. Likewise, the LQ model assumes that the hits to a target obey a Poisson distribution, Eq.(2.9). More on the statistical background of the LQ model will be presented in more detail in Chapter 5.

Information on the nature of the radiation incident on the targets can be derived from the LQ interpretation of the dose-response curve using parameters α and β. The efficiency of cell killing can be defined using the ratio α/β [Gy] - the dose at which the contributions of each type of hit to cell killing are equal. A low α/β ratio suggests that the rate of cell killing is relatively constant with dose whilst a high α/β ratio insists a more curved dose-response - this is where multiple hit events dominate. Revisiting our LET discussion from the previous section, the α/β ratio will increase with increasing LET, Fig.(2.13).

The LQ model can also be adapted to a fractionated radiotherapy treatment [11, 24] where a given dose D is separated into n adequately spaced treatments of dose d to allow the cells time to repair. This can be represented mathematically as

$$S = \left(e^{-\alpha d - \beta d^2}\right)^n = e^{-n(\alpha d + \beta d^2)} = e^{-D(\alpha + \beta d)}. \tag{2.14}$$

This method allows radiologists to tailor a patient's treatment as it is occurring to ensure optimal results.

Whilst the LQ model has remained very popular, its simplicity and relevance has been called into question. With more recent advances in radiobiology, models have been developed that account for fundamental biological processes that occur inside the cell and questions have been raised about whether the LQ model is globally valid, i.e. accurate regardless of cellular and radiation exposure conditions.

Comparisons with other radiobiological models (which will be discussed below) [20, 40] have shown that the LQ model does not perform as well as other models at reproducing specific cell survival data. This is attributed to the model only containing two parameters, even though the process of cell killing has been demonstrated to be far more complex. Other studies [19, 20] have shown that the LQ model loses its applicability at high doses where the β component to cell killing is dominant.

Another key parameter with a large influence on dose-response behavior, the LET, is a concept that has remained relatively unexplained despite a substantial contribution to the discrepancies in the LQ model [25, 30]. The idea of LQ applicability decreasing with increasing LET has previously been explored [18] and the idea of using a non-Poissonian distribution to describe the data has been put forth [68,69].

These discoveries have shown that whilst the LQ model is extremely useful at predicting dose-response behavior under certain parameters, its ability to reproduce experimental results is ineffective under some circumstances. This can be attributed to the fact that the LQ Model is not adequate at describing the complex processes that comprise radiation-induced cell death. Alternative models have since been proposed with the aim of improving on the LQ model and the selection of models in radiation therapy in general.

2.2.3.3 Repair-Misrepair (RMR) Model (1985)

The first notable improvement to the LQ model is the Repair-Misrepair Model (RMR) [31] which characterises the resultant DNA damages inflicted by radiation as being repairable with some probability. This probability is dependent on the concentration of specific repair enzymes available in the cell. These repair enzymes can also be deactivated by radiation. The result is a differential equation which describes the rate of linear repair and quadratic mis-repair (i.e. lethal lesions that contribute to cell death, U) inside of the cell according to coefficients λ and κ,

$$\frac{dU(t)}{dt} = -(\lambda + \kappa)U, \tag{2.15}$$

which leads to a survival equation containing a term to describe lethal lesions and repaired lesions. Assumptions worth noting in this model are that the distribution of lesions inside of the cell is Poissonian and for low doses the RMR model reduces to the LQ model.

2.2.3.4 Lethal-Potentially Lethal (LPL) Model (1986)

The next model to be discussed is the Lethal-Potentially-Lethal (LPL) Model [32]. This model assumes that there exists two types of DNA lesions, lethal and potentially-lethal. A lethal lesion is defined as a lesion that is irrepairable and will induce cell death. A potentially-lethal lesion is repairable and will do so at a rate ϵ_{PL}. The scenario of multiple lesions interacting (this becomes increasingly likely for high-LET radiation with dense ionisations about the particle track) is also accounted for and this interaction occurs at a rate ϵ_{2PL}, producing a lethal lesion. Again, this model assumes the distribution of lesions within a cell to be Poissonian. The result of this model is two differential equations that describe the rate of lethal (η_L) and potentially-lethal (η_{PL}) lesion development inside the cell during irradiation with dose rate $\dot{D}$. The equations are :

$$\frac{d\eta_L(t)}{dt} = \eta_L(t)\dot{D} + \epsilon_{2PL}\eta_{PL}^2(t), \tag{2.16}$$

and

$$\frac{d\eta_{PL}(t)}{dt} = \eta_{PL}(t)\dot{D} - \epsilon_{PL}\eta_{PL}(t) - \epsilon_{2PL}\eta_{PL}^2(t) \quad . \tag{2.17}$$

respectively. The survival fraction is estimated after the cells are allowed time to repair and after this time the remaining lesions of both types are now assumed to be lethal.

2.2.3.5 Microdosimetric-Kinetic Model (MKM) Model (1994)

The Microdosimetric-Kinetic Model (MKM) was proposed in 1994 [70] and adopts a formalism that the nucleus is a sensitive target divided into sub-volumes, each of mass m. Typically the sub-volumes are a series of shapes that fit together to form the nucleus which does not have to be uniform in shape.

After irradiation, a number of sub-volumes contain lesions which can be classified as Type I and II, alike to the LPL model. A Type I lesion is lethal and a Type II lesion is potentially-lethal meaning that it can either repair or become a Type I lesion. The number of lethal lesions present in the nucleus correlates with cell's probability of survival which assumes Poissonian distributed lesions, $S(D) = \exp(-L_f(D))$ where L_f is the number of lethal lesions in the cell.

The size of each sub-volume can influence the result substantially. In the high-LET domain, lesions are more likely to exist in close spatial proximity allowing them to interact. If a large sub-volume size is chosen and multiple lesions exist within the sub-volume, the model may mistakenly assume that they interact despite have a small chance of interaction in nature, meaning that the model can over-estimate the spatial proximity effect. Likewise, a small sub-volume size can classify lesions that are in close proximity separately where an interaction could occur in nature. In this case, the lethal lesion count can be under-estimated.

Further modifications to the MKM model have been developed to account for the stochastic nature of specific energies in the cell nucleus and its sub-volumes to predict the same cell response endpoints as the MKM model. The first iteration of the modifies MKM model is known as the stochastic microdosimetric model (SMK) [71]. This was closely followed by the double stochastic microdosimetric kinetic model (DSMK) which considers the mean and variance of the sub-volume specific energy, lowering computation time and memory requirements.

2.2.3.6 Local Effect Model (LEM) (1997)

The Local Effect Model (LEM) [44] adopts a similar formalism to the MKM model by approximating a cell as a series of sub-volumes. The expectation values of energy deposited in said sub-volume $\langle E \rangle$ is the only parameter that influences the cellular response within. The model achieves this by counting lethal events inside of each sub-volume where a single lethal lesion can induce cell death.

The average number of lethal events N_{lethal} in the nucleus is estimated by the local dose deposited in the sub-volume. The survival probability of the cell is then given assuming that the lethal events are Poisson distributed by $S(D) = \exp(-N_{lethal}(D))$.

The model does not involve cell specific parameters and is primarily used to predict cell survival. It has since been adapted in clinics across Europe [30].

2.2.3.7 Two Lesion Kinetic (TLK) Model (2001)

The Two-Lesion Kinetic Model (TLK) [33] is an amalgamation of the RMR and LPL models whilst accounts for the formation of DNA lesions with differing complexity, in addition to the repair and mis-repair of these lesions. The DNA lesions of varying complexity are the simple and complex double strand breaks (DSB). A simple DSB is defined as two DNA strand breaks existing within 20bp of each other whilst a complex DSB is defined as more than two DNA strand breaks existing within this same distance. The model is made up of four differential equations which describe the formation and repair/mis-repair of each DSB type (L_1 represents the number

of simple DSBs and L_2 the number of complex DSBs present), followed by the formation of lethal (L_f) and misrepaired (L_m) lesions. These equations are :

$$\frac{d\overline{L}_1(t)}{dt} = 2\dot{D}(t)Y\Sigma_1 - \lambda_1\overline{L}_1(t) + \eta\bar{L}_1(t)[\overline{L}_1(t) + \overline{L}_2(t)], \tag{2.18}$$

$$\frac{d\overline{L}_2(t)}{dt} = 2\dot{D}(t)Y\Sigma_2 - \lambda_2\overline{L}_2(t) + \eta\overline{L}_2(t)[\overline{L}_1(t) + \overline{L}_2(t)], \tag{2.19}$$

$$\frac{d\overline{L}_f(t)}{dt} = \beta_1\lambda_1\overline{L}_1(t) + \beta_2\lambda_2\overline{L}_2(t) + \gamma\eta[\bar{L}_1(t) + \overline{L}_2(t)]^2, \tag{2.20}$$

$$\frac{d\overline{L}_m(t)}{dt} = (1-\beta_1)\lambda_1\overline{L}_1(t) + (1-\beta_2)\lambda_2\overline{L}_2(t) + (1-\gamma)\eta[\overline{L}_1(t) + \overline{L}_2(t)]^2 \tag{2.21}$$

where β_1, β_2, λ_1, λ_2 and η are DSB rejoining parameters, $2Y\Sigma_{1,2}$ are DSB yield parameters and γ is a correction factor accounting for the formation of non-lethal chromosome aberrations.

The TLK model assumes the initial boundary conditions $\overline{L}_1(0) = \overline{L}_2(0) = \overline{L}_f(0) = \overline{L}_m(0) = 0$, meaning there are no damages in the cells prior to irradiation. The survival probability is estimated by taking the lethal lesions and applying them to $S = \exp(-L_f)$, assuming the Poissonian process. It has been shown that for low doses and dose rates, the TLK model can be approximated by the LQ model [72]. We will discuss this model in more detail when we apply it to a Monte Carlo simulation in Chapter 4.

Each of the models discussed above feature different methods of quantitatively describing dose response including the modelling of biological processes or the spatial mapping of energy depositions. However, the common thread between the models is they all rely on the Poisson distribution to compute survival, i.e. the fraction of cells that contain zero lethal lesions.

2.3 Conclusions

The models described in this section have some similarities in that they focus on several processes that are deemed to have the largest influence on a given endpoint (i.e. the number of lethal lesions, survival, etc.). The RMR, LPL and TLK models primarily focus on cellular response to radiation whilst the LEM and MKM models follow a more mechanistic approach to counting lethal lesions. The LQ model, whilst being a very useful model when fitting data to a dose-response curve due to its simplicity, appears to perform poorly when replicating experimental data. This is likely due to the lack of parameters involved, meaning that the model does not

adequately describe the system as a whole but only focuses on the number of "hits" to a single target.

Another noteworthy observation from this literature review is that nearly all radiobiological models developed in the last 100 years assume that lethal lesions obey a Poisson distribution. The result of this is that all commonly accepted radiobiological models that explain how lethal lesions are produced within a cell in the literature are based on the same statistical foundation.

The key conclusions made from this chapter are :

- an effective radiobiological model will incorporate as many of the parameters described in Section (2.1) as possible whilst optimising computation time, memory requirements and model simplicity.
- The modelling of LET behavior is still a relatively unexplored concept.
- The small parameter space of the LQ model "constricts" the accurate prediction of cell survival. It is therefore difficult to predict exactly how a group of cells will react to radiation because important cell response parameters are neglected.
- There exists hundreds of parameter combinations related to the cell environment and exposure conditions, meaning that care must be taken when comparing experimental results across different studies.
- The model used in this thesis should be rigorously tested against experimental data to ensure valid results and avoid unnecessary trade-off between computation time and accuracy of the results.
- In hadron therapy, radiation is often high-LET by nature. Considering evidence that the LQ model is less effective at high LET and dose [18, 20], its range of applicability may be limited in the case of hadron therapy.
- A good understanding of MC methods is required to achieve a good balance between simulation optimisation and achieving accurate results.

3

Background on Monte Carlo Methods for Particle Transport

Monte Carlo (MC) modelling is prevalent in not only particle therapy but areas of high energy physics and astrophysics. It has increased in popularity since its development in the 1970s [73]. Early MC codes were limited due to computing power available at the time, however the progression of high-performance and parallel computing (HPC) has allowed significant discoveries to be made across many fields including the sciences, engineering and finance.

Generally, MC methods are defined as the random sampling of a known probability distribution (or the selection of a uniform random number) as an input for a deterministic process[1]. The deterministic result is therefore made to replicate a stochastic system due to the randomness of the input parameter [74]. The results of such calculations can be analysed to estimate behaviors in complex systems.

In this chapter, an overview of hadron therapy MC techniques and high-performance computing is given. The software toolkits GEANT4 (Geometry and Tracking) and TOPAS (Tool for Particle Simulation), used extensively to conduct proton therapy simulations in this thesis, are introduced and their framework is described. Techniques of simulation optimisation in each toolkit are explained for our purpose of using in Chapter 4.

3.1 An Introduction to Monte Carlo Theory

We define the MC method by which stochastic processes are modelled by random sampling. If one wishes to estimate the frequency $f(x)$ of an observation, they could take N samples of x_i to count how often f_i occurs. From this one can determine the first and second moments,

[1] A process that will produce the same result with no randomness when the same input is provided.

$$\langle f \rangle = \frac{1}{N} \sum_{i=1}^{N} f_i \tag{3.1}$$

and

$$\langle f^2 \rangle = \frac{1}{N} \sum_{i=1}^{N} f_i^2. \tag{3.2}$$

The Central Limit Theorem gives the probability distribution for the expectation value $\langle f \rangle$,

$$p\big(\langle f \rangle\big) = \frac{\exp\big[-(\langle f \rangle - \bar{f})^2/2\sigma^2\big]}{\sqrt{2\pi}\sigma}, \tag{3.3}$$

where the variance in $\langle f \rangle$ is

$$\sigma^2 = \frac{\langle f^2 \rangle - \langle f \rangle^2}{N-1}. \tag{3.4}$$

That is to say that $\langle f \rangle$ will approach the true value $\bar{f}$ for large N [74].

3.2 Simulating Particle Transport

Using MC methods for particle transport is far more complex than the basic definition provided above but the basic premise remains the same. From Chapter 2, a particle propagating through a medium can undergo a series of interactions which can occur with different probabilities.

As an example, let us consider particle p traversing an arbitrary medium M, as shown in Fig.(3.1). Particle p can undergo three different processes, absorption, elastic scattering and inelastic scattering, defined with cross-sections Σ_a, Σ_e and Σ_i, respectively. These cross-sections can vary with energy and the medium the particle is traversing. Therefore the relative probability for each interaction type is denoted Σ_a/Σ_t, Σ_e/Σ_t and Σ_i/Σ_t, where $\Sigma_t = \Sigma_a + \Sigma_e + \Sigma_i$.

The particle begins at position x_0 and is assigned an energy E_0. By randomly sampling a probability distribution function (pdf), p is displaced by some distance Δx where it will undergo one of the three interactions. Before choosing an interaction type, it is worth checking if the particle is still inside medium M, because when it leaves M it is not worth tracking anymore as it would be a waste of computation time and memory. The interaction type is determined using a random number generator with the result dependent on the probabilities of each interaction occurring. If the interaction type is absorption, the particle no longer needs to be tracked. If the interaction is a scattering process, p can be assigned a new energy $E_1 \leq E_0$ and

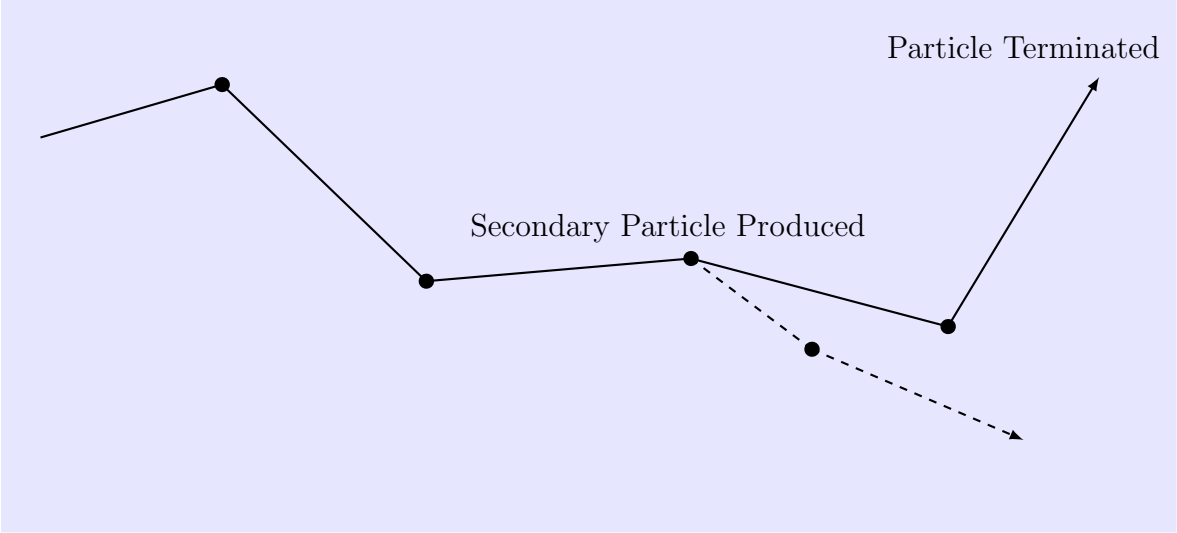


Figure 3.1: A single particle p traversing a medium M.

displaced by a new distance Δx determined by random sampling of the pdf. The process will repeat until the particle either leaves M or undergoes absorption.

This example may seem simple conceptually, but the situation becomes far more complex when simulating a "real world" situation with more interaction types and millions of events occurring. In some situations not all of these interactions can be simulated realistically, even with the computational advances made in recent years. The ability to simulate millions of events is important in reaching a satisfactory result of the simulation. This is because the uncertainty in a MC estimated observation is proportional to $1/\sqrt{N}$. In most applications it is not only the events themselves that need to be tracked, but information about the nature of each event must be derived as well (e.g. absorbed dose, energy deposited and particle fluence). This complicates the situation further. There are methods typically applied by MC users to find the best estimate of their result while optimising processing times and memory (i.e. N should be large enough to minimise the uncertainty while keeping the processing time as low as possible).

Light particles such as photons can be accurately simulated using the above method due to their sparsely ionising nature. Charged particles including protons, however, ionise the medium they are propagating through densely so simulating every step is not feasible computationally. In such situations, simplifications must be made to condense the simulation so that it can be managed computationally. For example, the simulations performed in this work, using the model described in Chapter 4, required 15 CPU cores with between 1 and 3 days of processing time and 12GB of memory on average.

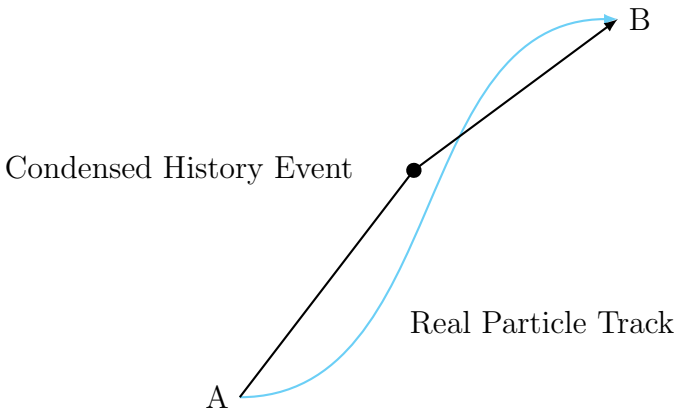


Figure 3.2: A particle traversing from A to B with the corresponding condensed history track.

As discussed in Chapter 2, protons are heavy particles, meaning that they typically do not scatter at large angles or lose large amounts of energy in a single interaction. This is an ideal situation to use condensed history (CH) particle transport. The basic premise of CH transport is that a series of very small steps where the chance of scattering or large energy changes are minimal can be condensed into a single large step event with little effect on the final result, Fig.(3.2).

3.3 Geant4 and TOPAS Monte Carlo Software Toolkits

Geant4 [75] is an open-source (i.e. all aspects of the software are accessible) software toolkit written in the computer language C++ which is capable of simulating the transport and interactions of particles propagating through matter. It is the combined effort of a worldwide collaboration. Originally developed by CERN, the toolkit's initial purpose was to simulate high-energy physics inside particle accelerators, such as the Large Hadron Collider (LHC). It soon found its way into other areas of physics, including space radiation and medical physics [75]. It is updated every year and the results it produces have been rigorously validated [76–78]. The popularity of Geant4 in medical physics has increased with time, becoming one of the most used toolkits for publications when compared with other particle transport codes such as FLUKA [79], MCDS [80] and PENELOPE [81].

TOPAS [47] is a MC toolkit that acts as a "wrapper" for the Geant4 toolkit and is specifically designed for medical physics and radiobiology applications. It is again written in C++ but the user is able to control all aspects of a simulation using a single file known as a *parameter file*. It is able to model all of the processes that Geant4 can, but the usability makes it simple to control complex simulations

common in medical physics.

The TOPAS MC toolkit (version 3.2.p2) together with the interaction processes of Geant4 (version 10.5.p01) were used for all simulations undertaken in this thesis. In coming chapters these simulations are described in more detail, whilst this section will focus on the framework of the toolkits themselves.

3.3.1 Code Framework

A Geant4 simulation is made up of a series of C++ classes containing functions - we call this the kernel [82] (Note: all terms written in red are Geant4 parameters and are defined below). We start by defining a geometry to contain all components of the simulation, known as the "**world**". Then we can introduce **physics lists**, **particle generators**, the **detector construction**, the **primary generator action**; and finally the "**run**", "**event**", "**tracking**" and "**step**" managers.

- **World**: the geometry in which all components of the simulation are contained.
- **Physics List**: a series of lists containing some physical processes. More than one of these can be called in a single simulation. Examples of the physics lists available in Geant4 are the Livermore and Penelope low energy lists, hadronic physics lists and Geant4-DNA list. Each is best suited to specific applications and implements ionisation, excitation and charge transfer interactions in the case of low energy physics.
- **Detector Construction**: the geometry in which particles will be tracked.
- **Particle Generator**: the source from which primary particles are generated. Available particles include protons, heavy ions, pions, electrons and photons among many others.
- **Run**: when one or more primary particle(s) are produced and tracked.
- **Event**: follows the production of a primary particle from when it is generated until it is terminated. This includes any secondary products generated.
- **Track**: the path followed by a single particle whether it is primary or secondary.
- **Step**: the snapshot of a particle traveling from one point to another, along with any changes to energy, momenta and direction that occurred between the two points.

The simplest of Geant4 simulations must contain the first four items listed above. Tracking customisation is optional but needed in most situations. Figure (3.3) depicts a visualisation of how Geant4 defines a track structure.

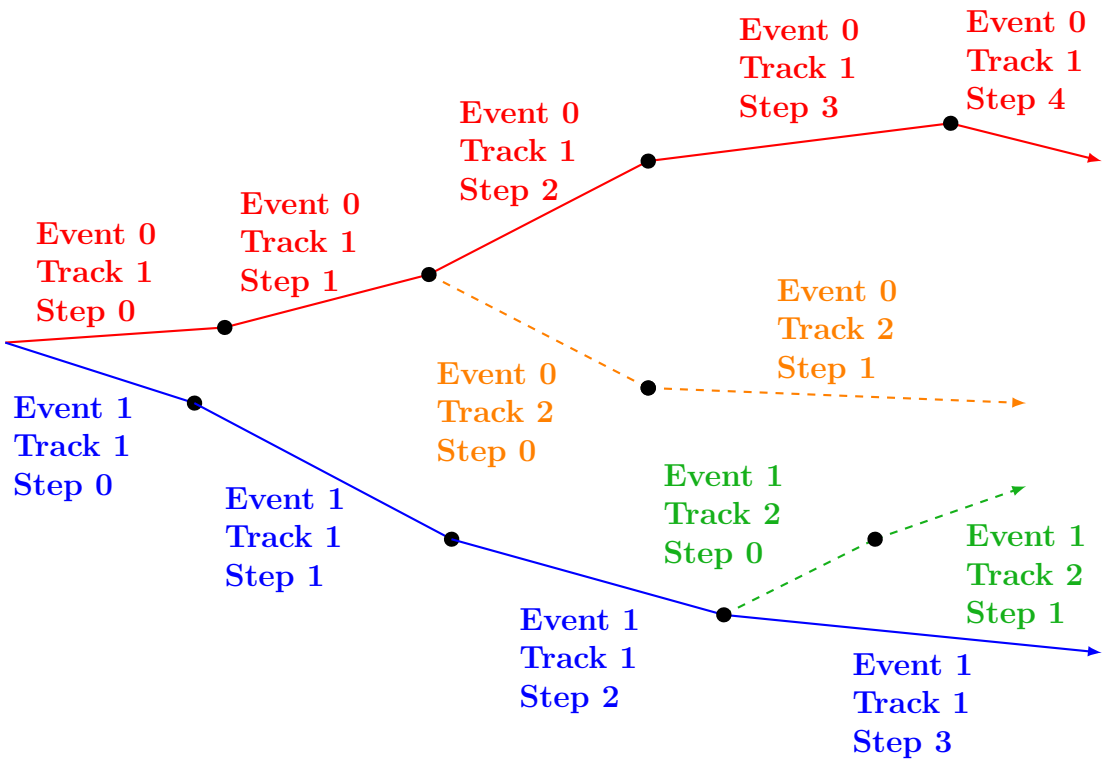


Figure 3.3: A Geant4 track structure of a single run.

Of course in a "real-world" situation, a particle does not propagate through a medium in steps. Steps are a method of optimisation due to the computing power required to track every event interaction. The **step size** can be defined in a Geant4 simulation but should be chosen carefully depending on the application. In radiobiology simulations, a small step size is a necessity. However in applications where the track structure is not important, the step size can be increased, resulting in shorter processing times and lower memory requirements.

Other optimisation methods in Geant4 are the introduction of **production cuts** to both particle range and energy. A **range cut** is defined as the minimum distance at which a particle will be tracked. This means that any particle produced with an estimated range less than the range cut will be terminated at the point that it is produced. Similarly, an energy cut will terminate a particle at the point it reaches an energy less than the cut value.

The choice of **physics list** can also affect processing time and memory requirements. Some **physics lists** are suited for specific applications. For example, the Geant4-DNA physics list is most suited to nano-dosimetry applications, whilst the standard physics lists are best suited for macro-scale simulations.

Geant4 also has the capability to run in multi-threaded mode. To initiate this, the command `-DGEANT4_BUILD_MULTITHREADED=ON` must be included when compiling the Geant4 program. This option will allow the central processing unit (CPU) to distribute the events across a number of threads, allowing them to run concurrently. Distributing these events in such a way will reduce simulation run time and memory requirements, allowing for a more optimised simulation.

All of these features of Geant4 can also be exploited in TOPAS, only they are altered inside a **parameter file** (Note: all TOPAS parameter names in blue are defined below). All aspects of the simulation can be controlled in one (or more) of these files. Additional **parameter files** can be used by calling them in the main **parameter file**. Each line in one of these files defines some parameter related to **geometry**, **particle generators**, **scoring**, **visualisation** and **time features** [83].

- **Parameter File**: a text file containing or directing to all parameters in the simulation.
- **Geometry**: all geometries contained in the simulation **world** along with their material compositions.
- **Particle Generators**: the source from which the particles are generated (similar to Geant4).
- **Scoring**: taking information such as particle fluence, dose and energy from each step in the simulation.
- **Visualisation**: self-explanatory - how the simulation is visualised. This should be turned off for large scale simulations because processing times are much higher when used with a large number of events.
- **Time Features**: needed if a time component is to be introduced into the simulation.

The lines in a parameter file must follow a specific format if it is to be correctly read by TOPAS [47]. These commands follow the same general format,

```
Parameter_Type:Parameter_Name = Parameter_Value [Unit (if required)].
```

The first thing to define in a TOPAS parameter is the type (s = string, b = boolean, i = integer, d = double and dv/iv = vector of the same parameter type), the second is the class that the parameter belongs to (Ge/ = geometry, Ph/ = physics, Sc/ = scorer, So/ = source, Tf/ = time features), the third is the name of the component (multiple geometries, sources, etc can be defined) and the final part is the parameter to be changed (the dimension, position, energy, etc). Certain parameters are also expected to carry units (dimensions and energies for example). Note that when defining a vector, the length of the vector must precede the entries themselves,

2212	293.13	-148.93	-1093.63	proton	proton_G4DNAIonisation	MyCulture/Nucleus0	0	0.01339
2212	293.131	-148.931	-1093.64	proton	proton_G4DNAIonisation	MyCulture/Nucleus0	0	0.01605
2212	293.131	-148.931	-1093.65	proton	proton_G4DNAIonisation	MyCulture/Nucleus0	0	0.01339
2212	293.132	-148.932	-1093.66	proton	proton_G4DNAIonisation	MyCulture/Nucleus0	0	0.01605
2212	293.132	-148.933	-1093.67	proton	proton_G4DNAIonisation	MyCulture/Nucleus0	0	0.01079
2212	293.132	-148.933	-1093.68	proton	proton_G4DNAExcitation	MyCulture/Nucleus0	0	0.01377
2212	293.133	-148.933	-1093.68	proton	proton_G4DNAIonisation	MyCulture/Nucleus0	0	0.01079
2212	293.133	-148.934	-1093.69	proton	proton_G4DNAIonisation	MyCulture/Nucleus0	0	0.01079
2212	293.133	-148.934	-1093.7	proton	proton_G4DNAIonisation	MyCulture/Nucleus0	0	0.01079
2212	293.134	-148.935	-1093.71	proton	proton_G4DNAExcitation	MyCulture/Nucleus0	0	0.01124
2212	293.134	-148.936	-1093.72	proton	proton_G4DNAIonisation	MyCulture/Nucleus0	0	0.01079
2212	293.135	-148.936	-1093.72	proton	proton_G4DNAIonisation	MyCulture/Nucleus0	0	0.01079
2212	293.136	-148.937	-1093.75	proton	proton_G4DNAIonisation	MyCulture/Nucleus0	0	0.01339
2212	293.136	-148.937	-1093.75	proton	proton_G4DNAIonisation	MyCulture/Nucleus0	0	0.01339
2212	293.136	-148.938	-1093.76	proton	proton_G4DNAIonisation	MyCulture/Nucleus0	0	0.0323
2212	293.137	-148.939	-1093.78	proton	proton_G4DNAExcitation	MyCulture/Nucleus0	0	0.01377
2212	293.138	-148.94	-1093.8	proton	proton_G4DNAIonisation	MyCulture/Nucleus0	0	0.01605

Figure 3.4: Example output from a TOPAS simulation with the particle ID number (red), event coordinates (blue), process and particle name (green), organelle name (purple) and energy deposited (orange).

for example a vector of four lengths must be written as `dv:Ge/Vector_Of_Lengths = 4 10.12 48.73 79.60 121.98 cm`.

Each parameter in TOPAS is customisable. **Geometries** can be built using simple, pre-defined shapes. Their dimension, rotation and position can be altered and they can be filled with materials defined in Geant4 or customised by the user. **Particle sources** can be customised by energy distribution (mean and spread), particle type(s), beam shape, position and scattering angles. Multiple **physics lists** can be defined in a single simulation. They can also be assigned to specific geometries or regions in a simulation. This method can be used for simulation optimisation. Simulation biasing is another optimisation technique where the production of specific particles is neglected, however care should be taken with this to ensure erroneous results are not generated.

Similar to Geant4, TOPAS can output a series of parameters including absorbed dose, deposited energy, event coordinates and the particle involved in the event, among many others, Fig.(3.4).

3.3.2 Geant4-DNA and TOPAS-nBio - Extensions for Low Energy Physics

Geant4 provides multiple physics lists suited to specific applications. The "standard" collection of physics lists are the most generalised and suited to macro-scale applications. The application type that is the focus of this thesis requires a very detailed account of the track structure down to the nano-scale. For this reason, the standard physics lists are not suited to the aims of the model developed in Chapter 4. For such applications, the low energy physics lists are the best option. These include the Livermore, Penelope and Geant4-DNA physics lists. Each of these physics lists cover a specific set of processes over different energy ranges (Tables (3.3) - (3.5)).

The Geant-DNA physics list is not only capable of simulating physical processes on the nano-scale level but it is able to simulate chemical processes using Geant4-DNA chemistry (specifically the radiolysis of water molecules). From Chapter 2, radiolysis reactions make a substantial contribution to indirect DNA damage and therefore could be useful in this thesis. One noteworthy disadvantage of the Geant4-DNA physics list is that it is currently available for liquid water phantoms only.

The Geant4-DNA chemistry extension uses a timeline similar to that described in Chapter 2 - meaning that it occurs over a user-defined virtual time $t \leq 1\mu s$ [83]. The chemical species are produced in user-defined time steps and their motion is dictated by random Brownian motion and diffuse according to the defined diffusion coefficients unique to the species (Table (3.1)). One of the defined reactions between chemical species will occur if they are within a given radius determined by the reaction rates in Table (3.2). Similar to the physical processes, appropriately large time steps must be chosen for optimisation purposes while having minimal impact on the results.

TOPAS-nBio is an extension of the TOPAS MC software toolkit [84] with a focus on radiobiological (nano-scale) simulations. It allows researchers to perform simulations using single or multi-cellular networks, in addition to sub-cellular networks (neuron and DNA-based simulations). It utilises the same framework as Geant4-DNA whilst providing more specialised geometries aimed at radiobiological research.

Species	**Diffusion Coefficient** ($10^{-9}\mathbf{m^2s^{-1}}$)
e^-_{aq}	4.9
$^\bullet OH$	2.2
$H^\bullet$	7.0
H_3O^+	9.46
H_2	4.8
OH^-	5.3
H_2O_2	2.3

Table 3.1: Geant4-DNA Chemistry Diffusion Coefficients [78]

Reaction	**Reaction Rate** ($10^7\mathbf{m^3mol^{-1}s^{-1}}$)
$H^\bullet + e^- + H_2O \longrightarrow e^-_{aq}$	2.65
$H^\bullet + {}^\bullet OH \longrightarrow H_2O$	1.44
$H^\bullet + H^\bullet \longrightarrow H_2$	1.20
$H_2O_2 + e^-_{aq} \longrightarrow H^\bullet + H_2O$	1.41
$H_3O^+ + e^-_{aq} \longrightarrow H^\bullet + H_2O$	2.11
$H_3O^+ + OH^- \longrightarrow 2\,H_2O$	14.3
$^\bullet OH + e^-_{aq} \longrightarrow OH^-$	2.95
$^\bullet OH + {}^\bullet OH \longrightarrow H_2O_2$	0.44
$e^-_{aq} + e^-_{aq} + 2\,H_2O \longrightarrow 2\,OH^- + H_2$	0.50

Table 3.2: Geant4-DNA Chemistry Reactions [78].

Particle	Physics Process	Process Class	Low Energy Limit	High Energy Limit
Gamma	Compton	G4ComptonScattering	250 eV	100 GeV
γ	Rayleigh	G4RayleighScattering	250 eV	100 GeV
	Conversion	G4GammaConversion	1.022 MeV	100 GeV
	Photo-electric	G4PhotoElectricEffect	250 eV	100 GeV
Electron	Ionisation	G4eIonisation	250 eV	100 GeV
e^-	Bremsstrahlung	G4eBremsstrahlung	250 eV	100 GeV

Table 3.3: Livermore Physics List Processes [85].

Particle	Physics Process	Process Class	Low Energy Limit	High Energy Limit
Gamma	Compton	G4ComptonScattering	250 eV	1 GeV
γ	Rayleigh	G4RayleighScattering	250 eV	1 GeV
	Conversion	G4GammaConversion	1.022 MeV	1 GeV
	Photo-electric	G4PhotoElectricEffect	250 eV	1 GeV
Electron/	Ionisation	G4eIonisation	250 eV	1 GeV
Positron	Bremsstrahlung	G4eBremsstrahlung	250 eV	100 GeV
e^-/e^+	Annihilation	G4eplusAnnihilation	250 eV	1 GeV

Table 3.4: Penelope Physics List Processes [85].

Particle	Physics Process	Process Class	Low Energy Limit	High Energy Limit
Proton	Nuclear Scattering	G4DNAElastic	100 eV	1 MeV
p	Electronic Excitation	G4DNAExcitation	10 eV	100 MeV
	Ionisation	G4DNAIonisation	0 eV	100 MeV
	Electron Capture	G4DNAChargeDecrease	100 eV	100 MeV
Electron	Elastic Scattering	G4DNAElastic	0 eV	1 MeV
e^-	Electronic Excitation	G4DNAExcitation	8 eV	1 MeV
	Ionisation	G4DNAIonisation	10 eV	1 MeV
	Vibrational Excitation	G4DNAVibExcitation	2 eV	100 eV
	Attachment	G4DNAAttachment	4 eV	13 eV
Gamma γ	Same as Livermore Model (Table (3.3))			

Table 3.5: Geant4-DNA Physics List Processes [86]

3.4 Conclusions

This chapter provides a solid background on MC theory with a description of how it is applied to particle transport simulations and the software toolkits currently available. The software used in this thesis are described in detail including the code framework, optimisation methods and capabilities.

The quality of results and optimisation trade-off effect was highlighted and areas where this balance can be improved are discussed. The key conclusions made from this discussion are :

- simulating particle transport with MC simulations can become computationally expensive with millions of events being processed simultaneously.
- optimisation methods including step sizes, production cuts and multi-threading capabilities should be explored and tested for a speed up of processing time.
- any optimisation performed on the simulation should be compared with the corresponding un-optimised version to ensure accurate results are still given.
- the geometry and physics sections of the simulation are the most memory intensive and should be the focus of further optimisation studies.

4

Developing a Monte Carlo Simulation of Radiation-Induced Cell Death

In Chapter 2 the question of what makes an effective radiobiological model is addressed by reviewing the intrinsic physical, chemical and biological aspects that influence cell irradiation and death. Current radiobiological models applicable to hadron therapy which employ these concepts to estimate biological endpoints including the survival fraction, DNA damage yield and chromosome aberrations are also discussed. To simulate the dose-response of a tumour computationally, MC simulation methods for particle transport were reviewed and methods for simulating detailed track structures with optimisation were investigated in the previous chapter.

In the current chapter the conclusions developed in Chapters 2 and 3 are used to design a Monte Carlo experiment to accurately formulate dose-response data for proton radiation. The model used in this thesis is based on previously developed models [48, 49, 87]. The chapter is organised as follows :

1. cell geometry and composition,
2. cell irradiation parameters and beam delivery methods,
3. the characterisation of DNA damage and how it is quantified,
4. the repair processes simulated inside of the cells post-irradiation,
5. cell survival calculation
6. and how the simulations are undertaken using HPC methods.

We begin by providing an overview of what is achieved in the model and how the model imitates a cell culture experiment, including geometry components; how

cell irradiation and repair is simulated; the output information that is derived from the simulations and the computational requirements, Figs.(4.1),(4.2). A discussion of how the cell geometry is implemented into the simulation and insight into the choice of cell dimensions and composition follows. Next the process of cell irradiation is described with insight into the Geant4 physics models that were chosen, how exposure conditions such as energy or LET are varied and how the beam was delivered is also described. An outline of the physical processes simulated inside the cells is provided. The methods used to characterise DNA damage of varying complexity and how the repair process was simulated is described next. We then discuss how the fate of a cell is determined (i.e. survival or death). We conclude the chapter by describing how the simulations are performed on the Phoenix and HPC1 supercomputing clusters (University of Adelaide, Adelaide, South Australia). All simulation parameters used to generate our results are given in Chapters 7 and 8.

4.1 Simulation Schematic

The primary endpoint of the MC simulation is to calculate the survival fraction as a function of dose and repair time, and in return RBE. This is achieved using a number of stages, as outlined in Fig.(4.3), consisting of :

- a cell placement and irradiation stage (performed with TOPAS-nBio [84] cell culture geometry),
- reducing the output from TOPAS to the relevant ionisations that contribute to DNA damage (performed using Bash Scripting),
- characterising and quantifying the DNA damage inside each cell (performed using MATLAB),
- the simulation of repair processes inside the cells,
- determining the fate of each cell (survival or death)
- the modelling of each cell's probability of survival at various repair times performed using a modified version of the MATLAB script developed in [48,87].

To observe the repair dynamics in the cell adequately, the fraction of cells surviving was measured up to 40 hours post-irradiation at 15 second (for $0 \leq t \leq 60$ seconds), 60 second (for 60 seconds $< t \leq 2$ hours) and two hour (for $2 \leq t \leq 40$ hours) intervals.

As a brief overview, each simulation consists of :

- a "world" filled with air,
- a cellular phantom (i.e. a box of randomly placed cells),

- a growth medium (water),
- a scorer (for recording particle track information),
- and a particle source.

The exact radiation exposure and geometry conditions implemented in TOPAS are discussed with more detail later in this chapter. To ensure that an adequate statistical analysis of the results can be performed, every simulation was assigned a different random seed to ensure statistical independence between results and each measurement was repeated multiple times for every dose point. In TOPAS, random seeds are implemented into a simulation using the command `i:Ts/Seed = n` where n is an integer in the range 0 to 2147483647.

At the end of the cell irradiation stage, TOPAS outputs a series of "pTuple" (in text format) and comma-separated value (CSV) files containing

- the xyz coordinates of the N randomly placed cells ($N = 530$ in this study),
- the xyz coordinates of every physical process undergone by all primary and secondary particles,
- the name of the physical process undergone at every coordinate (i.e. ionisation and excitation),
- the name of the organelle in which each event occurred,
- the energy deposited at the given xyz coordinates.

The TOPAS-nBio cell culture geometry source code originally did not provide the positions of the randomly placed cells. To implement this, the corresponding source code files were edited to output this information (the relevant changes made to the source code are listed in red in Appendix A). A "pTuple" scorer was included in the simulation parameter file to score the track information listed above. By default, the pTuple scorer in TOPAS takes a snapshot of the ionisations in the geometry up to a given time [1]. This was not suitable for the needs of the current model, where the track information must be recorded for the entirety of the radiation stage.

The remainder of the workflow involved post-processing of the TOPAS output using bash scripting commands and MATLAB scripts.

- The dose absorbed (Gy) by the cells was calculated

$$D = \sum_{i=1}^{n_{cells}} \frac{E_i}{m_i}, \tag{4.1}$$

[1] The default is 1μs.

where E_i is the energy deposited in the i^{th} cell (Joules) and m_i is the mass of the i^{th} cell (kg).

- The ionisation events (i.e. the events where the energy deposited is ≥ 10.79eV) are separated from the "pTuple" output file. A minimum energy deposition of 10.79eV was chosen because this is the ionisation threshold of liquid water [88], which makes up the cellular composition in the simulation (discussed with more detail in the following section).
- The number of ionisations are reduced according to a direct ionisation efficiency factor of 35%, the origin of which will be explained later in this chapter.
- The characterisation and quantification of DNA damage is performed using a series of MATLAB scripts, where the ionisations are reduced, clustered and used to determine the probability of survival for each cell in the simulation. This is discussed in more detail later in this chapter.

After all the simulation results are collected, the "pTuple" data file is compressed and the yield of each DNA damage type, LET (keV/μm) and absorbed dose (Gy) are written to a CSV file. Due to the detailed track structure simulated, the text file containing the track information ranged from 60 to 100GB in size and the reduced ionisation text file ranged from 2.5 to 25GB in size. This meant that each simulation had large memory, run time and storage requirements.

Ideally, we would have simulated indirect events using the Geant4-DNA Chemistry model because chemistry plays an important role in cell death, especially in the low-LET region [11]. However, the processing time and memory requirements placed on each simulation with only the direct effects implemented were already large and the chemical processes substantially increased these requirements. As a result, we could not implement indirect effects due to time and computation power constraints. Regardless, we explain how they would have been implemented into our simulation if they had not needed to be omitted. We intend to include the chemical processes in our future work.

These simulations were carried out on the Phoenix and HPC1 supercomputing clusters (University of Adelaide, Adelaide, Australia). The memory requirements and processing time increased exponentially for larger doses and varied according to particle LET. High-LET beams required an average of 15 CPU cores with a processing time ranging from 24 to 38 hours. Low-LET beams required $>$ 45 hours on average with the same number of CPU cores. This is because more particle tracks are required per Gy of low-LET radiation, which was in the range of 10^6 per Gy. High-LET radiation simulations required approximately 10^5 particles per Gy.

To ensure adequate statistics, the dose-response curves were formulated by measuring eight dose points, with 20 to 30 repetitions each. Using the processing requirement metrics above, each dose-response curve required 10^4 core-hours of processing

time. Two data sets were collected, one with the cells at zero depth in water and energy-varied LET and one with depth-varied LET (i.e. the position in the Bragg Peak is varied). Repeating the simulations substantially increased the storage requirements. The Intersect external digital storage system was used to store the data.

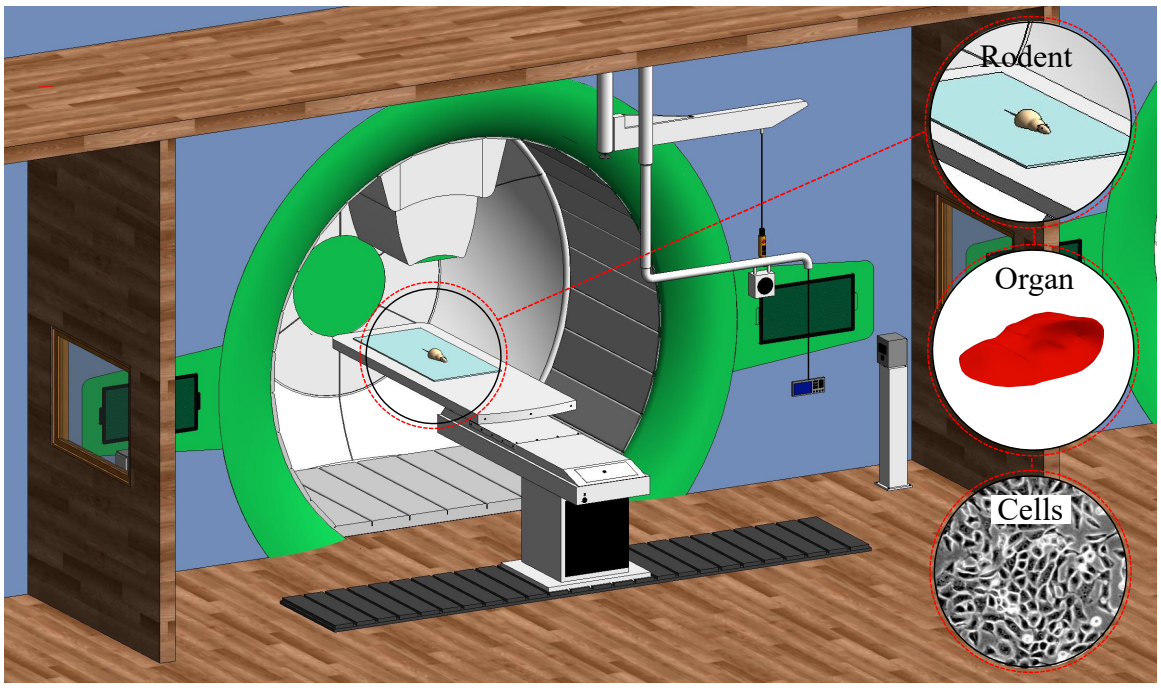


Figure 4.1: An experimental schematic of our simulations on V79 Chinese Hamster cells (Credit: Adamar Design).

Figure 4.2: A video of the simulation schematic shown in Fig.(4.1) (Credit: Adamar Design).

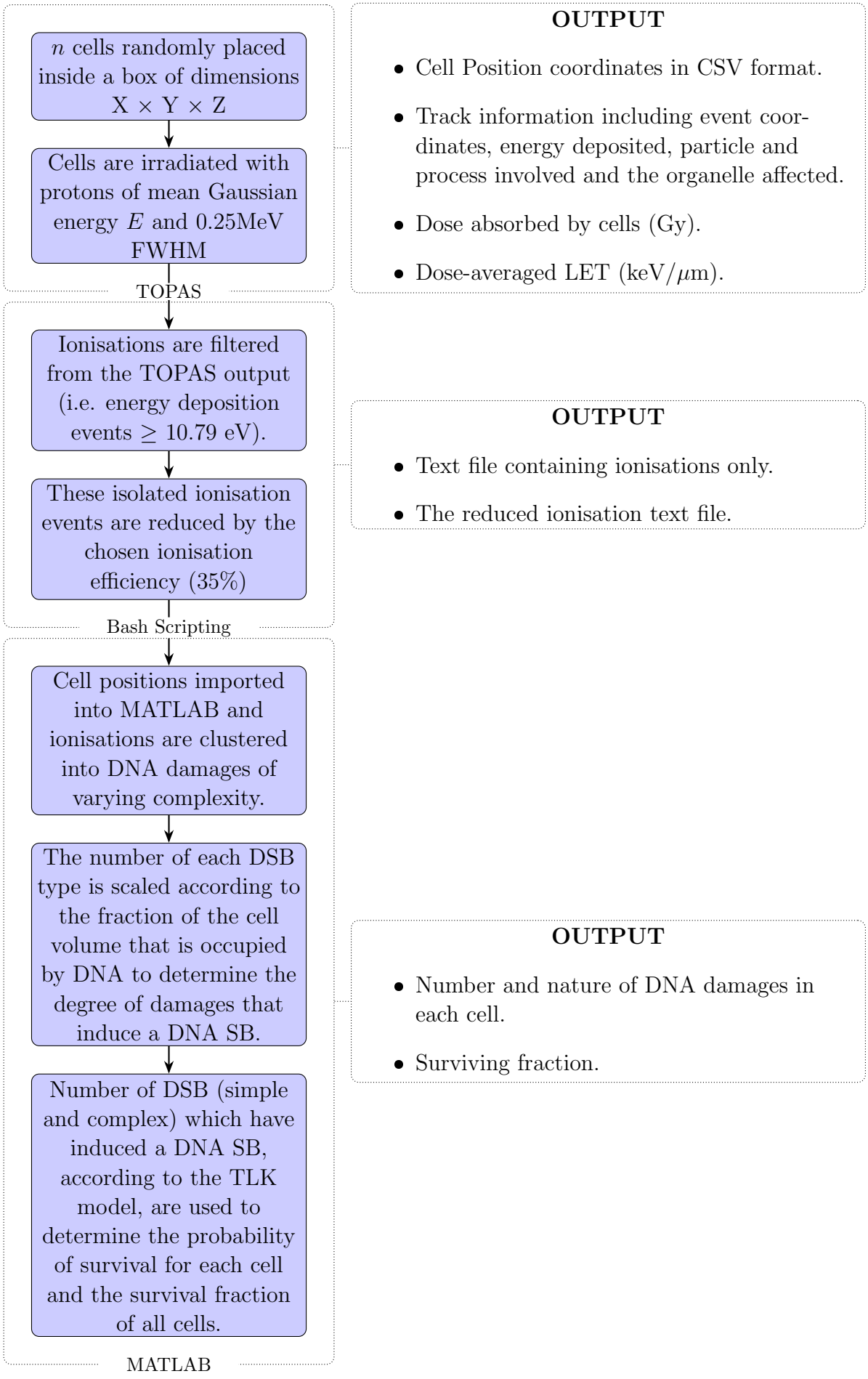


Figure 4.3: Simulation workflow.

4.2 Cell Geometry and Composition

A physicist's view of the cell is often very simplified compared to that of a biologist. In many radiobiological models involving cellular targets, the cell structure is more often than not, approximated by a sphere within another sphere to represent two cell organelles, the cytoplasm and nucleus. In reality however, mammalian cell structure is vastly more complex. A realistic cell cluster model would see each individual cell with a slightly different, non-uniform shape containing all organelles including the mitochondria and endoplasmic reticulum, Fig.(4.4). However, this would require large amounts of memory and processing time. In this thesis, the initial cell model considered was a spherical model with additional organelles as is demonstrated in [89] with the addition of mitochondria placed at random positions inside of the cytoplasm. It was later decided that the cell membrane, mitochondria and endoplasmic reticulum were not required because they receive a negligible fraction of the overall energy absorbed by a single cell and it does not change the dose delivered to the nucleus, nucleolus and cytoplasm significantly when they are removed. For this reason the cell model chosen consisted of a nucleus, a nucleolus and cytoplasm.

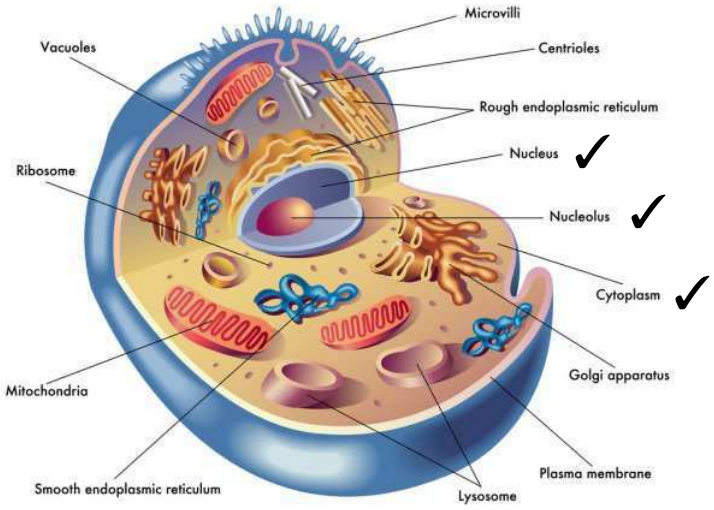


Figure 4.4: A realistic model of a cell.

4.2.1 Chemical Composition and Morphology

The chemical composition of each organelle within the cell should influence the ionisation yield due to the differences in mean excitation energy of the materials present. However, it has been previously shown [89,90] that the ionisation yield does not change substantially when a liquid water composition is used compared to that of a realistic organelle. It has also been noted in Chapter 2 that the Geant4-DNA

physics list is currently only available in liquid water phantoms. In an application where the track structure should be modelled as accurately as possible, the Geant4-DNA physics list is the preferred option over other low energy physics lists[2]. For this reason, our model has already been limited to liquid water phantoms. In this investigation the following parameters were implemented and tested :

- The first being a cell with liquid water organelles of the same density ($1.0 g/cm^3$) and a cell made up of liquid water with their densities scaled to match that of the corresponding organelle.
- The dimensions of the organelle used in this investigation are the same as those used in [89], except for the endoplasmic reticulum (ER) which was widened to account for the layering effect of the organelle surrounding the nucleus, Fig.(4.4). The dimensions are based on the those of the average human cell due to the lack of published data on this matter in the literature.
- The mitochondria organelles were not modelled in [89] but the capabilities of the TOPAS-nBio radiobiology extension [84] made this simple to implement. Each mitochondria is ellipsoidal in shape, meaning that the user must define it's major and semi-major axes dimensions. It is known that a single cell contains many mitochondria, therefore TOPAS-nBio allows users to randomly place a select number of mitochondria inside the cytoplasm of each cell. The placement of mitochondria is randomised in the same fashion as the random placement of cells utilised in the TOPAS-nBio cell culture geometry (discussed below).
- The mitochondria content inside of a cell modelled after the V79 cell line[3] has been investigated previously in the literature [91]. The results of this study showed that the mitochondria occupy 0.05% of the total cell volume with a mean number of 90 mitochondria per cell. To implement these mitochondria concentrations into our single cell simulation, a mean volume of the cell occupied by a single mitochondria organelle was calculated and the ellipsoid dimensions were selected based on this, Table (4.1). The density was chosen to be the same as the cell membrane ($1.1 g/cm^3$) due to membrane material making up most of its composition [92].
- The imperfect biology of the eukaryotic cell means that the morphology of a single cell does not necessarily represent all cells of the same type [94]. Therefore, to achieve the most realistic simulation possible, the shape of every cell should be unique. However, this is beyond the scope of this thesis and

[2]The Livermore and Penelope physics models.

[3]The V79 cell line is derived from the Chinese Hamster and is the most common cell line used in radiobiological research. It was chosen because the dimensions of most of its organelles have been previously reported in the literature.

Parameter Name	Parameter (unit)
a (Fig.(4.5))	0.65 (μm)
b (Fig.(4.5))	0.65 (μm)
c (Fig.(4.5))	0.95 (μm)
Number of Mitochondria	90
Percentage Occupied by Mitochondria	0.05 (%)
Density	1.1 (g/cm^3)

Table 4.1: Mitochondria Geometry Dimensions and Composition [93].

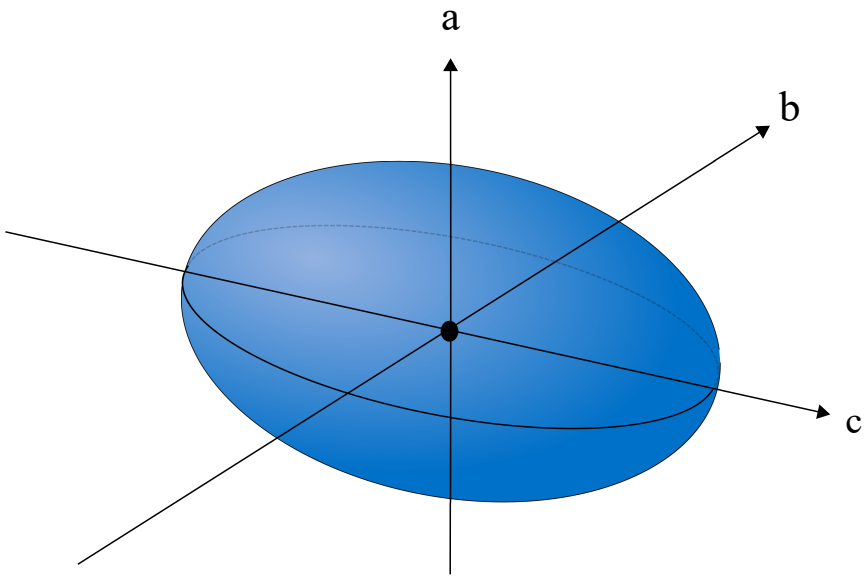


Figure 4.5: Ellipsoidal dimensions used to model the mitochondria.

would require computation power too high to make the simulation repeatable. Although this aspect could be subject to future investigations. In the current model, the cellular dimensions were approximated.

- It was decided that spherical cells would be the most suitable, due to the simplicity involved in avoiding cell overlaps.

TOPAS automatically induces an "overlap check" in all simulations. It is a simple check to perform for uniform shapes, however for non-uniform shapes this takes much longer with more computation time required. The use of ellipsoidal cells in Monte Carlo simulations has been investigated previously [89, 95] where the latter compared ellipsoidal and spherical shaped cells by comparing their S-values[4]. The results showed agreement within 20% between each shape. For this reason, it was decided that the additional processing time required to calculate overlap regions with ellipsoidal cells could not be justified because it would have a minimal effect on the final outcome.

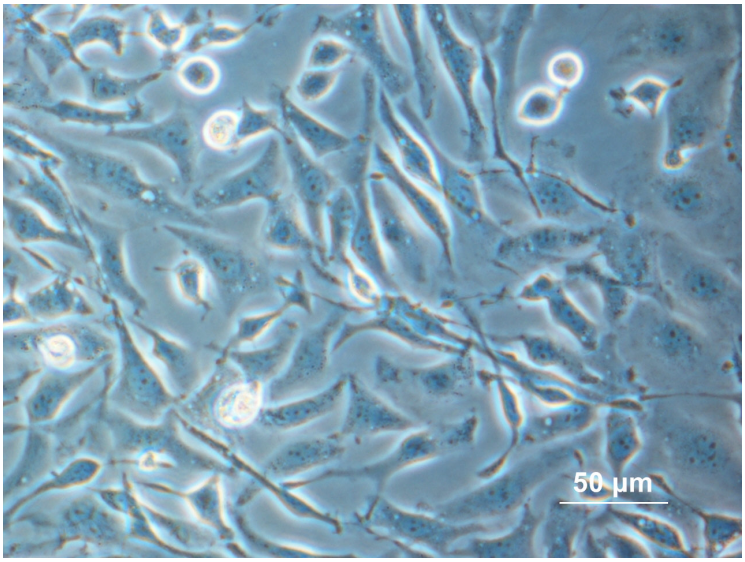


Figure 4.6: Chinese Hamster Ovary (CHO) cells in culture. (adapted from *https://bioprocessintl.com*).

4.2.2 Modelling Cell Organelles

This investigation was undertaken in TOPAS by irradiating a single cell phantom like that shown in Fig.(4.7). The cell was exposed to 1,000 protons of mean energy 2MeV (Gaussian with 0.25MeV FWHM) and the relative energy deposited in each

[4]The S-value is the mean absorbed dose to a target region per unit accumulative activity in some source region.

Organelle	Density (g/cm^3)	Radius (μm)
Nucleus	1.0	3.00
Nucleolus	1.1	2.00
Endoplasmic Reticulum	1.1	3.4, 3.6 (inner, outer)
Cytoplasm	1.0	9.00
Membrane	1.1	9.00, 9.01 (inner,outer)

Table 4.2: Cell Organelle Densities and Dimensions considered for testing purposes [89, 90].

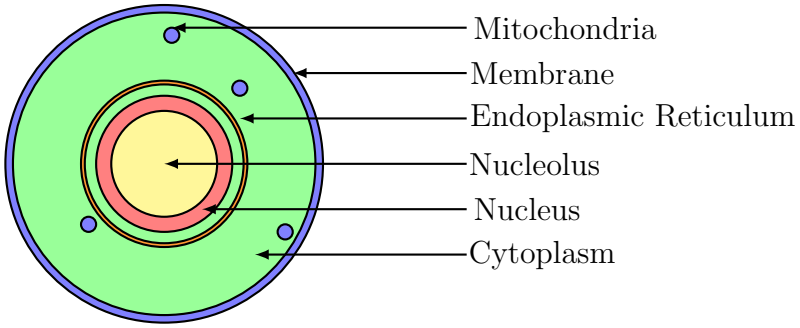


Figure 4.7: A single cell.

organelle was calculated. This simulation was undertaken with the scaled and non-scaled density liquid water cases described above (the densities used are shown in Table (4.2)).

The relative energy deposited in each cell organelle in the scaled and non-scaled density water phantoms is shown in Fig.(4.8). It is shown that the majority of energy is deposited in the cytoplasm because its large volume fraction in the cell, followed closely by the Nucleus and Nucleolus. The relative energy deposited in the ER and Membrane are negligible (<1%). Overall the scaled and non-scaled density cases do not appear to change the energy deposition significantly. This calls into question the need for modelling these two cell organelles and whether the energy deposition in the Cytoplasm, Nucleus and Nucleolus would change significantly if the ER and Membrane were removed. To test this, the same simulation described above was performed on a cellular phantom (with scaled-densities) with these organelles removed. The result was not shown to change significantly in Fig.(4.9). This is due to the negligible volume that these organelles occupy inside the cell. In light of our findings, it was decided that the scaled-density liquid water phantom would be used in subsequent simulations with the ER and membrane removed since they have a negligible effect on the final result.

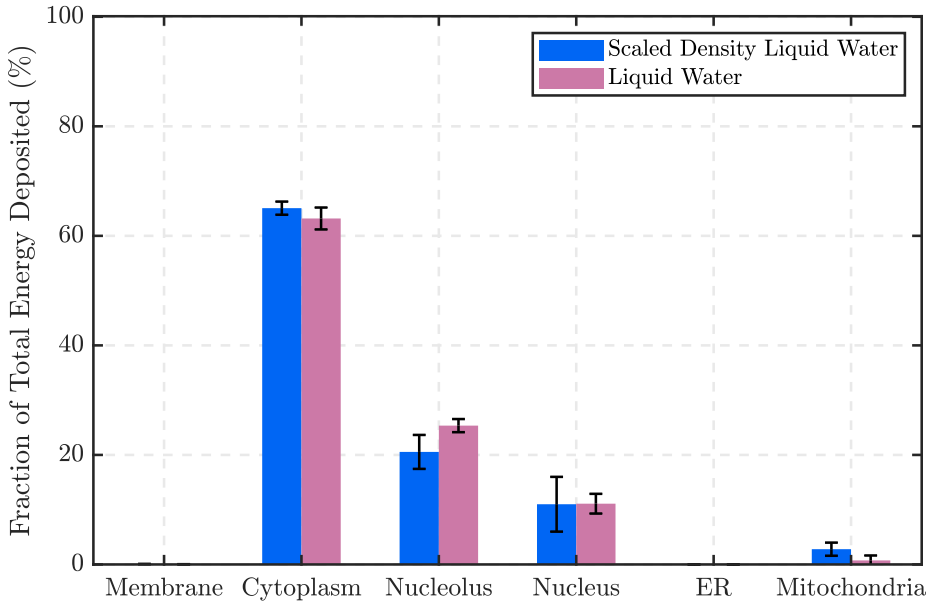


Figure 4.8: Fraction of the total dose delivered to each organelle per track.

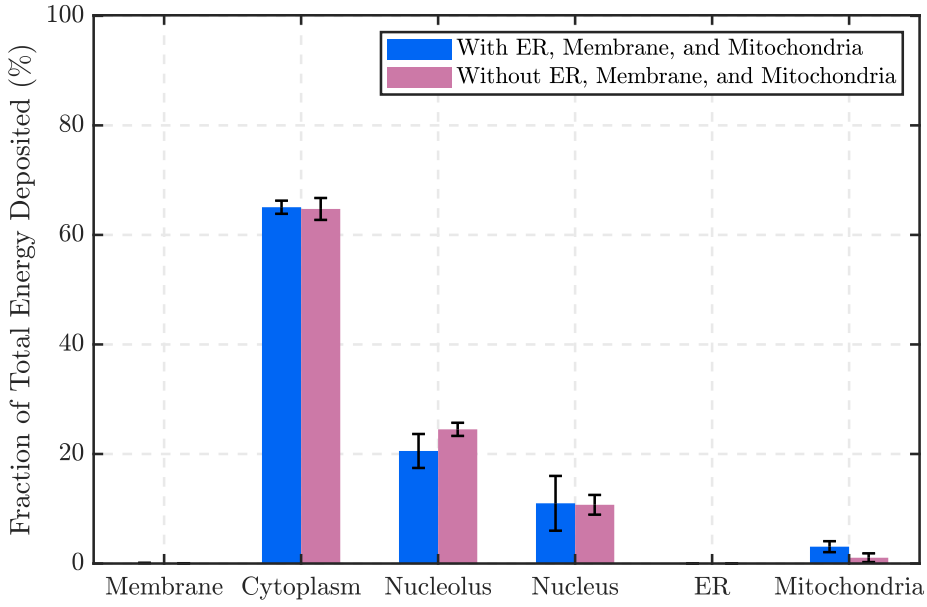


Figure 4.9: Fraction of the total dose delivered to each organelle per track with the inclusion of the endoplasmic reticulum and membrane.

```
s:Ge/MyCulture/Type="TsCellCulture"
s:Ge/MyCulture/Material="G4_WATER"
s:Ge/MyCulture/Parent="World"
d:Ge/MyCulture/Container_HLX= 100 um
d:Ge/MyCulture/Container_HLY= 100 um
d:Ge/MyCulture/Container_HLZ= 20 um
i:Ge/MyCulture/NumberOfCells = 20
d:Ge/MyCulture/CellRadius=10 um
d:Ge/MyCulture/NucleusRadius= 6 um
s:Ge/MyCulture/Nucleus/Material= "G4_WATER"
```

Figure 4.10: Defining a cell culture geometry in a TOPAS parameter file.

4.2.3 Implementing Cellular Geometry into the Current Model

In this study, the cell line of interest is the V79 cell derived from a Chinese Hamster. As previously mentioned, there is no single cell size for a given cell line as shown in Fig.(4.6). For this reason a single cell size is not reported in the literature but a typical range of sizes can be taken from the literature to determine an average size. The V79 cell line has been previously reported to have an average radius of 8.5μm with a mean nucleus radius of 7μm [96]. The dimensions of the nucleolus across different cell lines are not typically reported in the literature, so a radius of 5 μm was used in the simulation as was the case in [48].

The random placement of cells is a process built into the TOPAS-nBio extension framework and is known as the cell culture geometry (with class name `TsCellCulture`). In this class, a number of spherical cells with given dimensions and composition are placed inside a box at random. Each of these parameters is defined by the user inside the parameter file (see Fig.(4.10)).

This random placement is implemented in Geant4 using the toolkit's random number generator class `G4UniformRand()` which gives a uniform random number distribution in the range $[0, 1]$. This range is adapted to fit the cells of a given radius inside of a cube with half length dimensions HLX $\times$ HLY $\times$ HLZ by defining the coordinates for each cell to be

```
x = (2*G4UniformRand() - 1)*(HLX - CellRadius),
y = (2*G4UniformRand() - 1)*(HLY - CellRadius),
and z = (2*G4UniformRand() - 1)*(HLZ - CellRadius).
```

This random placement of cells is necessary to simulate the way in which cells are plated in a real cell culture experiment as shown in Fig.(4.11). Positioning the cells

in a lattice structure would not achieve this stochastic nature present in experiment, because the cells would have similar probabilities of being hit across all simulations. *Implementing a random seed as described above ensures that the cells are arranged differently in every simulation.* Each cell is checked for overlapping with previously placed cells by comparing the Euclidean distances between the centre of the cell being placed and all other cells. If the Euclidean distance is smaller than twice the radius of the cell (8.5μm in our simulations), the cell will be placed in different positions until no overlap is present.

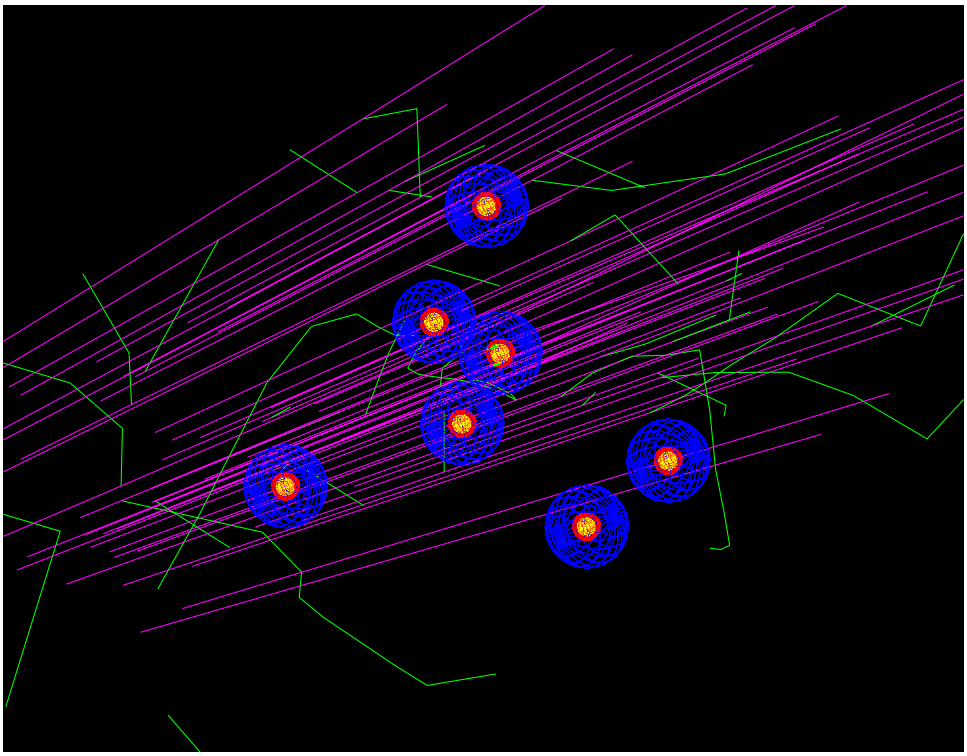


Figure 4.11: Randomly placed cells irradiated with a monoenergetic (1MeV) proton beam in a water medium using TOPAS. The magenta tracks represent proton events and the green tracks represent electron events.

The cell culture itself is positioned in the centre of a world composed of air with dimensions 20 $\times$ 20 $\times$ 20mm. The cells are arranged in a mono-layer formation[5]. Such behavior is derivative of cell culture experiments. A realistic tumour growth model was developed in [89], however the computation and memory time required to implement this into the current model would be too large considering the need for multiple repetitions of the each simulation to allow an adequate statistical analysis to be performed.

[5]Meaning that the cells are laid flat on a surface.

4.3 Cell Irradiation

Let us recall the following points made in Chapters 2 and 3.

- DNA breaks are the primary pathway to cell death and are caused by direct and indirect ionisation events on or in the vicinity of a DNA segment.
- Ideally all interaction processes with all cell organelles would be simulated in this model however this is not feasible because of the computational requirements and the need for simulation repeatability.
- The optimal trade-off between accuracy and simulation optimisation must be determined, meaning that aspects that have a strong influence on the result should be modelled. Aspects of optimisation with a negligible impact on the results whilst causing an increase in processing time and/or memory can be discarded from the simulation workflow.

The irradiation of the cell model developed in Section (4.2) was designed with these points in mind.

4.3.1 Beam Delivery

The proton beam is delivered as a pristine Bragg Peak of varied mean energy. Note that other beam delivery methods such as the spread-out Bragg Peak (SOBP) and pencil beam delivery methods can be tested. However, this could not be performed in this thesis due to time constraints.

The geometry of the simulation with dimensions and compositions is depicted in Fig.(4.12). The source generates the number of protons specified by the user. A time of delivery can also be defined using "sequential times" and is implemented by defining a initial and final time with n steps in between. *Time features allow a dose rate to be defined which was kept at approximately 60 seconds per Gy for all of our simulations.*

The dose was varied by the number of primary protons generated. We have discussed the number of primary particles required to generate 1Gy of absorbed dose for different LET values above. It is worth noting that the absorbed dose is linearly correlated with the number of primary particles produced. The beam itself is generated such that the entire cell culture volume is irradiated. The beam specifications are used in the simulations are provided in Appendix B.

The beam characteristics and profile within the world were kept constant throughout all simulations performed short of the mean beam energy. In a real world setting, the beam profile is adjusted using a series of scattering and modulating filters. These

devices play an important role in the sparing effect of healthy tissue. In the simulations, this was emulated by adjusting a series of position and scattering angles such that the cell culture fell entirely within the high dose region of the beam in the xy-plane, Fig.(4.13). The beam characteristics implemented in our simulations are summarised in Table (4.3) below.

Parameter Description	**Parameter (unit if applicable)**
Source Type	Beam
Particle Type	Proton
Gaussian Mean Beam Energy	1 to 50 (MeV) (giving an LET of 0 to 40 keV/μm)
Beam Energy Spread	0.25MeV FWHM
Beam Shape	Rectangle
Beam Position Distribution	Flat
Beam Angular Distribution	Flat
Beam Position Cutoff (XY)	820 $\times$ 820 μm
Beam Angular Cutoff (XY)	10 (degrees)
Beam Angular Spread (XY)	0.01 (degrees)

Table 4.3: TOPAS parameters implemented to control the shape and scattering angles of the beam. These parameters must be defined for a particle source in TOPAS.

4.3.2 Methods of LET Variance

The primary endpoint of these simulations is a series of dose-response curves with varying energy and therefore LET, measured up to 40 hours post-irradiation. The purpose of measuring cell survival at multiple times after irradiation is to observe the repair dynamics in cells as they progress through the cell cycle. The mean energy and hence LET in the cell can be varied using two methods. The first is to position the cell culture at a depth of 0 μm in liquid water where the mean energy is varied between 1.6 and 50MeV, resulting in proton LET variance between 1 and 40keV/μm. A total of 11 dose-response curves were generated using this method.

The second method of LET variance is performed by maintaining a constant mean energy[6] from the particle source whilst varying the depth of the cell culture in a liquid water medium. The range of a 20MeV proton beam in water is approximately 4.20mm, Fig.(4.14). As discussed in Chapter 2 the proton energy will decrease gradually in the entrance and proximal regions (0 to $\sim$3mm depths) and rapidly in the Bragg Peak region ($>$ 3mm). In this study, depths across all PDD regions were chosen, Fig.(4.15). A 20MeV proton beam was used for this study due to the wide range of LET that it provides. This allows for more comparison with the energy-varied LET study described above.

[6]In the context of this study, the energy is 20MeV

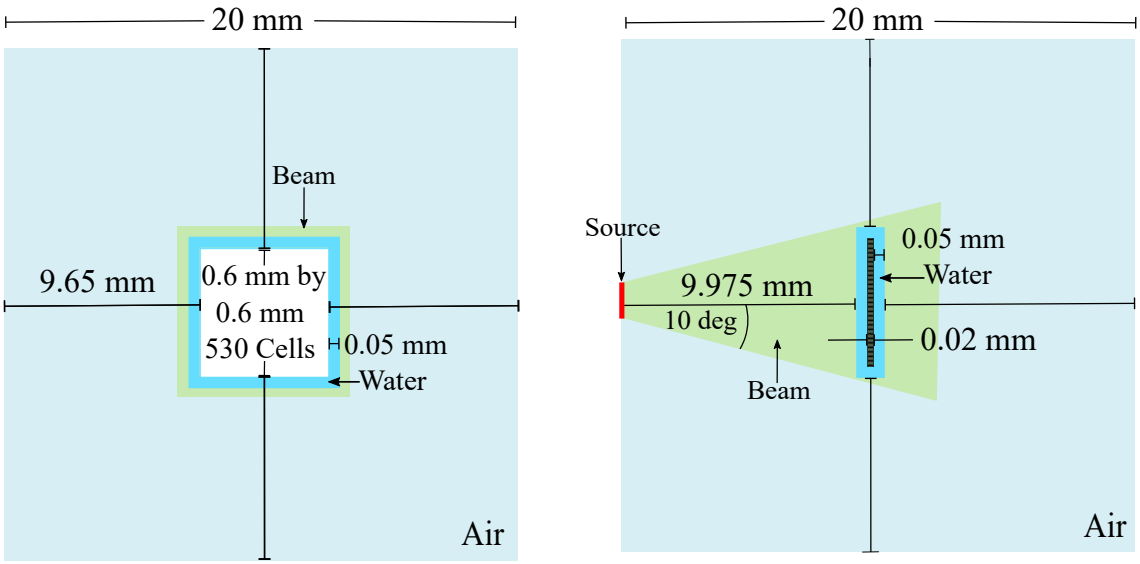


Figure 4.12: Experimental setup in the TOPAS simulation with dimensions indicated.

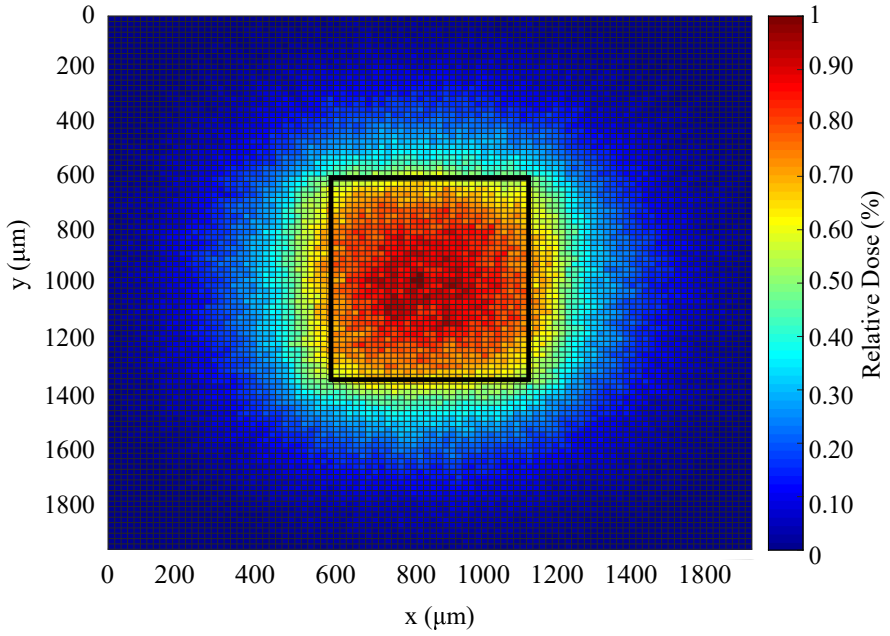


Figure 4.13: A heatmap of the beam profile within the world. The black box represents the outline of the cell culture.

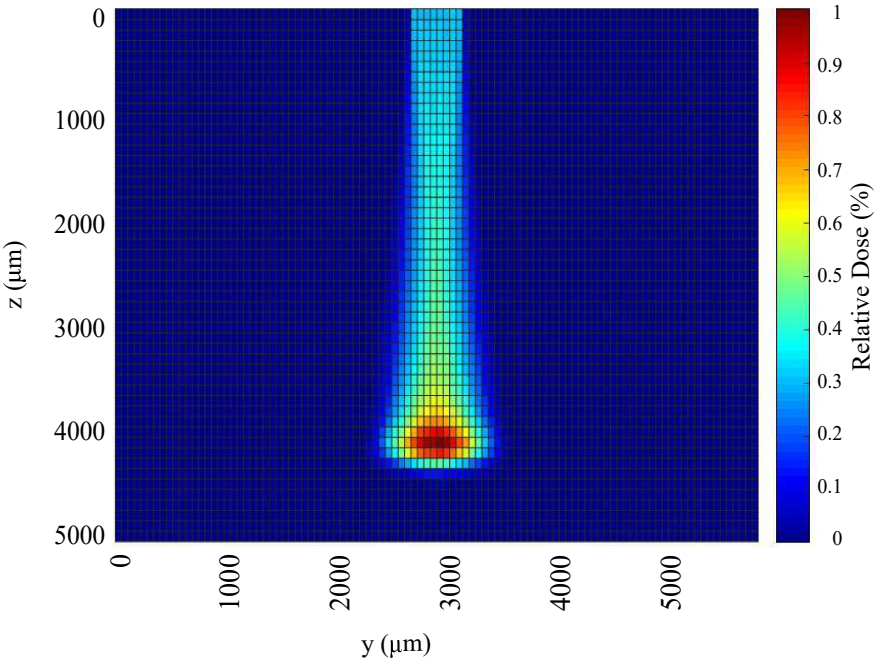


Figure 4.14: A 20MeV mean energy proton beam depth profile with the relative dose deposited in a voxelised water phantom (i.e. the entire medium is divided in smaller sub-volumes).

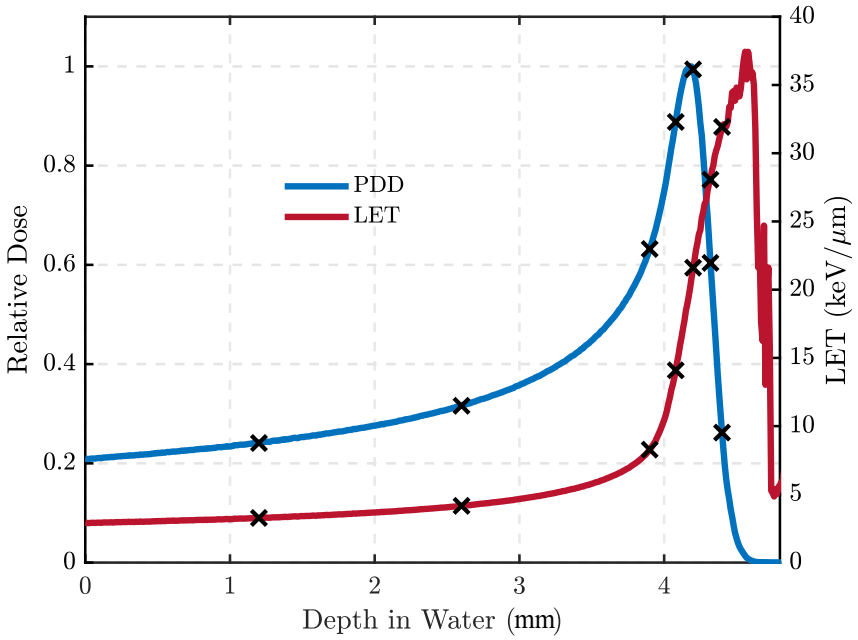


Figure 4.15: The 20MeV proton beam PDD and dose-averaged LET as a function of depth in liquid water with the depths studied using the current model indicated.

4.3.3 Ionising Radiation and Its Interactions with the Cellular Environment

4.3.3.1 The Physical Interactions Modelled in the Cells

The interactions of a particle traversing matter are implemented through the use of Geant4 physics models, three of which were considered for this thesis. These are the Geant4-DNA, Livermore and Penelope physics models. These models were chosen because they have been widely claimed to be the ideal for microdosimetry applications where clinical proton beam energies are used [76, 89]. The simulation used to test these models consisted of liquid water phantom of dimensions 1 × 1 × 1 cm (a cube) with a single proton of mean energy between 1 and 20MeV propagating through it. The output showed that substantially more ionisation events were recorded along the track when Geant4-DNA was used compared with the Livermore and Penelope models[7], Fig.(4.16). The Penelope physics list was quickly ruled out because it has a smaller energy range in which interactions can occur than the Livermore model. It was decided that Geant4-DNA would model DNA damage on the nano-scale with the most accuracy of the three candidates. However, a detailed track structure is not required in all geometrical regions in the simulation.

Recall that our key aim at this stage of the simulation is to estimate DNA damages as accurately as possible. This means that the only places where a detailed, nano-scale track structure is required, is in these regions of the cell that contain DNA. Bearing this in mind, the optimisation method (see the discussion in Chapter 3) of region-based physics lists was implemented into the simulation by using a combination of Livermore and Geant4-DNA physics models in the cell organelles that contain DNA (i.e. the nucleus). Damage to the cytoplasm and nucleolus do not play a significant role in cell death, therefore it was decided that this region did not require the computationally expensive Geant4-DNA physics model, Fig.(4.17). For this reason, only the Livermore physics list was used in these regions as well as the "world" where a detailed track structure would cost unnecessary amounts of memory and processing time with no gain in accuracy. Table (4.4) summarises all of the physical processes simulated inside of the cells by each of the chosen physics models.

[7] In this thesis an ionisation is defined as an energy deposition $\geq$ 10.79eV

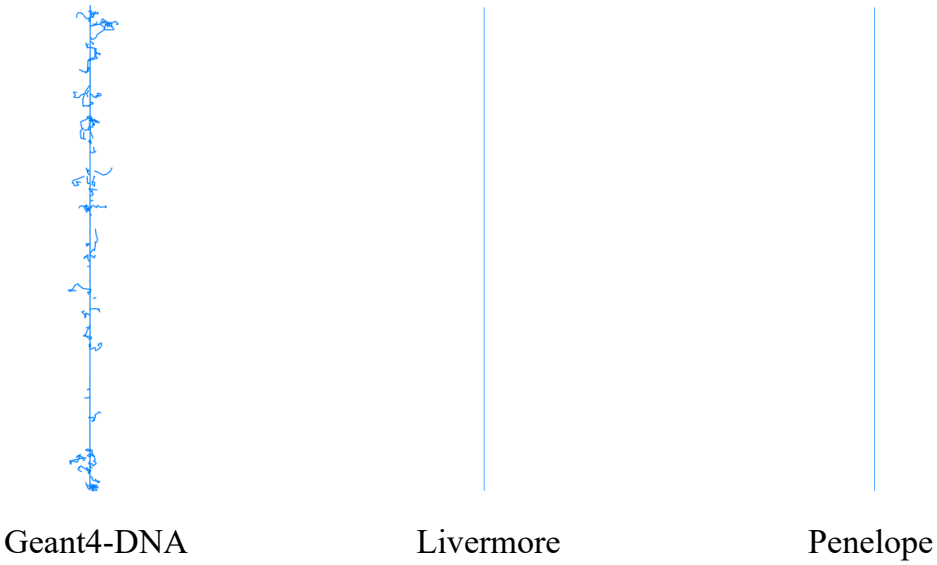


Figure 4.16: A 1MeV proton track of length 0.1μm in water using different low energy physics models in TOPAS.

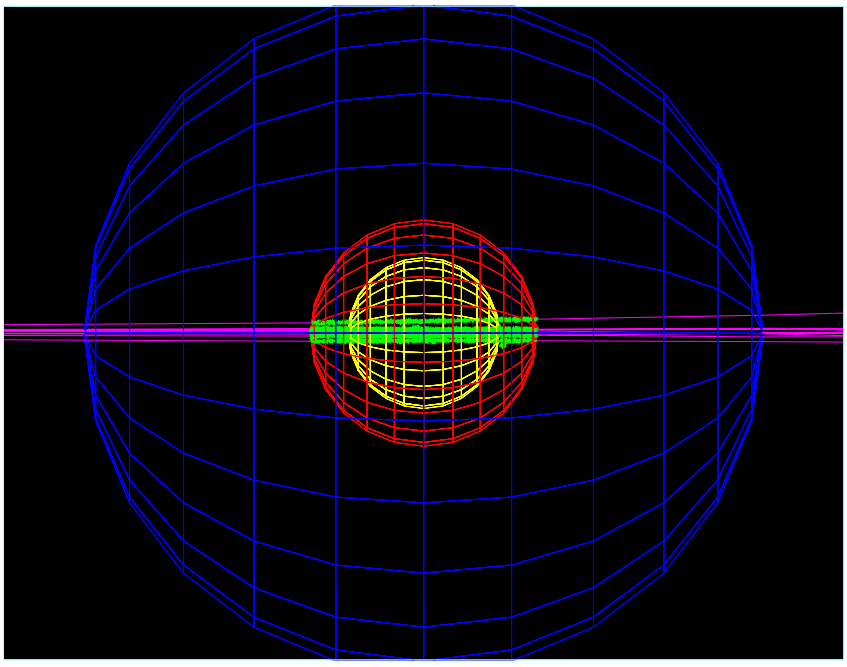


Figure 4.17: A single cell with the Livermore physics model implemented in the cytoplasm (blue) and a combination of the Livermore and Geant4-DNA physics models in the nucleus and nucleolus (red and yellow). The pink and green tracks represent protons and electrons respectively.

Particle	Process Type	Model	Energy Range	Reference
Photon γ	Photo-electric Effect	Livermore Photo-Electric Model	250 eV - 100 GeV	
	Compton Scattering	Livermore Compton Model	250 eV - 100 GeV	
	Rayleigh Scattering	Livermore Rayleigh Model	250 eV - 100 GeV	
	Conversion	Livermore Gamma Conversion Model	1.022 MeV - 100 GeV	
Electron e^-	Ionisation	Livermore Ionisation Model	10 eV - 100 GeV	
		Born Ionisation Model	11eV - 1 MeV	[97]
	Bremsstrahlung	Livermore Bremsstrahlung Model	10 eV - 100 GeV	
	Elastic Scattering	Champion Elastic Model[a]	7.4 eV - 1 MeV	[98]
		Screened Rutherford Elastic Model[a]	$\leq$ 1 MeV	[99]
	Electron Excitation	Born Excitation Model[a]	9 eV - 1 MeV	[100, 101]
	Vibrational Excitation	Sanche Excitation Model[a]	2 eV - 100 eV	
	Attachment	Melton Attachment Model[a]	4 eV - 13 eV	
Proton p	Nuclear Scattering	Ion Elastic Model[a]	100 eV - 1 MeV	
	Electronic Excitation	Miller Green Excitation Model[a]	10 eV - 500 keV	[102]
		Born Excitation Model[a]	500 keV - 100 MeV	[100, 101]
	Ionisation	Rudd Ionisation Model[a]	$\leq$ 500 keV	[103]
		Born Ionisation Model[a]	500 keV - 100 MeV	[100, 101]
	Electron Capture	Dingfelder Charge Decrease Model[a]	100 eV - 100 MeV	[97]

Table 4.4: Physical processes simulated in and around the cellular environment.
[a] Nucleus only.

4.3.3.2 Chemical Interactions

Recall the stages of radiation cellular response from Chapter 2. The physico-chemical and chemical stages contribute to DNA damage through indirect interactions. Geant4-DNA chemistry is capable of simulating these interactions through radiolysis reactions with water and is available in the TOPAS-nBio framework [104]. This begins with the ionisation of water molecules producing a series of highly reactive chemical species before undergoing dissociation into a number of species capable of interacting with the surrounding environment. The diffusion of each molecule is governed by the diffusion coefficients presented in Chapter 3 occurring via Brownian motion.

The simulation of chemical species in Geant4-DNA chemistry proceeds on a time scale, beginning at 1ps and continuing until 1μs, Fig.(4.18). This time dependence allows the user to select a time resolution for chemical species which is equivalent to the step size of physical interactions occurring over time instead of distance. If this were not implemented in the simulation, the processing requirements would be too large, even with the computational power currently available. Due to the processing time and large memory requirements, the chemical interactions could not be included in this study. Due to the geometry complexity combined with the high-LET radiation used and large number of primary particles generated, we experienced regular simulation crashes and the generated output files were too large for post-processing. A focus of future work will be to implement indirect effects such that the processing memory requirements are smaller[8].

The importance of indirect damages, particularly in low-LET particle beams, has been highlighted previously. Therefore, it should be noted that the low-LET survival fraction is most likely overestimated. Because of time constraints and large memory requirements we were unable to implement such processes into our simulations. However, we report how the indirect processes would be implemented had it been feasible.

The time-dependent yield of different radiolytic species suggests a higher concentration of charged, highly reactive species at the beginning of the chemical stage [104]. As time progresses, these species react to form neutral, less reactive species which have a smaller contribution to indirect DNA damage. Therefore, this means that a high time resolution is not necessary at later times in the chemical stage. Geant4-DNA chemistry and TOPAS-nBio allow different time resolutions to be defined for selected time intervals of the chemical stage. Considering the higher yield of reactive species at the beginning of the chemical stage, a higher resolution was used at this time. It was decreased further with increasing virtual time. For this reason, the chemistry stage was also ended at 2.5ns due to the high number of

[8]Post-processing includes characterising of DNA damages, calculations of repair and survival which are discussed later in this chapter.

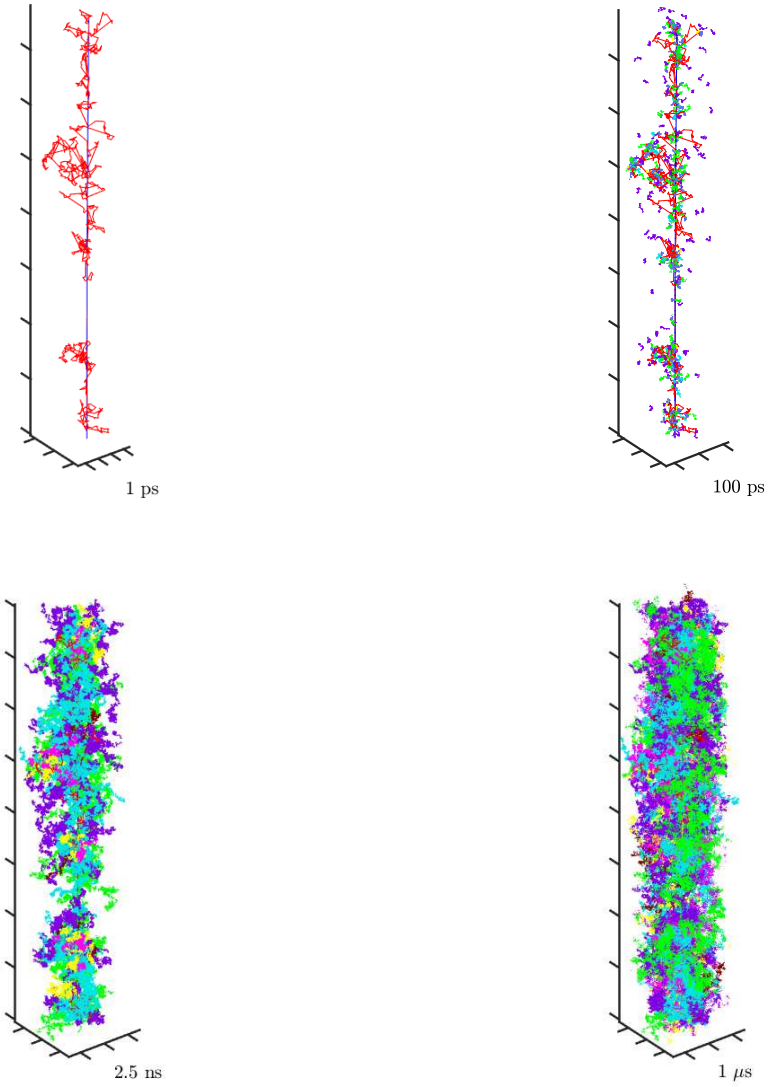


Figure 4.18: A single 1MeV proton propagating through a liquid water medium with Geant4-DNA physics and chemistry models implemented at virtual times 1ps, 100ps, 2.5ns and 1μs. Each species are coloured as follows: proton (blue), electron (red), H_2O (cyan), Hydrogen Peroxide - H_2O_2 (magenta), Hydronium - H_3O^+ (green), Hydrogen - $H^\bullet$ (yellow), Dyhydrogen - H_2 (orange-yellow), Hydroxide - OH^- (orange), Hydroxyl - $OH^\bullet$ (pink-purple) and Solvated electron - $e_{aq}{}^-$ (purple).

primary particles required per simulation. The suggested time steps and resolutions that would have been implemented if indirect effects were simulated are summarised in Table (4.5).

Virtual Time (ps)	Time Resolution Step (ps)
1 - 10	0.1
10 - 10^2	1
10^2 - 10^3	3
10^3 - 10^4	10
10^4 - 10^6	10^2

Table 4.5: Appropriate time steps for radiolysis reactions if chemistry processes were implemented.

4.3.3.3 Characterising DNA Damage of Varying Complexity

In experiment we define a cell as dead if it has lost all "reproductive ability". A cell is classified as such when it cannot proliferate for an extended period time or it has undergone apoptosis or necrosis. A cell can die as a result of a single lethal lesion however its probability of death increases with the number of lethal lesions [23]. In the current study cell death is defined according to the number of lethal lesions present with repair time which is discussed with more detail in following sections. However, to determine the number of lethal lesions present in each cell we must first characterise the damage complexity inside of each cell. The next three sections follow the procedure provided in Fig.(4.19).

DNA damage can come in various forms and complexities, meaning that some repair quickly and others slowly with some becoming susceptible to mis-repair. It is the number of DSB damages containing ≥ 2 strand breaks inside a cell that will be used to determine its probability of survival [33]. In the simulations used in the current study an ionisation cluster is defined as a series of damages that exist within a 3.4nm radius[9], Fig.(4.20). One limitation of this model is that the DNA structure is not physically simulated in the nucleus. Instead, a MATLAB code adapted from [48] is implemented to spatially map ionisations onto a DNA segment by scaling the total damage in the nucleus by the fraction of the cell volume that is occupied by DNA. The nucleus is made up entirely of water with the nucleolus inside and it is determined which damage breaks the DNA during post-processing.

The DNA damage complexity is estimated using the `clusterdata` algorithm available in MATLAB. First the ionisation events were sorted into the cells in which they occur. The ionisations that occured inside of the DNA deficient cytoplasm and nucleolus were removed from further processing by checking if the distance between

[9] 3.4nm = 10 base pairs on a DNA segment.

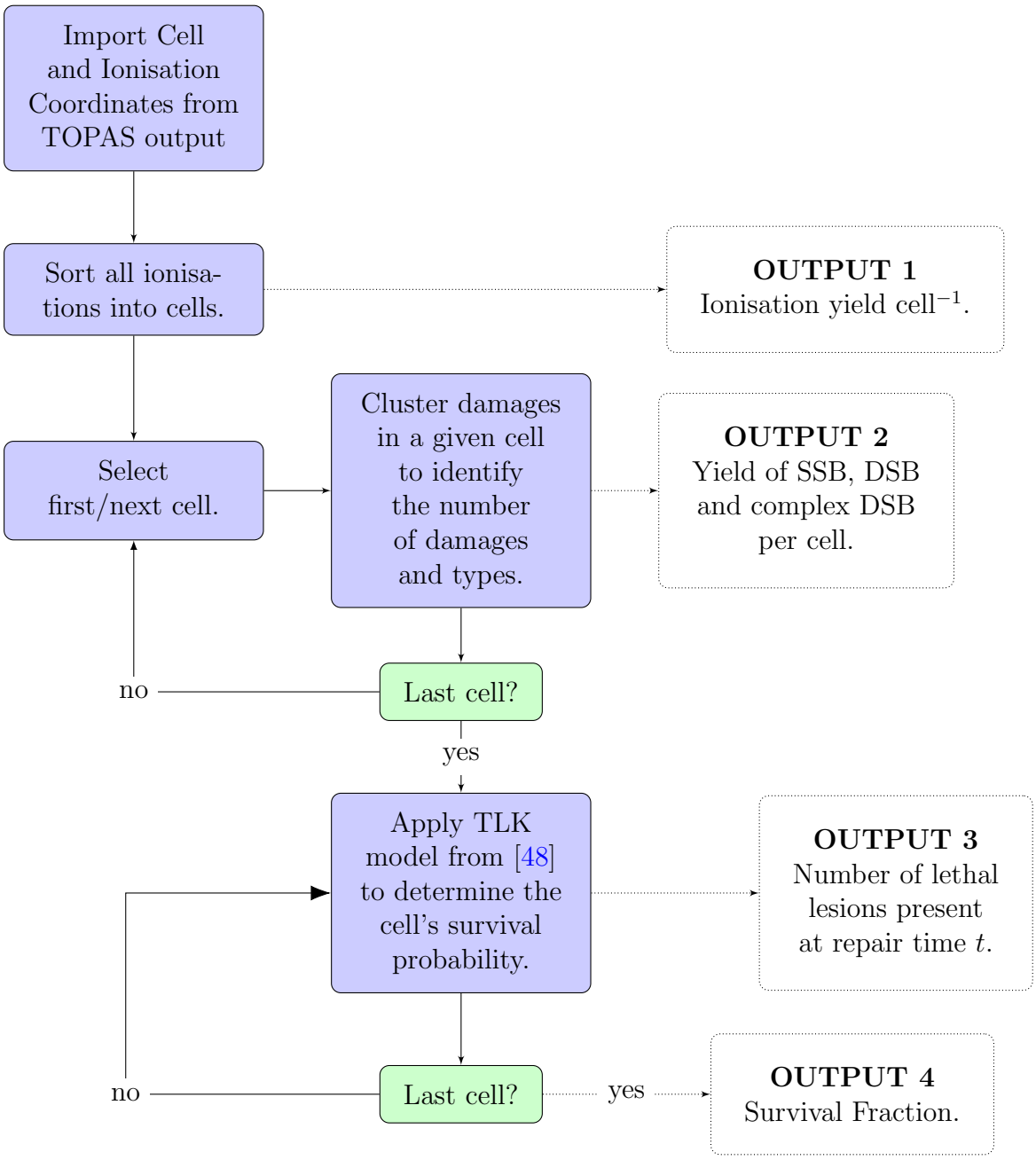


Figure 4.19: Flowchart describing the characterisation of DNA damages, the counting of lethal lesions and calculation of survival fraction.

centre of the cell and the x, y and z coordinates of the ionisation events were in between the radius of the nucleolus (5 μm) and the nucleus (7 μm). This was done by computing the Euclidean distance between every ionisation event and the centre of each cell. All distances less than the i^{th} cell radius were placed in the i^{th} position of a $1 \times n$ cell array where n is the number of cells, Fig.(4.21). Once sorted into cells, the ionisations are clustered using the command,

```
c = clusterdata(X,'criterion','distance','cutoff',3.4e-3...
,'distance','euclidean','linkage','centroid','savememory','on');.
```

where `X` represents the ionisations present in a single cell. The additional options perform the following operations.

- `criterion` : the variable that dictates how the points are clustered.
- `cutoff` : the maximum radius at which the points can be clustered[10].
- `distance` : defines how the distances between each pair of points is calculated. In this case the Euclidean distances are used.
- `linkage` : defines how the clusters are separated. Centroid linkage was used to pinpoint the centre of each cluster which was determined by calculating the cluster's unweighted centre of mass.
- `SaveMemory` : defines whether the distances between points are calculated at one time and stored in a matrix or they are calculated once without being stored in the memory. When the memory requirements to store the matrix are too large, this option should be set to "`on`".

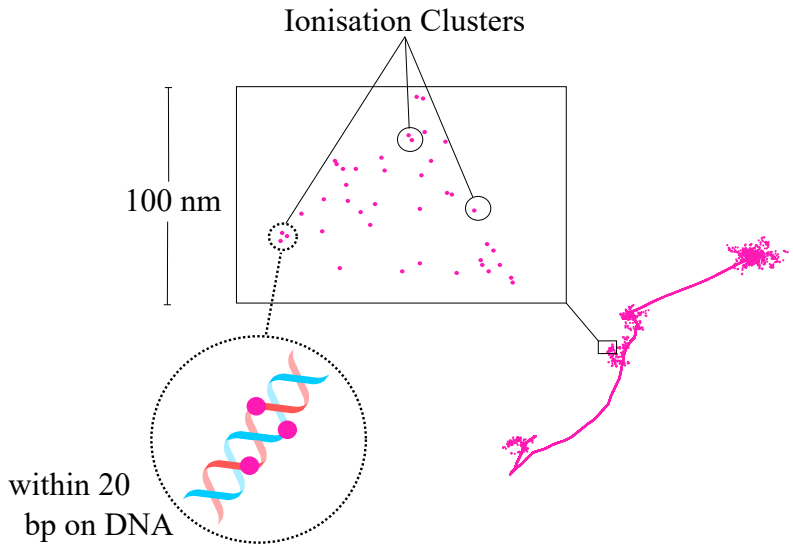


Figure 4.20: The clustering of a group of occurrences damages of from an ionisation track.

[10]In this case it is the length of 10 DNA base pairs.

The output of this function is a series of integers representing the cluster indices[11], Fig.(4.20). The `centroid` option enables cluster centroid positions in close vicinity to be merged into a single cluster. This prevents the same ionisation events from being assigned to multiple clusters and allows major DSB clusters to be defined[12], Fig.(4.22).

```
for i = 1:length(Cell_Position)

        %Euclidean distance between each energy deposition and cell
            position

        Euc_Dist = sqrt((((Cell_Position(i,1)) - (Ionisations(:,1))
           ).^2)...
        + (((Cell_Position(i,2)) - (Ionisations(:,2))).^2)...
        + (((Cell_Position(i,3)) - (Ionisations(:,3))).^2));

        mask = (Euc_Dist <= Cell_Radius);

        Ionisations_Sorted{i} = Data(mask, :);

end
```

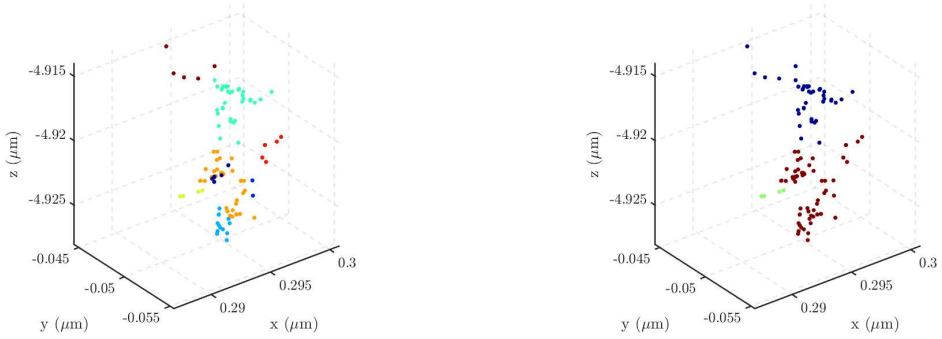
Figure 4.21: The code used to sort ionisation events into cells.

After each ionisation event has been assigned a cluster index, the relative ionisation complexity of each cluster is defined as a single strand break (SSB), simple double strand break (DSB) or complex double strand break (DSB). Once these results are known, they can be scaled according to the fraction of the cell volume that is occupied by DNA. The model approximates the DNA segment to be a cylindrical volume of radius 1nm (10Å) and height equal to the genome length (GL) of the DNA segment. The GL range is unique to the cell line and can vary between cells according to different cell conditions [105]. The literature also reports different genome lengths for the V79 cell [48, 106–108]. To account for the variation of GL in the V79 cell under different conditions, every cell in the simulation was assigned a random, Gaussian distributed GL according to GL$\sim \mathcal{N}(3.8, 1.5)$ Gbp. Using this, the relative fraction of each cell that is occupied by DNA can be calculated with

$$V_{Fraction} = \frac{r_{DNA}^2 L_{DNA}}{\frac{4}{3} r_{Cell}^2}, \tag{4.2}$$

where r_{DNA} is the radius of the DNA cylinder (10Å), r_{Cell} is the radius of the cell and $L_{DNA} = \text{GL} \times 0.34\text{nm}$[13].

[11] i.e. `max(c)` from the above function would give the number of clusters in a single cell
[12] multiple DSBs merge to form a single DSB of higher complexity
[13] The bases in a DNA segment are spaced 0.34nm apart.



(a) A series of ionisations sorted into minor clusters.

(b) Ionisation minor clusters from (a) merged into major clusters.

Figure 4.22: Ionisations clustered in MATLAB using the `clusterdata` function.

The number of DSBs are scaled by $V_{Fraction}$ to estimate the number of ionisation clusters that occurred on or in the vicinity of a DNA segment, Fig.(4.20). The efficiency with which ionisations induce strand breaks can be related to the concentration of oxygen within the cell [29, 57]. It has been discussed previously in this thesis that a lack of oxygen in the cellular environment can decrease the radiation sensitivity of a cell.

4.3.4 Efficiencies of Strand Break Induction on DNA.

The role of oxygen (pO_2) concentration inside a cell has been previously shown to have an effect on the DNA damage yield efficiency, decreasing with a lack of oxygen due to a lower availability of reactants. This concept has been previously modelled using probabilities of direct and indirect ionisations causing strand breaks on the DNA segment or in its first hydration layer [57], Fig.(4.23). Their objective was to model SB induction with the oxygen enhancement ratio which can vary between cell lines. The cell line of focus in this thesis (the V79 Chinese Hamster cell) has been experimentally shown to have an OER of ∼3 [93, 109].

Recalling the definition from Chapter 2, OER can be related to the probability of a SB being induced under oxic ($[pO_2] = 0$ mmHg) and hypoxic ($[pO_2] = 760$ mmHg) conditions as follows

$$OER([pO_2]) = \frac{Pr([pO_2] = 760\text{mmHg})}{Pr([pO_2] = 0\text{mmHg})}, \tag{4.3}$$

according to [57].

The model developed in [57] assumed that under oxic conditions, the SB efficiency for both direct and indirect ionisation events is 100%. Assuming an OER of 3, the range of direct SB efficiency in V79 cells is $0.33 \rightarrow 1$ using Eq.(4.3). The

direct SB efficiency used in [48] was 30%. In our simulations the cells were assumed to have a small concentration of oxygen to more accurately simulate the hypoxic[14] conditions typical in a tumour.

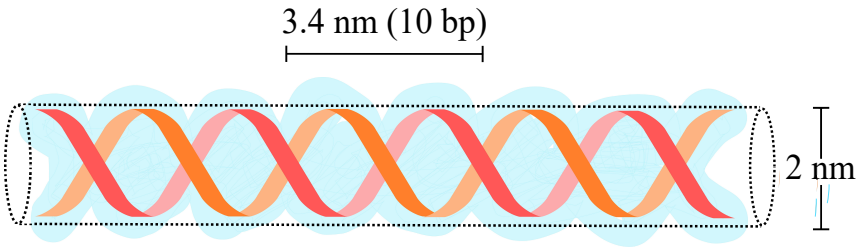


Figure 4.23: A DNA segment with its first hydration layer.

The decrease in radio-sensitivity due to a lack of oxygen was implemented into the current simulation by randomly sampling 35% of the output file from TOPAS containing the ionisations. This was achieved by counting the total number of ionisations occurring inside the cells and using the Bash Scripting command `shuf`, to randomly select 35% of the lines from a file,

```
shuf -n X filename
```

where `X` is the specified number of lines to be randomly chosen. The result of this process is a series of ionisation events that can be considered efficient under the oxygenation effect. This command is performed on the TOPAS output prior to the clustering stage performed in MATLAB, Fig.(4.3).

4.4 Cell Death & Repair

In the final stage of the simulation we model the response of each cell to the amount of damage present inside. We have previously discussed the repair mechanisms that cells undergo in Chapter 2. In this part of the simulation we adapt a MATLAB code from Refs. [48, 87] which model the repair of DNA damage and predicts the probability of survival for a single cell at repair time t_{rep} according to the number of lethal damages present. The TLK model [33] from [48, 87] is used to determine the probability of survival of each cell in our simulations for repair times up to 40 hours. Ultimately, this model from allows us to determine the damage yield in each cell and the survival fraction

$$SF = \frac{N}{N_0}, \tag{4.4}$$

[14] Lack of oxygen.

where N is the number of cells that survived at repair time t_{rep} and N_0 is the total number of cells in the simulation (i.e. $N_0 = 530$). It is assumed that 100% of the cells are alive prior to irradiation. We discuss how the process of determining cell death and repair is undertaken in more detail below.

4.4.1 The Two-Lesion Kinetic (TLK) Model

The TLK model [33] was briefly introduced in Chapter 2, however it will be discussed with more detail in this section because it is used in the current study to calculate the survival fraction. We know that the primary mechanism for cell death is through DNA DSB formation where a DSB is defined as a cluster of damage on a single DNA strand within a distance of 10 to 20 bp. A number of these damage clusters will repair, mis-repair or become lethal[15] which affects the cell's probability of survival.

Firstly, the DNA damages capable of causing a lethal lesion is sorted into two categories, (i) the simple DSB (Type I) and (ii) the complex DSB (Type II). The formation, repair and mis-repair of each type as a function of time t per cell is defined by two first-order differential equations :

$$\frac{d\overline{L}_1(t)}{dt} = 2\dot{D}(t)Y\Sigma_1 - \{\varepsilon_1 + \lambda_1\}\overline{L}_1(t) - \overline{L}_1(t)[\eta_1\overline{L}_1(t) + \eta_{1,2}\overline{L}_2(t)] \tag{4.5}$$

and

$$\frac{d\overline{L}_2(t)}{dt} = 2\dot{D}(t)Y\Sigma_2 - \{\varepsilon_2 + \lambda_2\}\overline{L}_2(t) - \overline{L}_2(t)[\eta_{1,2}\overline{L}_1(t) + \eta_2\overline{L}_2(t)] \tag{4.6}$$

where $\overline{L}_1(t)$ and $\overline{L}_2(t)$ are the number of simple and complex DSB present at time t, respectively. The number of lethal and non-lethal (mis-repaired) lesions (L_f and L_m, respectively) present in a cell at repair time $t = t_{rep}$ are described using,

$$\begin{aligned}\frac{d\overline{L}_f(t)}{dt} = {} & [(1-a_1)\beta_1\lambda_1 + \varepsilon_1]\overline{L}_1(t) + [(1-a_2)\beta_2\lambda_2 + \varepsilon_2]\overline{L}_2(t) + \\ & \gamma_1\eta_1\overline{L}_1(t)\bar{L}_1(t) + 2\gamma_{1,2}\eta_{1,2}\overline{L}_1(t)\overline{L}_2(t) + \gamma_2\eta_2\overline{L}_2(t)\overline{L}_2(t)\end{aligned} \tag{4.7}$$

and

[15] Recall that a lethal lesion cannot repair and leads to cell death. A mis-repaired lesion causes instabilities within the cell but may cause cell death and with a lower probability than lethal lesions.

$$\frac{d\overline{L}_m(t)}{dt} = (1-a_1)(1-\beta_1)\lambda_1\overline{L}_1(t)+(1-a_2)(1-\beta_2)\lambda_2\overline{L}_2(t)+ \\ (1-\gamma_1)\eta_1\overline{L}_1(t)\overline{L}_1(t)+(1-\gamma_2)\eta_2\overline{L}_2(t)+2\gamma_{1,2}\eta_{1,2}\overline{L}_1(t)\overline{L}_2(t). \tag{4.8}$$

The parameters are defined as follows,

- $\dot{D}(t)$ is the dose rate with respect to time.
- Y represents the number of base pairs (bp) per cell.
- $\Sigma_{1,2}$ is the expected DSB yield of Type I and II DSBs per base pair per Gy (the factor of 2 converts this value to nucleotides[16] per Gy because there are two nucleotides per base pair).
- $L_{1,2}$ represents the formation of Type I and II DSBs.
- $\lambda_{1,2}$ is the probability of Type I and II DSB repair respectively per unit time [hr^{-1}]
- $\eta_{1,2}, \eta_1$ and η_2 are the binary interaction probabilities (i.e. the probability of two or more separate DSBs interacting during the repair process) between Type I and II, Type I only and Type II only, respectively.
- $a_{1,2}$ is the fidelity parameter for linear repair.
- $\varepsilon_{1,2}$ represents the physio-chemical fixation of Type I and II DSBs[17].
- $\beta_{1,2}$ is the probability of incorrect repair of each DSB type per unit time [hr^{-1}].
- $\gamma_{1,2}$ is a correction term that accounts for the formation of non-lethal chromosome aberrations [33].

The lower and upper bounds placed on the model are,

$$0 \leq \beta_1, \beta_2, a_1, a_2, \gamma \leq 1 \\ \text{and } 0 \leq \lambda_1, \lambda_2, \eta_{1,2}, \eta_1, \eta_2, \Sigma_1, \Sigma_2. \tag{4.9}$$

This version of the TLK model contains many biological parameters, making its application a very cumbersome task. For this reason, it is suggested in [33] that the following simplifications can be made.

- The parameters $\eta_{1,2}, \eta_1$ and η_2 can be condensed into a single binary mis-repair parameter η. Similarly for $\gamma_1, \gamma_{1,2}$, and γ_2.

[16] Recall that a nucleotide contains a single DNA base binded to the strand.

[17] Physio-chemical fixation occurs when the "sticky ends" of a broken DNA segment interacts irreversibly with other proteins or DNA-bound molecules forming a chromosome break [33].

- $\varepsilon_1 = \varepsilon_2 = 0$, indicating no DSB repair fixation.
- $a_1 = a_2 = 0$, indicating that a repaired DSB will always result in a lethal or non-lethal mutation.

Making use of the simplifications recommended reduces the model to a much more manageable 10 parameters resulting in Eqs.(4.5)-(4.8) being re-written as follows :

$$\frac{d\overline{L}_1(t)}{dt} = 2\dot{D}(t)Y\Sigma_1 - \lambda_1\overline{L}_1(t) + \eta\bar{L}_1(t)[\overline{L}_1(t) + \overline{L}_2(t)], \tag{4.10}$$

$$\frac{d\overline{L}_2(t)}{dt} = 2\dot{D}(t)Y\Sigma_2 - \lambda_2\overline{L}_2(t) + \eta\overline{L}_2(t)[\overline{L}_1(t) + \overline{L}_2(t)], \tag{4.11}$$

$$\frac{d\overline{L}_f(t)}{dt} = \beta_1\lambda_1\overline{L}_1(t) + \beta_2\lambda_2\overline{L}_2(t) + \gamma\eta[\bar{L}_1(t) + \overline{L}_2(t)]^2, \tag{4.12}$$

$$\frac{d\overline{L}_m(t)}{dt} = (1-\beta_1)\lambda_1\overline{L}_1(t) + (1-\beta_2)\lambda_2\overline{L}_2(t) + (1-\gamma)\eta[\overline{L}_1(t) + \overline{L}_2(t)]^2 \quad . \tag{4.13}$$

Now the parameters in the above model can be calibrated to experimental dose-response curves by reducing the weighted Pearson's χ^2 goodness-of-fit parameter,

$$\chi^2 = \frac{1}{N^2}\sum_{i=0}^{N} w_i[X_i - P_i]^2. \tag{4.14}$$

A strategy suggested in [33] is to calibrate parameters η, λ_1, λ_2 and Σ_2 to an existing dose-response curve with the remaining parameters set to reasonable values given in Eq.(4.9). In the context of this thesis, this is achievable with Type I and II DSB yields being predicted before the TLK model is applied. The calibration of the parameters in the TLK model to experimental data in this thesis will be implemented in Chapter 7.

The nature of the differential equations in the TLK model can make them difficult to solve analytically. In [48], Euler's method was applied to solve Eqs.(4.11)-(4.13) numerically in each cell at repair time t_{rep} (with $\dot{D}(t_{rep}) = 0$ because the measurements are made after irradiation has stopped). In this method, iterations are performed for a small time step Δt_{rep} from $t_{rep} = 0$ to the maximum repair time, subject to the boundary conditions $L_f(0) = L_m(0) = 0$[18]. The expressions are as follows :

[18]No damages have been classified as lethal or non-lethal at $t_{rep} = 0$ because they have not been given time to repair.

$$L_1(t+\Delta t) = L_1(t) - (\lambda_1 L_1 + \eta L_1(L_1+L_2))\Delta t \quad , \tag{4.15}$$
$$L_2(t+\Delta t) = L_2(t) - (\lambda_2 L_2 + \eta L_2(L_1+L_2))\Delta t \quad , \tag{4.16}$$
$$L_f(t+\Delta t) = L_f(t) + (\beta_1\lambda_1 L_1 + \beta_2\lambda_2 L_2 + \gamma\eta[L_1+L_2]^2)\Delta t \quad , \tag{4.17}$$

where L_1 and L_2 are the number of Type I and II damages present at time $t+\Delta t$, respectively. The number of lethal damages at time t are calculated using Eqs.(4.15)-(4.17) for every time step $t+\Delta t$, until the desired repair time t_{rep} is achieved.

The survival probability of each cell can be calculated using $S = \exp(-\overline{L_f})$ and compared with a uniform random number to decide whether the given cell is dead or alive. Then the survival fraction is calculated by taking the ratio of the number of cells that survive and the total number of cells in the simulation N_0. We then calculate the mean of 10 repetitions of the repair and survival fraction calculations. This same approach is adopted in this study. The MATLAB code used to perform these calculations is adapted from Appendix D in [87]. It is important to note that this script contains an error in the expressions of L_1 and L_2 and it has been corrected for use in our studies.

4.4.2 Methods of DNA Repair

The TLK model parameters do not explicitly model the DNA repair processes described in Chapter 2 (NHEJ and HR). Instead, the biological repair mechanisms are sorted into first- and second-order processes where a first-order process describes two corresponding broken DNA strand ends rejoining and a second-order process describes the rejoining of ends from different DNA strands, Fig.(4.24). This is because the probability of each rejoining process can change with Type I and II damage yield. It is known that the damage yield changes with dose and particle LET, suggesting that the binary repair parameter η should have particle LET dependence due to the increased probability of interaction between individual DNA damage clusters. Conclusively, whilst the repair pathways are not modelled individually in the TLK model, they can be linked to the defined repair probability parameters.

4.5 Assumptions Made in the Current Study

As with many models across all fields of science, the current model accompanies a series of assumptions. Modelling a highly complex, stochastic system such as nanoscale multi-cellular irradiation can be computationally intensive and require a large amount of storage, particularly when measuring multiple biological endpoints, as we do here. This section provides an overview of the assumptions made in the

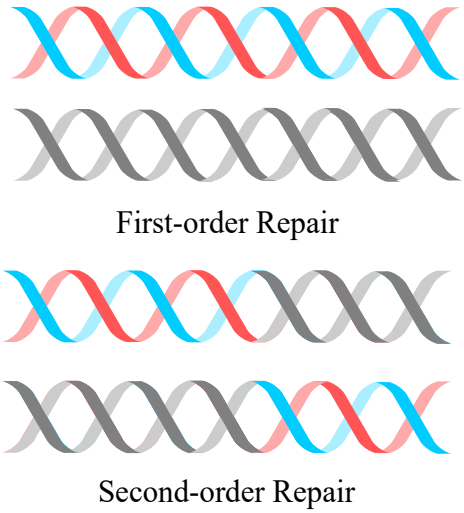


Figure 4.24: Methods of DNA repair accounted for in the TLK model.

current model from which our measurements in Chapter 7 are made.

Our assumptions arise from the geometry, irradiation and DNA repair mechanisms. The cellular geometry is extremely simplified in comparison with a realistic mammalian cell, Fig.(4.4). The morphology of the V79 cell are assumed to be spherical while a realistic cellular geometry should be non-uniform and unique in shape to other V79 cells. The cellular dimensions are also taken to be homogeneous inside the cell culture. Cell morphology is highly dependent on the conditions in which the cells are kept and can be completely different from another group of the same cell line kept in different conditions. The DNA genome is also assumed to be normally distributed due to a wide range of values being reported in the literature [46, 107, 108]. This results in differing response to ionising radiation [110]. In the current work, time permitted only one set of cellular conditions to be simulated.

The cellular response to ionising radiation is assumed to stem from DSB damages to the DNA genome inside the nucleus. This is a common assumption made in radiobiological models, likely the result of optimisation for processing times. We assume that the major source of radiation-induced cell death is DNA damage inside the nucleus, however we have previously discussed that the mitochondria contain small amounts of DNA. Furthermore, the cell cycle phase and other biological conditions can impact the DNA damage yield after irradiation.

The DNA damage repair mechanisms are very complex and additionally are dependent on conditions within the cellular system. The TLK model is the key source for the post-irradiation repair stage of our simulations. The TLK model assumes two DNA damage complexity types. The differential equations used to describe the change in Type I (L_1) and Type II (L_2) damage demonstrate that they should decrease with time, Eqs.(4.15)-(4.17). This is consistent with experiment because

most of the damages will repair when sufficient time is given for them to do so. The equation describing the number of lethal damages present with time, L_f (Eq.(4.13)), suggests that the lethal damage will increase with time. In experiment, one would expect the number of lethal damages to decrease as the Type I and II damages undergo repair. It is therefore a property of the TLK model that the lethal damages are accumulated with time according to the number of Type I and II damages present. Therefore, the experiment and the model do not conflict with each other.

The time during and after irradiation are separated completely in the current model. This means that the repair processes are assumed to begin as soon as irradiation has stopped. In experimental conditions, this is not the case because the cell will begin to repair DNA lesions immediately after they are induced. Ideally, one could run the irradiation and damage repair simulations concurrently and the results can be collected over a single time scale. However, the MATLAB and TOPAS scripts could not be run concurrently. This could be made possible in future work.

4.6 High-Performance and Parallel Computing Methods for Particle Transport Simulations

For computation-intensive simulations such as that described in this chapter, it is often beneficial to utilise high performance computing (HPC) infrastructure and parallel computing processes to improve performance. When performing simulations involving multiple processes that are independent from each other, parallel processing can significantly decrease processing times by distributing these individual processes among multiple workers (CPU or threads) to run simultaneously, Fig.(4.25). This is therefore highly applicable to particle transport simulations.

The multi-threaded capabilities of Geant4 and TOPAS allow individual particle events to process over a number of threads selected by the user. To enable this option in a TOPAS simulation, the user simply needs to include the command `i:Ts/NumberOfThreads = n` where n is the number of requested workers. While logic may tell us that increasing the number of threads will always result in code speed up, this is not always the case. As a benchmark example, identical TOPAS simulations were performed with the parameters outlined in Tables (4.6) and (4.7) (with mean proton energy of 2MeV and 200,000 primary particles) on a PC. The PC used was an Intel i7-8700 (3.20 GHz, 6 CPU cores and 12 threads) and the simulations were run identically across 1 to 12 threads. The CPU time as a function of the number of threads is shown in Fig.(4.26) and indicates an increase in processing time when more than 6 threads are used. Therefore, in this situation the optimal number of threads to use is in the range of 1 to 6. Recalling that the PC used for these simulation contains 6 cores, it is possible that the optimal number of threads

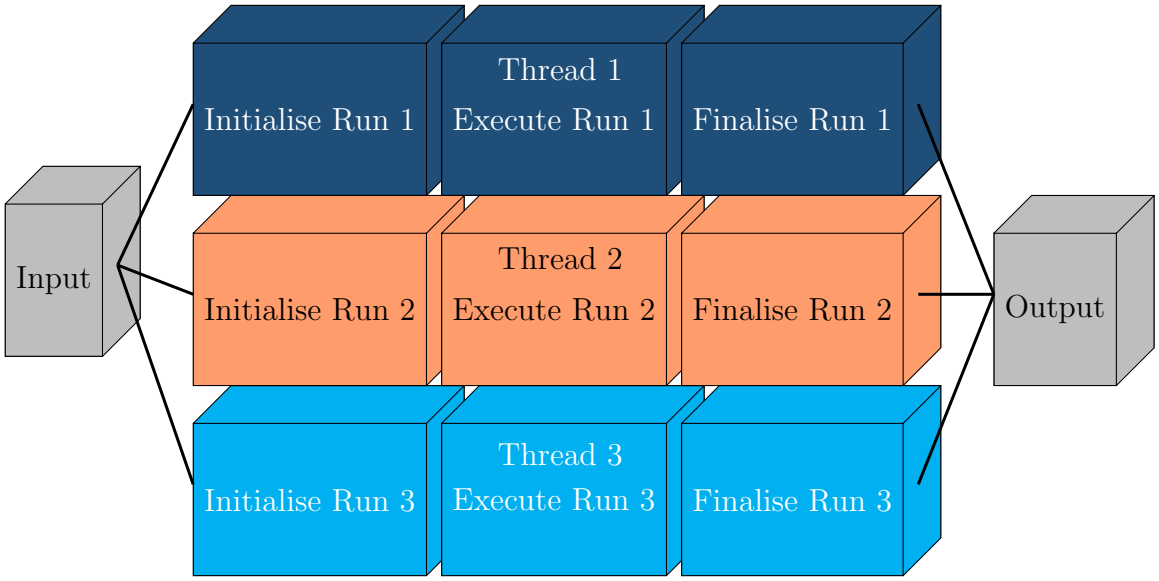


Figure 4.25: The concept of multi-threading across 3 workers applied to a Geant4/ TOPAS simulation.

to request per simulation should be within the number of CPU cores available. A similar method is to be applied when running simulations on different machines.

The effect of this speed up can be further enhanced using a HPC cluster which allow multiple software with high memory and CPU requirements to run simultaneously. These simulations can be submitted to a HPC in the form of batch jobs. On a shared HPC service such as the HPC1 computing cluster at the University of Adelaide (used to generate the results in this thesis), jobs are submitted to a queue with a series of parameters representing memory and CPU requirements for the job, Fig.(4.27). The job is prioritised among other jobs in the queue according to the relative number of resources requested and what is currently available on the cluster. Geant4 and TOPAS simulations can benefit from this infrustructure and with the large number of resources that can be requested, multiple simulations and repetitions can be run simultaneously.

In the simulations performed in this thesis, TOPAS was run on HPC1 in multi-threaded mode. As multi-threaded simulations do not operate in the same way in a HPC environment as they do on a PC when more than one node is used, a single node was requested for all simulations. The number of cores was dependent on the size of the job, therefore high dose points required more cores than low dose points. To avoid the issue shown in Fig.(4.26), the command `OMP_NUM_THREADS=$SLURM_NTASKS` was included in the batch job script so that the number of threads requested in the TOPAS parameter file was always equal to the number of cores requested. This ensures optimal job performance in a HPC environment.

Overall the memory and processing requirements for this study exceeded 100,000 core hours with over 5,000GB of memory in total. The use of HPC techniques made this possible.

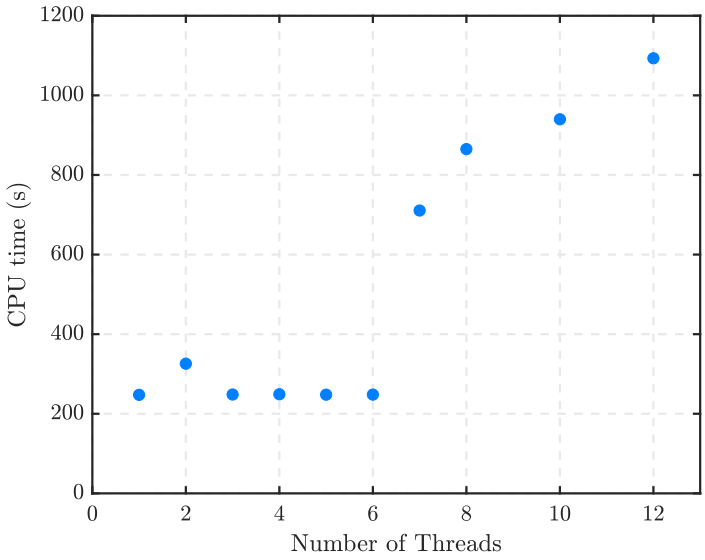


Figure 4.26: CPU time of identical simulations run on an Intel i7-8700 3.20 GHz 6 core (12 thread) PC as a function of the number of workers.

```
#!/bin/bash

#SBATCH -p batch                # Partition (the queue your job
    will be added to )
#SBATCH -N 1                    # Number of nodes
#SBATCH -n 15                   # Number of cores
#SBATCH --time=24:00:00         # Time allocation, D-HH:MM
#SBATCH --mem=25GB              # Memory pool for all cores

#SBATCH --mail-type=END         # Send email if job completed
#SBATCH --mail-type=FAIL        # Send email if job fails
#SBATCH --mail-user=melissa.mcintyre@adelaide.edu.au    # Email to
    which notification will be sent
```

Figure 4.27: Commands to be entered at the beginning of any bash script when submitting to a Slurm manager.

4.7 Overview

In this section, an overview of the parameters selected for the collection of results in this thesis is given in Tables (4.6) and (4.7) along with any concluding statements about the overall performance of the model.

The track structure information is critical to the accuracy of results. Methods of optimisation are limited in nano-scale simulations. The key optimisation method used in the irradiation stage was the implementation of region-based Geant4 physics models where the most detailed track structures were simulated only in organelles containing DNA. A minimum step size was not implemented in these simulations due to the scale of the simulation. A maximum step size of 1mm was implemented to ensure the accurate tracking of particles on their path toward the cell culture. The chosen irradiation conditions are summarised in Table (4.6).

Parameter Name	**Parameter Value**
Particle	Proton
Mean Energies	1.6 to 50MeV
Energy Spread	Gaussian (0.25MeV FWHM)
Field Size	Square (820 $\times$ 820 μm)
Dose Rate	60 seconds per Gy
Lower Production Cut	250 eV
Maximum Step Size	1 mm
Minimum Step Size	None
Physics Lists	Geant4-DNA and Livermore

Table 4.6: Radiation Exposure and Tracking Parameters of our simulations.

The geometry of the cell was an area with plenty of optimisation potential. In this stage, the resultant ionisation yield inside the nucleus was prioritised over other organelles due to the DNA content and its influence on cell death. Spherical cells were used to save computation time during the TOPAS' "overlap checking" phase due to their uniform shape. The endoplasmic reticulum, mitochondria and cell membrane were removed because they have a negligible contribution to the dose delivered to the nucleus when removed. The simulations performed in this thesis are focused on the V79 Chinese Hamster cell line. The properties of the cellular geometry used in subsequent simulations is outlined in Table (4.7).

Parameter Name	Parameter Value
Cell Container Size	0.8 × 0.8mm (Monolayer)
Number of Cells in culture	530
Cell Origin	Chinese Hamster (V79)
Cytoplasm Composition	Liquid Water (1.0g/cm^3)
Nucleus Composition	Liquid Water (1.0g/cm^3)
Nucleolus Composition	Liquid Water (1.1g/cm^3)
Cytoplasm Radius	8.5μm
Nucleus Radius	7.0μm
Nucleolus Radius	5.0μm

Table 4.7: Geometry Parameters of our simulations.

5

A Statistical Analysis of Experimental Hadron Therapy Data

Mathematical modelling is applied to explain phenomena taking place in diverse fields including science, engineering and finance. Such models are formulated using a series of observations made from a system in response to an independent variable. The verification and accuracy of these models are assessed using statistical analysis. It is the quality and rigour of the analysis that determines the model's effectiveness of describing the system in question. Explaining complex systems using models in the medical field is of particular importance since it has a big impact on human health.

In the current study our aim is to model the response of a multi-cellular system after exposure to increasing doses of ionising proton radiation. Chapter 2 outlines how the physical, chemical and biological systems involved influence cell response. Our analysis utilises a continuous random variable D representing the absorbed dose to measure the cellular response in the form of its survival probability.

The Linear-Quadratic (LQ) model is one of the most frequently used models in radiation research and hadron therapy centres to explain *in vivo*[1] or *in vitro*[2] cellular response which includes physical and biological effects described in previous chapters. The model's prevalence in radiation research is possibly owed to its simplicity and applicability for specific combinations of parameters including cell environment and radiation exposure conditions. Simultaneously, this can result in the model appearing to under-perform because **the results only describe cellular response**

[1] *In vivo* experiments are performed with or inside of a living organism.
[2] *In vitro* experiments are performed outside of a living organism - including cell culture experiments.

under a single combination of conditions whilst lacking predictability [30].

In this chapter we perform a rigorous statistical analysis of fits using the LQ model on experimental dose response data summarised in Appendix D, using an extensive regression analysis. We begin by describing the mathematical formalism of cell response to varying doses of ionising radiation used extensively in the literature in terms of the Poisson process. After a description of the statistical techniques used in our study, we test LQ model fits to the experimental data using simple linear and non-linear regression methods. We achieve this by testing the fits to the experimental world data against the assumptions of regression analysis under various biological and exposure conditions (the exact conditions considered are outlined in coming sections). As a result we arrive at the following conclusions :

- The goodness-of-fit of the LQ model to the data is not rigorously determined
- The residual errors are not Normally distributed which is a requirement for χ^2 and other analyses
- The "assumptions of regression analysis" are not met
- There is a high degree of correlation in the data
- The validity of the LQ model is restricted to certain conditions

5.1 The Statistics Behind the LQ Model

The focal point of early radiobiological models were single target models for photon radiation and these simple foundations are still utilised in models today (see Section (2.2.3)). The foundations of the LQ model stem from Radiation Action and Target Theories developed by Chadwick and Leenhouts [35] and Kellerer and Rossi [23]. Such theories describe a system where individual cells or DNA segments are radiation sensitive "targets" which are characterised as being "hit" or "not hit" by an incoming ionising particle during or after irradiation. This formalism is based on several postulates;

- The lethality of damage induced in or on the target influences the resultant response of the system. Lethality is classified according to the number of cells that exist in one of three states [111] :
  - containing lethal lesions - are irreparable and will influence cell death,
  - containing potentially-lethal lesions - can influence cell death if repaired incorrectly
  - and containing repairable lesions - can repair correctly.

  The survival probability of a system of targets is correlated with the fraction that exist in the first state, i.e. they contain lethal lesions.

- Given sufficient repair time, a majority of the lesions induced will repair. However, some lesions are too complex to undergo repair and become lethal. Even a small number of lethal lesions inside a cell are capable of inducing death.
- The process of damage induction and repair is stochastic in nature.

Overall, the targets of our system follow a two binary outcome where they are characterised as containing or not containing a lethal lesion after a given repair time has elapsed. The lethality postulate is described using an i.i.d. (independently identically distributed) binary random variable $X^{(p)}$. The probabilities of a target containing and not containing a lethal lesion is given by

$$\begin{aligned} P(X^{(p)} = 1) &= p, \\ P(X^{(p)} = 0) &= 1 - p, \end{aligned} \tag{5.1}$$

respectively, where p is a positive number $0< p <1$. If we take a sample of N random variables $X_1^{(p)}, X_2^{(p)}, \cdots, X_N^{(p)}$ the resultant number of hits can be written as

$$S_N^{(p)} = X_1^{(p)} + X_2^{(p)} + X_3^{(p)} + ... + X_N^{(p)} = \sum_{i=1}^{N} X_i^{(p)}. \tag{5.2}$$

This is the mathematical definition of a Bernoulli experiment and is a random variable in itself taking values between 0 and N. Now if we introduce a variable k such that $0 \leq k \leq N$, then the probability that the sum takes the value k is given by the so called binomial probability distribution defined as

$$P(X_1^{(p)} + X_2^{(p)} + X_3^{(p)} + ... + X_N^{(p)} = k) = \binom{N}{k} p^k (1-p)^{N-k}. \tag{5.3}$$

Let us first take a sequence of positive numbers $p_N \to 0^+$ as $N \to \infty$ with $N.p_N = \lambda$. Then we perform the limit $N \to \infty$ in Eq.(5.3)

$$\binom{N}{k} p_N^k (1-p_N)^{N-k} \Rightarrow \frac{(Np_N)^k}{k!} \frac{(N)_k}{N^k} (1-p_N)^{-k} \left(1 - \frac{Np_N}{N}\right)^N \to \frac{\lambda^k}{k!} e^{-\lambda}, \tag{5.4}$$

as $N \to \infty$ where $(N)_k = N.(N-1)...(N-(k-1))$. The limit resulting in Eq.(5.4) is the probability distribution function for a Poisson random variable. S_N converges to S as $N \to \infty$ where the limiting random variables have the distribution Eq.(5.4) namely for any integer number k.

Finally, we can rewrite Eq.(5.3) in its simpler form

$$P(\lambda, S_N = k) = \frac{\lambda^k}{k!} e^{-\lambda} \quad . \tag{5.5}$$

In our physical picture we may think of a random experiment such as radiating a target with particles repeated infinitely often and adjusting the success probability p_N to the limit $N \to \infty$ by keeping $N.p_N = \lambda$ fixed, and then finally ask for the outcome that S takes on the integer value k. The survival probability is given by the integer value $k = 0$ order, namely targets that were not hit by a particle.

$$P(S = 0) = \frac{\lambda^0}{0!} e^{-\lambda} = e^{-\lambda}. \tag{5.6}$$

Once we determine the only parameter λ from this survival probability, all other probabilities for $k > 0$ are given by Eq.(5.5).

We can relate λ to the dose absorbed by the target. Cellular response of a target to radiation is stochastic in nature and is described by the Poissonian process. Firstly, consider the total number of targets N_0 - assuming no "hits" exist prior to irradiation (i.e. $S = 1$) - and the targets left without "hits" after irradiation N. The survival probability for a target after absorbing dose D is

$$P(D/D_0 = \lambda, 0) = \frac{N}{N_0} = e^{-\frac{D}{D_0}} = e^{-\lambda} \quad , \tag{5.7}$$

where D_0 is the dose required to achieve an average of one hit per target and λ is the number of lethal and potentially-lethal lesions per target. The probability of a single target being hit is then described by

$$1 - P(D/D_0, 0) = 1 - e^{-\frac{D}{D_0}} = 1 - e^{-\lambda} \quad , \tag{5.8}$$

where λ are the number of lethal lesions present per target. A survival curve is constructed using the $k = 0$ order probabilities as a function of absorbed dose D as depicted in Fig.(5.1).

To explain hitting multiple targets within a cell to induce cell death, the single target theory in Eq.(5.8) must be expanded to account for multiple targets and hits. This is observed in Fig.(5.2), where the experimental data describing cell survival often encountered deviates from a purely exponential relationship due to the "shoulder" effect present in the low dose region (< 2Gy) (Fig.(5.2)). We can use Eq.(5.8) to derive an expression that gives the probability of a single hit to "m" biological targets and hence the cell's probability of survival,

$$S = 1 - (1 - e^{-\frac{D}{D_0}})^m = 1 - (1 - e^{-\lambda})^m. \tag{5.9}$$

In single hit single target and single hit multiple target theories it is assumed that the survival fraction is independent of dose-rate - i.e. the only dependence is on the total dose absorbed before measuring the survival fraction $SF(D)$ but not on the time interval over which the radiation is delivered. However, in reality this assumption does not hold [112].

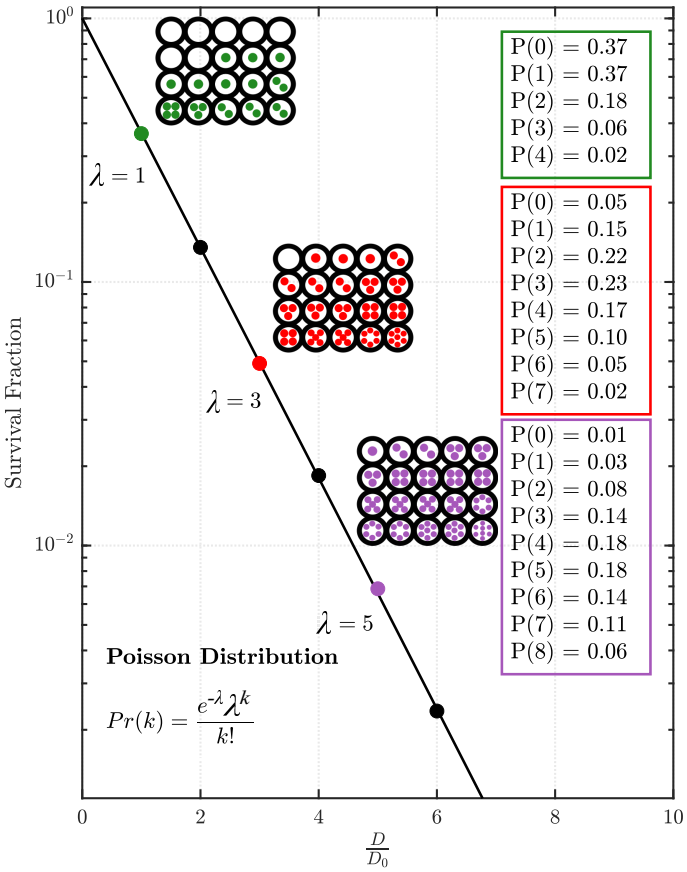


Figure 5.1: A survival fraction constructed from Poisson distribution probabilities, $1 - \exp(-\lambda)$, for the mean number of lethal lesions produced in the system, λ. Graphical depictions of the cell damage probabilities are shown for select SF. Credit to [111] for the original figure idea. Note: probabilities < 0.01 are not shown.

Chadwick and Leenhouts [35] and Kellerer and Rossi [23] arrived at the LQ model by combining single target single hit and multiple target double hit theories[3]

$$S = e^{-\alpha D - \beta D^2} \quad . \tag{5.10}$$

This means that a cell can be damaged by a single lethal hit or by multiple hits which are not lethal individually but can become lethal together [43].

It has been discussed in Chapter 2 that the response of a cell following exposure to ionising radiation is extremely complex when compared with that of tissue response. Cellular response is not only dependent on the ionising radiation itself but is governed by inter-cellular repair mechanisms and conditions. These include, but are not limited to, oxygen concentration, cell cycle phase and availability of nutri-

[3]The D and D^2 terms respectively.

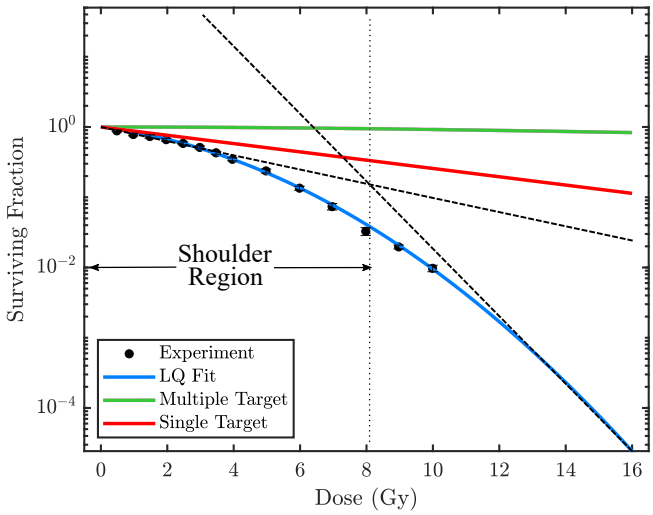


Figure 5.2: A comparison of the single target, multiple target, and LQ fit for clonogenic V79 cell survival data after being irradiated with a 70MeV proton beam from Wouters et al. [113].

ents. Furthermore, cell response is also governed by chemical processes, including the production of hydroxyl radicals which are another pathway to DNA damage leading to cell death. An effective model will account for as many of these processes as possible. Various attempts have been made to improve upon the LQ model by incorporating these processes into α and β [31, 41].

Surely a model can appear to fit a sample well by observation (consider Fig.(5.2) as an example) however a statistical analysis of a given model can paint a very different picture. A series of diagnostics can be employed to quantify the effectiveness with which a model describes the data on a statistical level. Regression analysis is among the most common methods used to perform this analysis.

5.2 An Overview of Regression Analysis

In this study, two types of regression analysis will be considered in performing a rigorous analysis of the experimental data:

1. Simple linear regression (SLR).
2. Non-linear regression (NLR).

To summarise, regression analysis is used to study the behavior of a variable $\mathbf{Y}_i$ in response to a predictor variable $\mathbf{X}_i$ according to a parameter space $\hat{\delta}$ with an error term $\epsilon_\mathbf{i}$ to account for perturbations in the system being modelled. Alternatively, it

is a mathematical description of the correlation between two variables and is given as

$$\mathbf{Y}_i = f(\mathbf{X}_i, \hat{\delta}) + \epsilon_i \quad . \tag{5.11}$$

This is the general definition of a regression model in its simplest form. In some systems the relationships can be more complex and involve more than one independent variable.

We will utilise regression analysis to determine how effectively the LQ model describes dose response data under different biological and exposure conditions. A range of techniques to assess goodness-of-fit are employed and a regression analysis is performed by testing the fits against the assumptions of regression modelling.

The databases of experimental data originates from Prof. Cynthia Keppel and Dr. Pawel Ambrozewicz of the Thomas Jefferson National Accelerator Facility (JLab) and the Particle Irradiation Data Ensemble (PIDE) [46]. The databases contain experimentally obtained dose-response curves with different combinations of physical and biological parameters. In addition to cells across different species of origin, cellular conditions and varying radiosensitivities, the data includes different radiation exposure conditions including:

- radiation type (heavy ions, X-rays and protons),
- primary proton energy,
- beam type (pristine BP or SOBP),
- particle LET,
- depth in medium (H_2O, PMMA and Nylon).

When undertaking the analysis, we must remain mindful of the small sample size of the data. In most cases, 6 observations are made with some containing a minimum of 3 and a maximum of 24 observations. The idea of "large N" is dependent on the application and there is no single rule of thumb for what "N" should be. However, extremely small data sets make it difficult to perform regression analysis and observe trends within the data when assessing the assumptions. For this reason we must be mindful of how we test the data and how we interpret the results.

Analysis performed in this chapter is important to radiobiological modelling because the parameters derived from the experimental LQ fits are dependent on the physical and biological parameters of the system. When these fits are not determined rigorously, quantities such as the relative biological effectiveness (RBE) and oxygen enhancement ratio (OER) (see Chapter 2) cannot be accurately determined, resulting in unreliable conclusions being drawn from the studies. An effective LQ model fit should be determined by considering the following :

- knowledge of all relevant biological and radiation exposure parameters,
- applying effective fitting procedures to derive the fit parameters,
- rigorously testing goodness-of-fit,
- accurately evaluating fit uncertainties.

5.2.1 The Assumptions of a Regression Model

Linear regression modelling is the most common type that assumes a strictly linear response from the predictor variable. However non-linear regression models are also used frequently at times providing an improved fit in compared to linear regression.

More often than not, scientific phenomena are non-linear in nature[4] [114]. Typically, non-linear models exist in two forms : intrinsically linear models and non-linear models that are linearisable. The LQ model falls into the latter category where it is non-linear in its familiar form however we can apply a transformation to linearise the model by taking the logarithm of Eq.(5.10).

$$-\frac{\log(S)}{D} = \alpha + \beta D \quad . \tag{5.12}$$

Thus we have arrived at a linear form of the LQ model which will allow us to perform a rigorous linear regression analysis.

We must assess these fits according to the assumptions of regression analysis, all of which are briefly outlined below with the methods used to test them. The procedure of each test is provided and discussed in more detail in Appendix E. The list of assumptions may be longer in the literature, however the focus of this chapter will be condensed to the four key assumptions discussed below. If these assumptions are not met, then our estimate of the model parameter space is not trustworthy and their standard errors are not reliable.

Briefly, the "assumptions of regression" to be tested and the methods used are the following:

- **Linearity (SLR)/Non-Linearity (NLR)**;
  - Plotting the fit residual errors against the independent variable.
  - Observing the goodness-of-fit metrics.
- **Homoscedasticity (constant variance of residuals)**;
  - Breusch-Pagan test [115],

[4] Growth and decay models are good examples of this.

  - White test [116].

- **No autocorrelation of residuals**;

  - Durbin-Watson test [117] for linear regression.

- **Normality of residuals**;

  - Kolmogorov-Smirnov test [118, 119] using critical values obtained from [120],
  - Anderson-Darling test [121] using the critical values calculated as a function of sample size and significance level [122].

5.3 A Regression Analysis of Experimental Dose-Response Data

Improving hadron therapy treatment requires a deep knowledge of the radiobiological mechanisms involved with cell death. The literature has highlighted a need to diverge from the assumed clinical proton RBE of 1.1 and apply radiobiological weighting to the models used in hadron therapy. The validity of the model against the experimental world data fits reported in the literature should be performed rigorously to determine biological factors such as RBE and OER accurately. It is shown in our results and the conclusions of [17] that the published LQ fits suffer from several deficiencies :

- the goodness-of-fit of the LQ model to the data is not rigorously determined,
- the residual errors are not Normally distributed which is a requirement for χ^2 and other analysis,
- the "assumptions of regression analysis" are not met,
- there is a high degree of correlation in the data,
- the validity of the LQ model is restricted to certain conditions.

We test the ability of the LQ model and the validity. Additionally we investigate how a poor outcome of the LQ model influences the RBE. We obtain the optimal LQ parameters α and β with the best goodness-of-fit achievable by comparing the metrics described in Appendix E with those from the fits in the literature (denoted as $\alpha_{Original}$ and $\beta_{Original}$ throughout this chapter).The method applied to fit Eqs.(5.10) and (5.12) to the data is a **weighted** ordinary least squares (OLS) fitting by solving

$$\min\left[\sum_{i=1}^{n}\frac{(-\frac{\log S_i}{D_i}-\frac{\log \hat{S}_i}{D_i})^2}{\epsilon_i^2}\right] = \min\left[\sum_{i=1}^{n}\frac{\left(-\frac{\log S_i}{D_i}-[\hat{\alpha}D_i+\hat{\beta}D_i^2]\right)^2}{\epsilon_i^2}\right] \quad (5.13)$$

$$\min\left[\sum_{i=1}^{n}\frac{(S_i-\hat{S}_i)^2}{\epsilon_i^2}\right] = \min\left[\sum_{i=1}^{n}\frac{\left(S_i-e^{-\hat{\alpha}D_i-\hat{\beta}D_i^2}\right)^2}{\epsilon_i^2}\right] \quad , \quad (5.14)$$

for the linear and non-linear fits, respectively for $\hat{\alpha}$ and $\hat{\beta}$ where S_i is our i^{th} survival fraction observation with associated "standard error" ϵ_i for each dose point D_i.

The full list of experimental data selected for this study considers the radiation exposure and biological conditions listed in Appendix D. Here we provide a brief description of some of the experiments considered as an example :

- **Belli et al. (1998) [26]:** The RBE-LET relationship is investigated using proton radiation of incident energy 0.57 to 5.01MeV and an LET range of 7.7 to 37.8keV/μm incident on V79 Chinese Hamster cells. The cell cycle distribution is not reported however we can assume that it is asynchronous when the cycle is not stated. The oxygen content is also not specified. The findings showed that the RBE increased with LET until $\sim$ 30.5keV/μm, after which is begins to decrease.

- **Belli et al. (2000) [60]:** Four cell lines of human origin (namely SCC25, SQ20B, M10 and HF19) with varying radiation sensitivity are exposed to proton radiation of incident energies between 0.76 and 5.04MeV and an LET range of 7.7 to 33.0keV/μm. The cell cycle distribution is not stated nor is the oxygen content. In the current study cell cycles are assumed to be asynchronous in phase for this study. The results showed that the SQ20B cell line was the most resistant to γ-radiation but the most sensitive to high-LET proton radiation. In order the M10, SCC25 and HF19 cell lines were shown to decrease in sensitivity to proton radiation.

- **Bettega et al. (1998) [61]:** The RBE-LET relationship was investigated by irradiating the C3H10T12 rodent cell line using protons and deuterons (we do not have access to the deuteron data) of energies ranging from 0.72 to 3.18MeV (and an LET range of 11.0 to 33.2keV/μm). The results supported those of Belli et al. (1998) [26] where the RBE increased with LET. However the LET range is not high enough to reach the peak that is observed in [26].

- **Folkard et al. (1989) [96]:** The RBE is determined for protons of mean incident energy 0.76, 1.15 and 1.9MeV (and an LET range of 17 to 32keV/μm) incident on V79 Chinese Hamster cells. The cells are assumed to be asynchronous in cell cycle phase with no oxygen concentration stated. The results showed an increase in RBE with LET, albeit the peak observed in [26] is not present because the maximum LET in this study is 32keV/μm.

- **Howard et al. (2018) [123]:** The RBE is measured for four depths in water with pristine 71 and 160MeV Bragg Peaks incident on one rodent and two human cell lines (CHO, A549 and T98 respectively). The depths correspond to an LET range of 0.99 to 7.34keV/μm. The results show an increase in RBE with LET and that the two human cell lines had higher RBE values than CHO cells.

- **Wouters et al. (2015) [38]:** The RBE is measured for V79 Chinese Hamster cells exposed to two modulated proton SOBPs of energy 160 and 230MeV (and a LET range of 1.03 to 4.75keV/μm) at five depths each in water. The results show a dose dependence on RBE where it decreases in the high dose region. As a function of depth in the SOBP RBE was found to increase with LET.

In order to carry out discussions in this chapter, we utilised the above experimental data, however our complete analysis results over all the experimental world data in Appendix D available can be found in Appendices F, G and H. The performance of the LQ model is investigated for varying LET using the data of Belli et al. [26], Bettega et al. [61], Baek et al. [124], Gueulette et al. [125] and Folkard et al. [96]. A discussion on the behavior of different cell lines and tissues by comparing Howard et al. [123] and Belli et al. [60] is performed. Finally the behavior for depth studies is investigated by comparing Howard et al. [123] and Wouters et al. [38].

The following sections address the assumptions from Section (5.2.1) individually using the methods described for our linear and non-linear regression fits respectively (the methods are described in Appendix E).

5.3.1 Goodness-of-Fit Metrics and Linear Regression Analysis on the LQ Model

In this section we perform a linear regression analysis and goodness-of-fit (gof) metrics on experimental data listed in Appendix D by fitting to the linearised LQ model Eq.(5.12).

5.3.1.1 GOF Metrics

To commence the analysis of our linear regression fits, we consider the dose as a function of survival fraction (SF) as reported in the literature of the experimental world data (listed in Appendix D) with the transformation $SF \rightarrow -\log(SF)/D$. The errors on the survival measurements are reported as the standard error of the mean (SEM) in the literature. We perform the fit by solving Eq.(5.14) using the least square method. As our benchmark study, we present the fits from Belli et al. [26] in the current chapter (see Figs.(5.3)–(5.5)) and select fits from other studies in Appendix H to aid discussion. These figures depict the linear fit from Eq.(5.12) parameter estimations and gof metrics, however the fit itself is presented as absorbed

dose versus survival on a logarithmic scale. We present our fits in this way, despite fitting a linear function, because the transformation $SF \rightarrow -\log(SF)/D$ results in extremely large errors which do not accurately represent the data.

We have reported the fitting parameters from the original studies for comparison with the fits performed in this study in Appendix F. Our observations that the fits reported in the literature agrees with the regression study of Prof. Cynthia Keppel and Dr. Pawel Ambrozewicz from JLab [17] from which the following conclusions can be drawn :

- the goodness-of-fit of the LQ model to the experimental data is not determined rigorously,
- the "assumptions of regression analysis" are not satisfied,
- the fit residual errors are not Normally distributed, which is a requirement of the χ^2 test,
- the fits to the data are highly correlated,
- the applicability of the LQ model is restricted to a series of conditions.

Firstly, let us consider Figs.(5.3)–(5.5) which depict the linear regression fits to the data of Belli et al. [26]. Moving systematically, the low-LET region, Fig.(5.3), yields good agreement with the LQ model and our fit is an improvement on the published fits (the χ^2 value is lower in all our fits). Figs.(5.3(a)) and (5.3(b)) possess the expected "shoulder region" shown in Fig.(5.2) with an increase in cell killing efficiency for higher doses. The gof metrics suggest that the model fits well within the error bars ($\chi^2 = 0.459$ and 0.926, respectively). Although, when we calculate $-\log(SF)/D$ the variance in the survival measurements begin to show, which is reflected in the R^2 values of 0.716 and 0.770 for Figs.(5.3(a)) and (5.3(b)), respectively. The RMSE indicates that the Euclidean distances between the observations and predictions are quite small in Fig.(5.3(a)) but slightly larger in Fig.(5.3(b)) due to the presence of an outlier (the sixth dose point).

Additional plots are presented in Section (H.1) of Appendix H. Figures (H.1(a)), (H.1(b)) and (H.2(a)) depict fits to data from Wouters et al. [38] and Howard et al. [123] in the low-LET region. Figure (H.1(a)) yields an almost perfect fit with little deviation in the residuals ($\chi^2 =< 0.001$, $R^2 = 0.99$, RMSE≈ 0 and mean residual error $= -5.4 \times 10^{-6}$) for an LET of 1.1keV/μm. Figure (H.1(b)) also yields good fit to the linearised LQ model ($\chi^2 = 0.36$, $R^2 = 0.95$, RMSE $= 0.007$ and mean residual $= 0.003$) for an LET of 2.26keV/μm. Considering this and the χ^2 values supplied in Table (F.1), the linearised LQ model appears to yield a good fit to the data in the low-LET region across different cell lines. Figure (H.2(a)) also fits very well ($\chi^2 = 0.36$, $R^2 = 0.95$, RMSE $= 0.03$ and mean residual $= 0.003$) for an LET of 4.35keV/μm. Although the small sample size of the data makes it difficult

to observe behavior in the residual errors, there is a larger scatter about the mean (mean residual error = 0.008) compared to Fig.(H.1(a)).

Next we will consider Figs.(5.4(a)) and (5.4(b)), the intermediate LET region. In Fig.(5.4(a)) the "shoulder behavior" discussed above is still present in the data and the fit is still very reasonable. The fit is within all the error bars ($\chi^2 = 0.399$) and most of the variance in the observations is accounted for in the fit (R^2=0.818). The good fit is further reflected in the Euclidean distances between the fit and the observations which is small (RMSE = 0.631). However, as we increase the LET to 30.5keV/μm in Fig.(5.4(b)), the behavior begins to change. The LQ parameter β decreases to $\sim$0, resulting in a purely exponential fit (only contributions from α). The large χ^2 value in comparison to lower LET is large, indicating that the LQ model does not fit well within the error bars. There is a high degree of variance not accounted for by the LQ model ($R^2 = 0.001$) and is further reflected in the high RMSE value of 1.341. Furthermore, we note that the adjusted R^2 is negative in Fig.(5.4(b)). This suggests that the linearised LQ model performs poorly at explaining dose-response data in the high-LET region.

In Fig.(H.3) we fit the linearised LQ model to the data of Bettega et al. [61] (C3H10T12 cells, 3.18MeV with LET 11keV/μm) and Belli et al [60] (HF19 cells, 1.49MeV with LET 19.6keV/μm) the relationship demonstrates the same linear behavior in the dose response curve on the logarithmic scale as for Figs.(5.4) with similar χ^2, R^2 and RMSE metrics. Although we acknowledge the behavior of $\beta \to 0$ is observed at a higher LET for the V79 cell line than those considered here. This can be the result of differing radio-sensitivities for each cell line due to differing intrinsic biological parameters relating to the cell environment.

Finally in Fig.(5.5), the high-LET region yields the worst fit of the LQ model where the LQ parameter β becomes negative. Furthermore, when $\beta < 0$ the argument becomes negative and therefore the function turns from concave to convex (see Fig.(5.5)). In Fig.(5.5(a)), the χ^2 value is greater than 1 (1.999) and the variance of the observations is not well explained by the model, $R^2 = 0.565$. The highest LET of our benchmark study, depicted in Fig.(5.5(b)), contains an dose point that impacts the resulting χ^2 (the 5^{th} point). Following close analysis of this data set, we have reason to believe that this points is an outlier. Therefore, we performed the fit with this point omitted, although we still present it in Fig.(5.5(b)). By omitting the outlier, we reduced the χ^2 from 2.100 to 1.746. This is further evidenced by the high RMSE value of 1.414. This trend continues in Fig.(5.5(b)) where the χ^2 is now 2.1 and less than half of the data variance is adequately explained by the LQ model $R^2 = 0.42$. Furthermore, the RMSE has increased to 1.449, indicating a large error in the fitting. If we observe the fits themselves, Fig.(5.5) shows that its curvature inverts and that the LQ argument becomes a convex decreasing function.

In Fig.(H.4) we fit the linearised LQ model to the data of Belli et al. [60] (SQ20B cells, 0.88MeV with LET 30keV/μm) and Folkard et al. [96] (V79 cells, 0.76MeV with LET 32keV/μm). The gof demonstrates the same behavior in the dose response curve on the logarithmic scale as for Fig.(5.4(b)) which corresponds to V79 response at an LET of 30.5keV/μm. We expect the behavior in Figs.(5.4(b)) and (H.4(a)) because they correspond to the same cell line under similar exposure conditions. The fits themselves are mostly exponential and result in large χ^2 and RMSE values as expected. This shows similarities in the V79 cell response across different studies, thus supporting our observation that the effectiveness of the LQ is dependent on the cell line. With radiation of LET 30keV/μm incident on the SQ20B cell line, Fig.(H.4(b)) demonstrates a strong loss in cell killing efficiency with β becoming negative. However, the gof metrics indicate a reasonable fit ($\chi^2 = 0.897$, $R^2 = 0.9$ and RMSE = 0.947). This alone may suggest the linear model is a good fit, however the residual plot shows that there is a high degree of scatter about the fit. We will expand on this in the next section where we discuss the assumption of linearity.

In all the fits presented in Appendix F we found that the χ^2 value was smaller (better) for our fits than those reported in the literature. The extent to which the fit is improved is depicted in Fig.(5.6) for some of the world data, where the LQ parameter estimates $\hat{\alpha}$ and $\hat{\beta}$ are plotted against each other with their standard deviations (1σ, 2σ and 3σ) presented as an ellipse to show the error in the estimates. The parameter estimations from the literature are also plotted and the result of the χ^2 test is presented for our fits and the published fits with the correlation between α and β. Fig.(5.6) shows that in many cases the published LQ fit parameters differ from those performed in this analysis by more than one standard deviation, 1σ. This indicates that in most cases, the published results can be improved and should be analysed rigorously for statistical anomalies as we do in this section. The fits to the remaining experimental world data in Appendix F demonstrate the same depleting goodness-of-fit of the LQ model in the high-LET region as we do for Belli et al. [26] (Guan et al. [27] and Belli et al. [60] are good examples of this for different cell lines). Logically, the quality of the data can also impact the goodness-of-fit. When outliers are present, the gof metrics fluctuate. The data of Bettega et al. [61] consists of outliers in their low-LET dose-response measurements (see Fig.(H.7(a))) which is reflected in the gof metrics.

All the fits that are presented in this investigation demonstrate a high degree of correlation (>90%) between the LQ model parameters, α and β, Fig.(5.6). This suggests they are not completely independent of each other. Recall from Chapter 2 that the LET is related to the LQ parameters by α/β (high α/β corresponds to low LET and low α/β to high LET). There are clearly multiple α and β values that give the same measure of α/β. This could possibly explain the observed α–β parameter correlation. We also note that the α–β correlation is negative implying that as α increases, β will decrease because the number of single hits will decrease

as the number of multiple hits increases. The α–β correlation is determined from the cross terms in the covariance matrix.

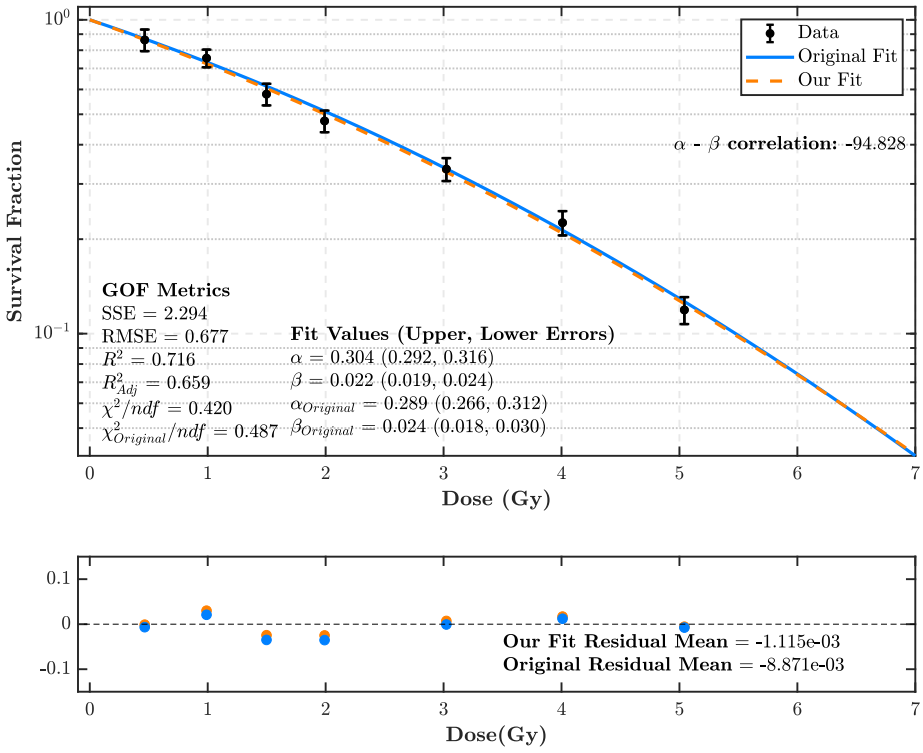
Overall, the linearised LQ model yields a reasonable fit in the low-LET region where the gof metrics are the best. In the mid-LET range, the gof metrics indicate a worsened fit compared to low-LET. However, in the high-LET region the gof metrics suggest a very poor fit of the linearised LQ model to the data.

5.3.1.2 Testing for the Assumption of Linearity

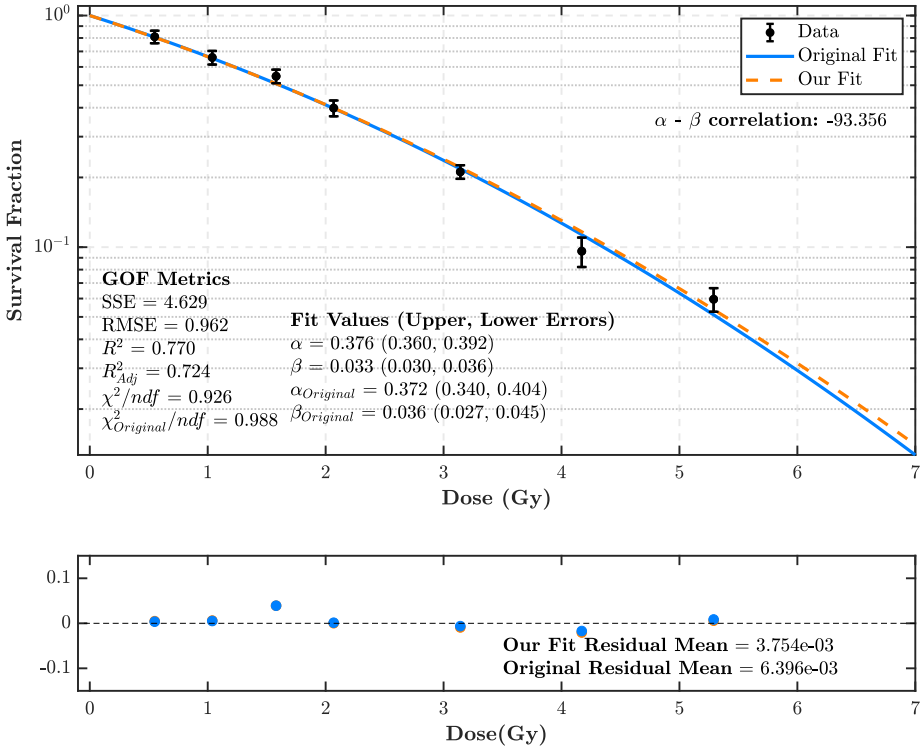
The linearity test assesses the functional form of the fit model, namely if the true relationship is linear. The most effective method of testing the experimental data against the regression assumption of **linearity** is to produce residual plots (below Figs.(5.3)–(5.5)) and to analyse the gof metrics. Due to the large number of data sets considered here, not all plots are presented in this section. However, we have provided all the fit results performed on the experimental data in Appendices F and H. The basic information on the cellular and exposure conditions for each study is provided in the captions of each figure and Appendix F, however more detailed information on each experiment is provided in Appendix D. It is difficult to make direct comparisons between cell line responses across different experiments because cell culture preparation techniques can vary and affect the response [25]. Belli et al. (1998 and 2000) [26, 60] are good candidates to make this comparison because the cell culture methods and radiation exposure techniques are similar and multiple cell lines are studied. In the previous section we discussed the gof metrics which we can incorporate into our discussion of linearity as our fit function, Eq.(5.12), is linear.

Let us revisit Figs.(5.3)–(5.5) which depict our linearised LQ model fits to the data of Belli et al. [26] with the residual plots presented below each figure. If the LQ model can adequately explain the data, performing the transformation $SF \rightarrow -\log(SF)/D$ should yield an increasing linear relationship with absorbed dose. In turn, the residual errors between the observations and fit should be random fluctuations about the fit line.

Working systematically, the LQ fit to the lowest LET data of Belli et al. [26] depicted in Fig.(5.3) yields a very good fit to each data set as discussed in the previous section with respect to the gof metrics. Visually, the fit follows the expected trend of a decreasing concave function when presented as survival versus dose on the logarithmic scale. Therefore, when linearising the dose-response relationship we can expect the data to closely follow an increasing linear relationship. The residual plots below Figs.(5.3(a)) and (5.3(b)) further evidences the strong linear relationship of this data set with very small, random fluctuations about zero and no apparent trend with absorbed dose.

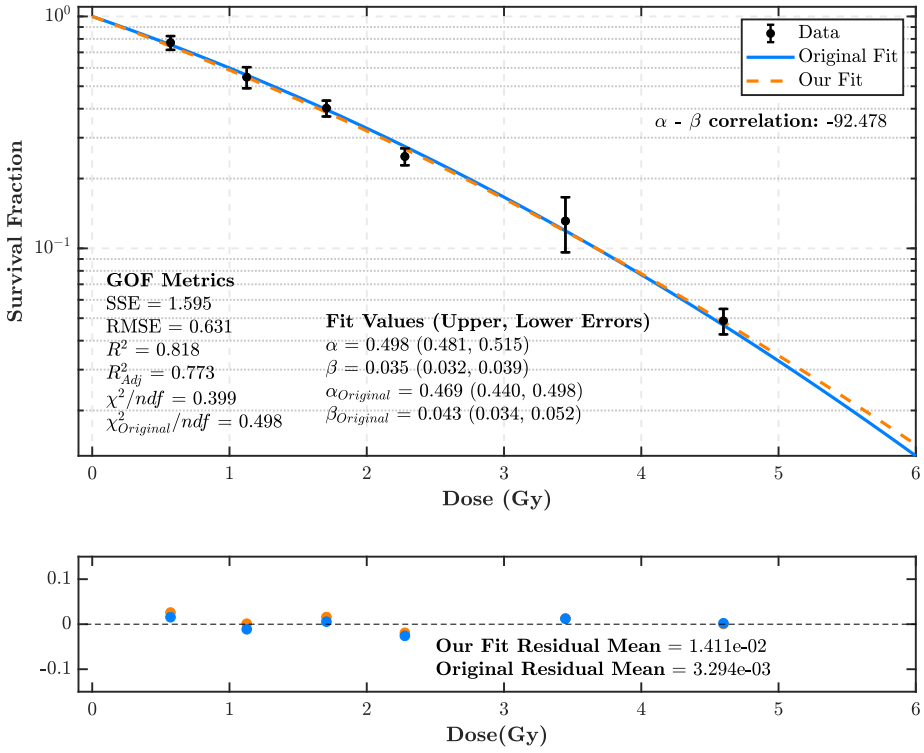


(a) 5.01MeV (LET = 7.7 keV/μm)

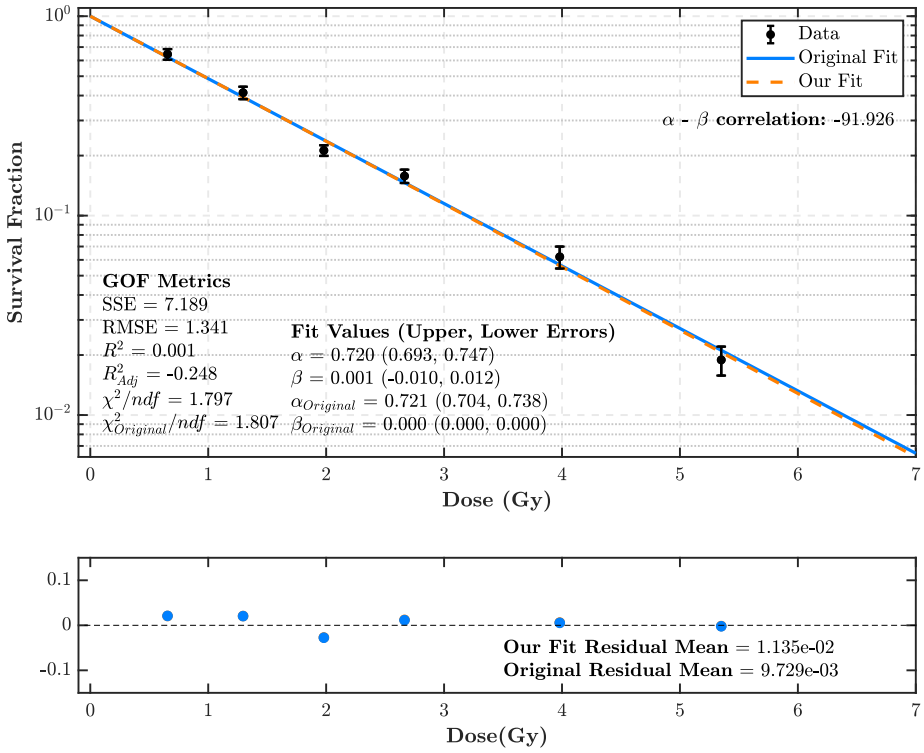


(b) 3.20MeV (LET = 11.0 keV/μm)

Figure 5.3: Linearised LQ fits on data from Belli et al. (1998) [26] with protons of varying energy incident on V79 cells. The second panel is a plot of the residuals between the observations and the linear LQ model.

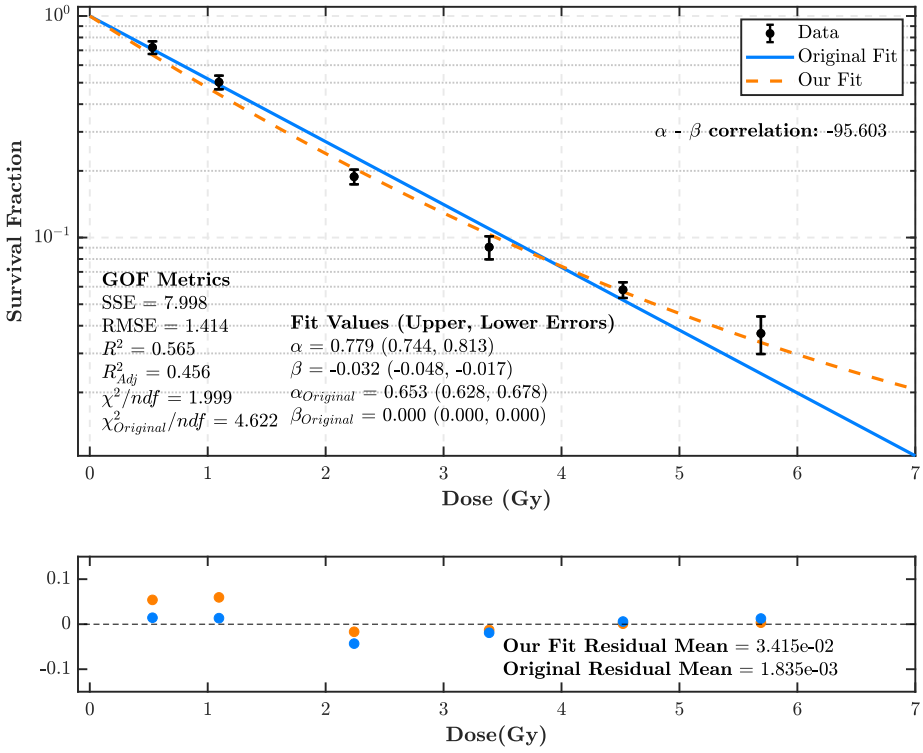


(a) 1.41MeV (LET = 20.0 keV/μm)

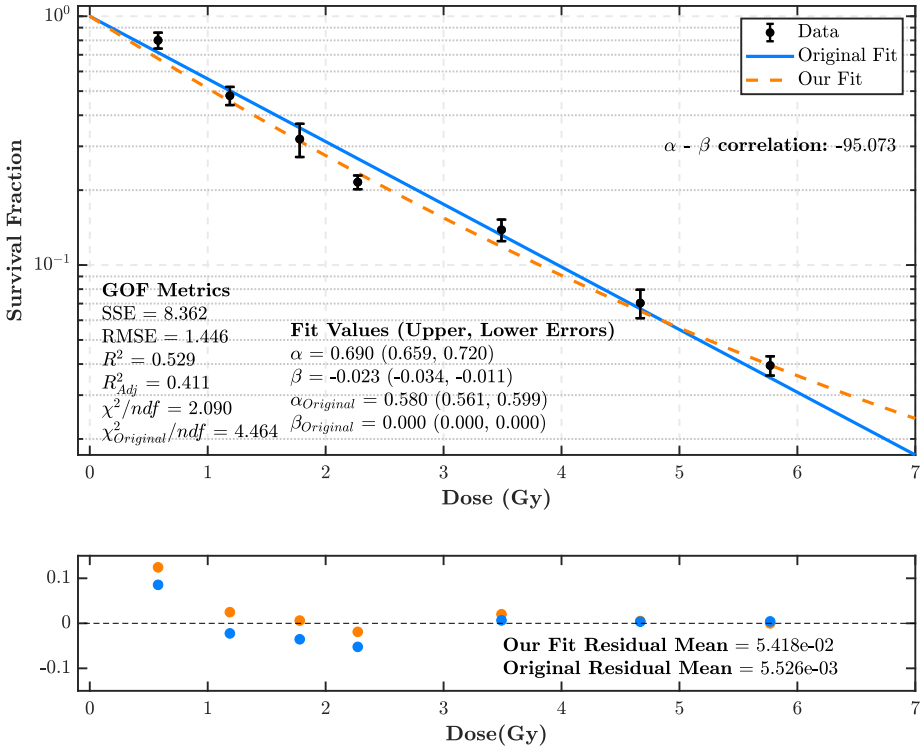


(b) 0.76MeV (LET = 30.5 keV/μm)

Figure 5.4: Linearised LQ fits on data from Belli et al. (1998) [26] with protons of varying energy incident on V79 cells. The second panel is a plot of the residuals between the observations and the linear LQ model.

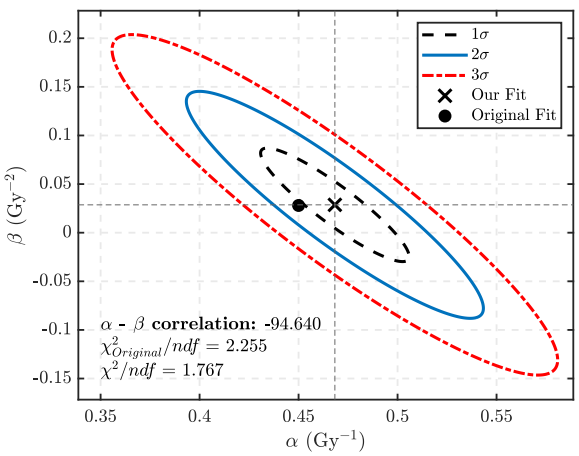


(a) 0.64MeV (LET = 34.6 keV/μm)

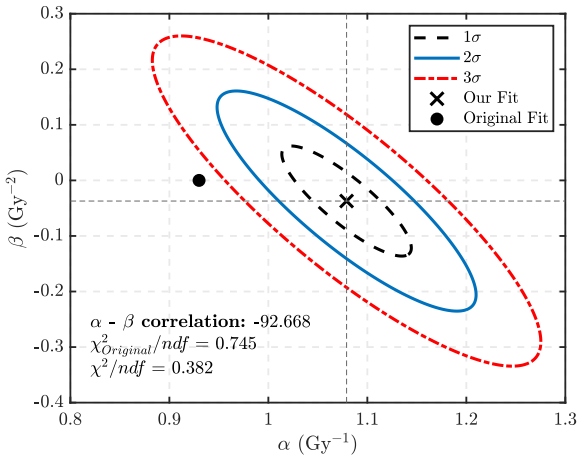


(b) 0.57MeV (LET = 37.8 keV/μm)

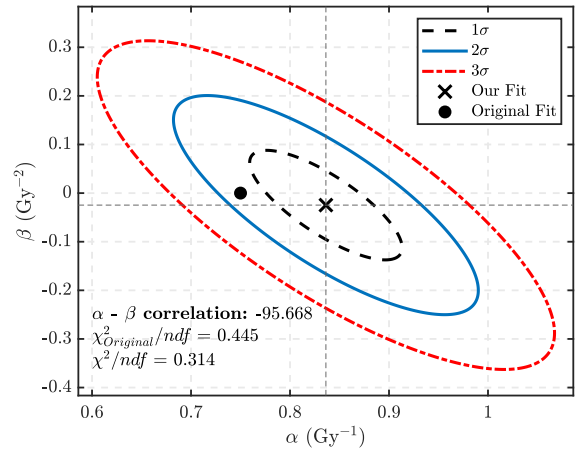
Figure 5.5: Linearised LQ fits on data from Belli et al. (1998) [26] with protons of varying energy incident on V79 cells. The second panel is a plot of the residuals between the observations and the linear LQ model.



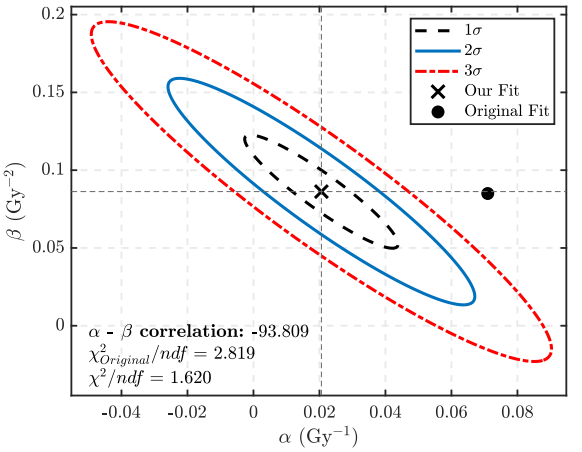
(a) Folkard et al. (1996) [126]: 1.83 MeV protons 18 keV/μm incident on V79 cells.



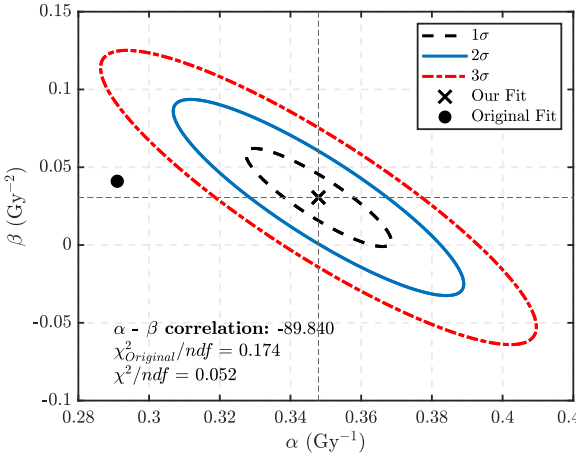
(b) Belli et al. (2000) [60]: 0.79 MeV protons 33.0 keV/μm incident on M10† cells.



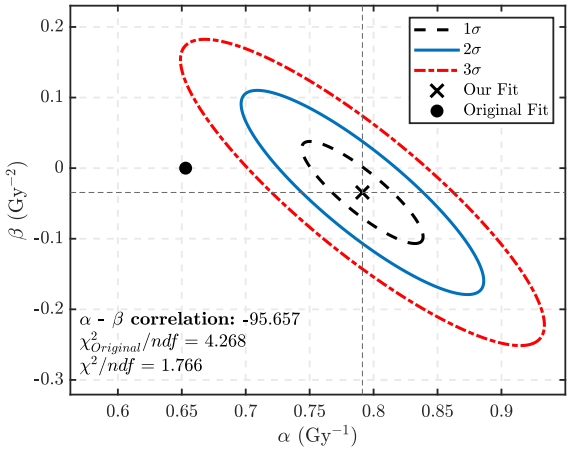
(c) Bettega et al. (1998) [61]: 0.74 MeV protons 33.2 keV/μm incident on C3H10T12 cells.



(d) Gueulette et al. (1996) [125]: 85 MeV protons in mid-SOBP incident on CHO cells.



(e) Baek et al. (2008) [124]: 190 MeV protons in mid-SOBP incident on HSG† cells.



(f) Belli et al. (1998) [127]: 0.64 MeV protons 36.4 keV/μm incident on V79 cells.

Figure 5.6: Our LQ model parameters (×) with standard deviation error ellipses against the parameters reported in the original studies (•). Note that cell lines originating from a human are denoted by a dagger (†) under each plot.

The linear behavior in the low-LET region is also consistent across other cell lines. This is better visualised in the fits from each study in the low-LET region presented in Figs.(H.1) and (H.2). In Fig.(H.1(a)) we present a LQ fit for the data of Wouters et al. [38] corresponding to V79 cells exposed to 160MeV SOBP with LET 1.1keV/μm to which the LQ model yields a very good fit. A strong linear relationship is present in this data because the residual plot shows very little scatter about zero. Figure (H.1(b)) which corresponds to our LQ fit to the data of Howard et al. [123] for the A549 human cell line exposed to a 160MeV pristine Bragg Peak with LET 2.26keV/μm also demonstrates strong linearity upon observation of the residual plot. Figures (H.2(a)) and (H.2(b)) depict fits to data of Howard et al. [123] and Belli et al. [60] for the CHO (71MeV, LET = 4.35keV/μm) and SCC25 (5.01MeV, 7.7keV/μm) cell lines. Although the LQ model fits the data quite well, the residual plots show that the errors are larger in the low dose region than high dose. This indicates that the CHO and SCC25 cell lines may be more sensitive to low-LET radiation than the V79 and A549 cell lines because there is a higher amount of scatter from the linearised LQ relationship. Overall, we can conclude that linearity is present in cells when exposed to low-LET radiation, although this appears to be somewhat dependent on the intrinsic biological conditions of the cell line considered.

As the LET increases, the increasing linear relationship of Eq.(5.12) appears to breakdown for most LQ fits as shown in our fits to the data of Belli et al. [26] in Fig.(5.4). It is in this region that $\beta \to 0$. Whilst Fig.(5.4(a)), still yields a reasonable fit to Eq.(5.12), we begin to observe more fluctuation about zero in the residual plots compared to Fig.(5.3). In Fig.(5.4(b)) linearity begins to dissipate as shown by the residual plot where there is a slight bias with most of the observations sitting above zero. Similar to Fig.(5.4), Fig.(H.1) depicts fits to the data of Bettega et al. [61] and Belli et al. [60] corresponding to our LQ fits to the C3H10T12 (3.18MeV, 11keV/μm) and HF19 (1.49MeV, 19.5keV/μm) cell lines. In both fits we observe $\beta \to 0$ resulting in Eq.(5.12) equating to α. Similar to Fig.(5.4(b)), there is much scatter in the residual plots in the low dose region for both cases. Although we note that both figures correspond to lower LET than that of Fig.(5.4(b)), it is possible that the C3H10T12 and HF19 cell lines deviate from LQ behavior at lower LET than the V79 cell line similar to the CHO and A549 cells discussed above.

In the high-LET region we show in Figs.(5.5(a)) and (5.5(b)) (Belli et al. [26] 34.6 and 37.8keV/μm incident on V79 cells, respectively) that the LQ relation becomes a convex decreasing function as β turns negative. The extremely low R^2 in each plot indicates that the model is a very poor fit for the data under high-LET exposure conditions. For example, Fig.(H.3(a)) (Bettega et al. [61], 11keV/μm incident on C3H10T12 cells) shows a very gradual positive slope where the low dose points are more scattered than the high doses as indicated in the residual plot. The reason for the bias toward high dose is the weighting of the fit to the data points with smaller errors whilst the data points at low doses have consistently larger errors. Therefore

the fit is more weighted to mid to high doses and not low doses which has the highest contribution to non-linearity and is reflected in the gof metrics. Figure (H.4(b)) yields a negative slope with $\beta = -0.032 \pm 0.003$ and another poor fit ($\chi^2 = 0.897$, $R^2 = 0.90$, RMSE = 0.947 and mean residual error = -0.005). The same behavior is present in Fig.(H.4(a)) which both demonstrate extremely poor fits, with non-linearity present in the low dose region due to the larger scatter about zero in the residual plots.

By studying the experimental data we observed that the LQ model does not explain the data in the high-LET region, while at low-LET there is more linearity. This is exemplified by the data in Belli et al. [26] and similar observations are made with the remaining experimental data considered in Appendix F and H. Overall we can conclude that the linear regression assumption of linearity is satisfied in the low-LET region (LET < 10keV/μm) and worsen as the LET increases. The impact on the LQ model's performance is that it does not give a detailed picture of the biological and chemical processes such as DNA lesion repair with time, but only explains the "hit" and "miss" process to the targets. Therefore, in the high-LET region where a more complex ionisation track structure is involved, the LQ model fails to adequately explain the data.

5.3.1.3 Testing for the Assumption of Normally Distributed Errors

The assumption of Normally distributed residual errors is tested using the Kolmogorov-Smirnov (KS) and Anderson-Darling (AD) tests for unknown mean and variance with critical values derived for *in-sample* testing [120]. The test results for the KS and AD tests are presented for all the experimental world data considered in the current study in Appendix G. The details of this "assumption" and the technical discussions are detailed in Appendix E. The hypothesis testing results show that many of the linear regression fits (in fact approximately half of the fits considered) fail one of both the KS and AD tests and therefore do not satisfy this key assumption (Table (G.1) of Appendix G). Some experimental data also presents a Type II error (i.e. non-rejection of a false null hypothesis). Namely the residual errors were not rejected by either test when they should have been. For example, some of the data contains outliers such as Fig.(5.5(b)). The failure rate appears to be high for depth studies despite many of the fits demonstrating linearity. In the energy varied case the failure rate is higher in the high-LET region, although not all of the tests fail.

If we carry on with our analysis of Belli et al. (1998) [26], we observe that in the linear case the highest LET (37.8keV/μm) fit in Fig.(5.5(b)) fails both the KS and AD tests. The remaining fits from Belli et al. [26] pass both tests despite some data demonstrating a poor linear fit (namely Figs.(5.5(a)) and (5.3)).

Upon observation of the fits in Table (F.1), the data from Wouters et al. [38] fails one or both of the KS and AD tests despite the good linear fit. We know that the KS test is most powerful with larger sample sizes which could be a possible explanation for the failure rate in the data from Wouters et al. [38] because the sample size is much larger compared to other data considered in this study. The KS test is also known to be most sensitive to deviations from the test distribution in the tail regions, whilst this is not the case for the AD test. There are a few cases in Table (F.1) where the data fails the KS test and passes the AD test (for instance the 230MeV, 1.03keV/μm and 4.02keV/μm fits to the data of Wouters et al. [38]). From this we may suggest that the residuals deviate from the Normal distribution away from the tail region. We plan to focus future efforts on why the fail rate of the data from Wouters et al. [38] is so high however this could not be investigated further in the current study due to time limitations.

To conclude, approximately half of the data from the literature fails either or both of the KS and AD tests for Normally distributed residual errors. This observation is consistent with what was observed in [17]. The consequence of this is that the errors of the fit are not trustworthy and seriously biased. This deficiency is most prominent in the high-LET region.

5.3.1.4 Testing for the Assumption of Homoscedasticity

The assumption of homoscedasticity (constant variance of residual errors) can be assessed visually and by making use of the Breusch-Pagan (BP) and White Tests described in Appendix E. For all the experimental data referred to in Appendix D, these tests are performed and the results are stated in Appendix G. The residual plots are presented in the panel underneath Figs.(5.3) and (5.4) and in Section (H.1) of Appendix H. **Note:** the Breusch-Pagan test assumes Normally distributed errors [115] and therefore the fits that fail the KS and AD tests cannot be considered for the BP test. The White test does not require this condition to be satisfied [116] and can therefore be considered for the fits that fail the KS and AD tests.

Figs.(H.1(a))–(H.2(a)) depict the variance of residual errors for the linear fit to low-LET data. In all instances we observe that the residual errors of the fits performed in this study scatter very little about the mean with deviations of no more than 0.05. For LET $<$ 8keV/μm, the fits that pass the KS and AD tests pass the BP and White tests at the 95-percentile. The only exception to this are the fits performed on the data from Wouters et al. [38] which consistently fail the White test. All but two of the fits from Wouters et al. [38] presented in Table (F.1) (namely 160MeV, 1.1keV/μm and 230MeV, 1.03keV/μm) do not possess Normally distributed residual errors hence rendering the BP test redundant. The White test however fails for this data as well. This behavior has not been observed for other experimental data, however time constraints have not allowed us to investigate this anomaly further but will be a focus of future work. The low-LET region demon-

strates the highest degree of linearity and hence it is not surprising that the fits demonstrate homoscedasticity under the BP and White tests.

In the mid to high-LET range ($> 8\text{keV}/\mu\text{m}$) the BP and White tests pass often. Interestingly, some of the high-LET fits fail the linearity condition in the residual plots below the fit (Figs.(5.5(a)) and (5.5(b)) are good examples of this). In fact the data in Fig.(5.5(a)) fails the BP test ($p = 0.041$) and passes the White test by a small margin ($p = 0.058$). Unfortunately we cannot conduct the BP test for the highest LET fit of Belli et al. [26] because the residual errors are not Normally distributed (it fails the KS and AD tests). Regardless, the White test passes implying that the residual errors are homoscedastic despite conflicting with the residual plots - this indicates the presence of a Type II error. It is also possible that the White test is not effective on small sample sizes because it is difficult to detect trends with so few data points.

There is no clear trend in behavior of the BP and White test with respect to cell line and depth. Both tests pass often when considered in terms of cell line and depth with the exception of Wouters et al. [38]. Overall, we observe conflicting results in light of the residual plots and hypothesis testing. The BP and White tests overwhelmingly imply that the residual errors are homoscedastic despite residual plots suggesting otherwise - particularly in the mid to high-LET range. We are limited due to the small sample size and non-Normally distributed residual errors of most data. Therefore, we recognise that this assumption should be investigated further with more data and rigorous testing against the conditions of the White test - particularly the data from Wouters et al. [38].

Considering the evidence provided by the residual plots and regression tests, the residuals are correlated with the absorbed dose when observed visually. This is despite many LQ fits passing the BP and White tests, thus suggesting the presence of a Type II error. The residual plots show that the dose-residual error correlation is more prominant in the high-LET region, whilst in the low-LET region most of the residual errors are randomly scattered about zero. The consequence of correlated residual errors is that the estimated LQ fit parameters and the corresponding errors are not trustworthy.

5.3.1.5 Testing the Assumption of No Autocorrelation

The assumption of the existence of no autocorrelation in the residual errors of the LQ model fits is best tested analytically. For larger sample sizes this effect can be observed graphically, however the sample sizes of the data considered in this study are too small to do this. Therefore we make use of the Durbin-Watson (DW) test and the results are stated in Appendix G. Ideally we would also like to implement the Ljung-Box test for autocorrelation [128] because it accounts for lag values > 1.

However this test could not be used because the sample sizes of the data considered here are too small, meaning that lags > 1 are not achievable.

Upon observation of Table (F.1), the DW test indicates no autocorrelation of the residuals for less than half of the data considered, many of which are in the low-LET range (with the exception of data from Folkard et al. [96]). Many of the higher range LET data ($> 10\text{keV}/\mu\text{m}$) report an inconclusive DW test result. Therefore we cannot say with certainty whether the residual errors are autocorrelated or not.

Interestingly, the data from Wouters et al. [38] indicates no autocorrelation in half of the fits (for the 160MeV beam with LET 1.1, 2.41 and 3.2 $\text{keV}/\mu\text{m}$; and 230MeV with LET 1.95, 2.95 and $4.02\text{keV}/\mu\text{m}$) considered whilst indicating positive correlation in three fits (160MeV: $2.06\text{keV}/\mu\text{m}$ and $4.74\text{keV}/\mu\text{m}$ and 230MeV: $4.02\text{keV}/\mu\text{m}$) and an inconclusive result in the remaining three (see Table (G.1) of Appendix G). Again, we cannot confirm what is causing this anomaly and future efforts will aim to uncover the reason for the poor performance of this data in light of hypothesis testing despite yielding an excellent fit to the linearised LQ model.

Overall, it is difficult to determine whether autocorrelation is present in the residual errors due to the large number of inconclusive results given by the DW test and the lack of alternative methods of testing for autocorrelation with small sample sizes. It is our expectation that data with larger sample sizes will clarify whether the LQ model fits satisfy the autocorrelation assumption. In such cases, the fit parameter estimates from the least squares fits are not trustworthy. Further testing of this assumption will therefore require data large sample sizes.

5.3.2 Non-Linear Regression Analysis

In this section we test the goodness-of-fit metrics when the experimental data in Appendix D are fit to the non-linearised version of the LQ model and the assumptions outlined in Section (5.2.1). We compare the behavior in goodness-of-fit metrics and hypothesis testing results with the linear case from the previous section as a function of LET, cell line and depth in the Bragg Peak.

5.3.2.1 GOF Metrics and Non-Linearity

The estimated LQ model parameters using the non-linear fit are similar in the low-LET region but differ in the mid to high-LET region, as shown in Table (5.1). An improved fit is observed in the high-LET region when using the non-linear LQ model fits compared to its linearised form. The reason being the lack of the linearity in the high-LET data as discussed in the previous section.

Let us continue the analysis of our benchmark study Belli et al. [26]. Similar to the linear case, Fig.(5.7(a)) yields an excellent fit ($\chi^2 = 0.42$, $R^2 = 0.99$ and RMSE = 0.02) to the data. If we compare this with Figs.(5.8(b)), (5.9(a)) and (5.9(b)) for LET $>$ 30keV/μm the fit appears to worsen until the highest LET is reached (the gof metrics of the highest LET are $\chi^2 = 1.82$, $R^2 = 0.96$ and RMSE = 0.05). Although the coefficient of determination is close to 1, the χ^2 and RMSE metrics show that the residual errors increase by a large amount in the high-LET region.

A comparison of the LQ model performance in the linear and non-linear cases are provided in Table (5.1). The behavior in the LQ model fit parameters in the low-LET region between the linear and non-linear cases do not fluctuate much. The parameters disagree more in the high-LET region due to the improved gof discussed above. In the case of Belli et al. [26] the linear and non-linear parameters agree within their error bounds. We note in both cases that the β parameter becomes negative for LET $>$ 30keV/μm. In the case of Wouters et al. [38] the fit parameters do not change and have negligible error. This is not surprising because the data demonstrated near perfect linearity when fit to the linearised LQ model and hence should also yield an excellent fit to the non-linear LQ model.

We also note that our fits to the data are consistently improved on those reported in the literature, as was the case using the linearised LQ model. The R^2 value is consistently better in comparison to the linear case upon observation of Table (F.4) (with the exception of Wouters et al. [38] where it remained between 0.99 and 1 for both cases). The most improvement is observed in the high-LET region where the R^2 increased to $>$0.95 for the non-linear cases presented in Table (5.1). Likewise, the χ^2 and RMSE metrics show substantial improvement on the fit. Recalling the non-linearity of the data points in the low dose region this was excepted.

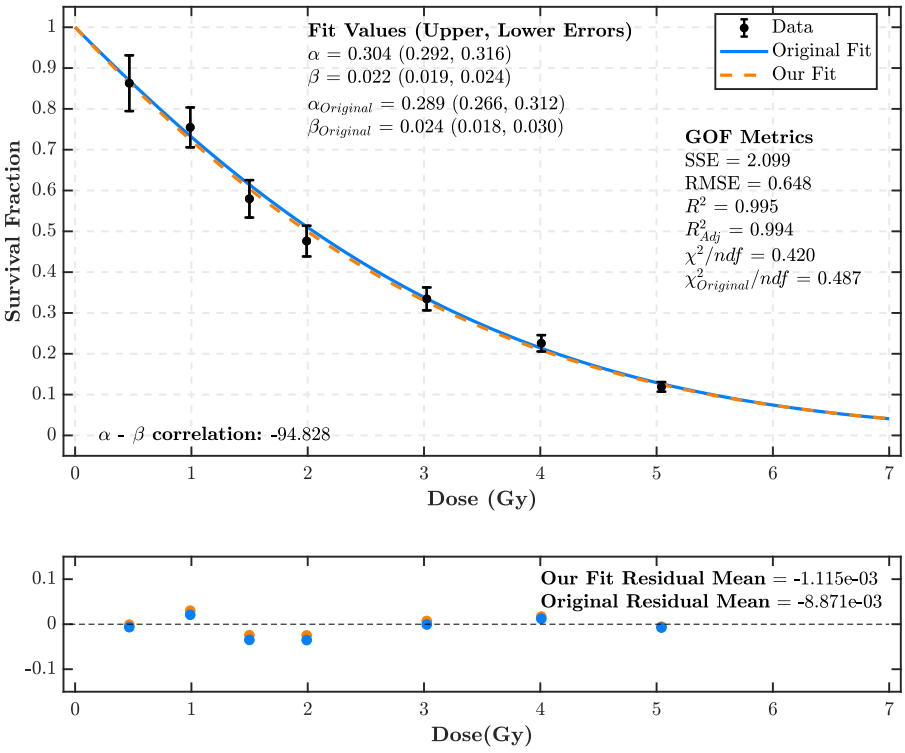
5.3.2.2 Testing for the Assumption of Normally Distributed Errors

The assumption of normally distributed residual errors is tested using the Kolmogorov-Smirnov and Anderson-Darling Tests for unknown mean and variance, with critical values derived for *in-sample* testing [120].

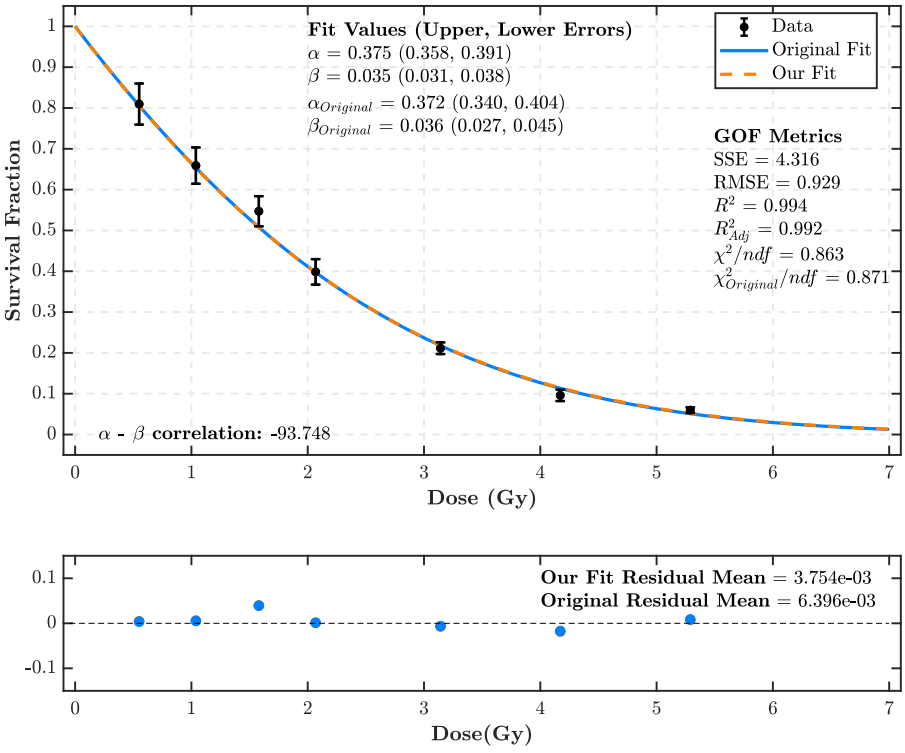
To continue with our analysis of the data from Belli et al. [26], both tests pass in the low-LET region and fail for the two highest LET (34.6 and 37.8 keV/μm). The remaining mid to low-LET fits pass both tests. The data from Wouters et al. [38] still fails both tests (with the exception of the AD test for 160MeV/2.41keV/μm). The results in Table (G.2) show that a just under half of the experimental data considered fails the KS and AD tests - only a slight improvement compared to the linear case. The power of the tests for data with a small sample size decreases and hence Type II errors are also observed.

Table 5.1: A summary of our fits to data from Belli et al. [26] and Wouters et al. [38] discussed in this section. The parameters and gof metrics for the non-linear fit are presented first with the corresponding linear fit are presented below in **bold**.

Energy	**LET**	$\alpha\pm$ **error**	$\beta\pm$ **error**	χ^2	$\mathbf{R^2}$	**RMSE**
Belli et al. (1998) [26] - V79 Cells						
5.01	7.7	0.304 ± 0.012	0.022 ± 0.003	0.419	0.995	0.648
		0.302 ± 0.012	**0.022 ± 0.003**	**0.459**	**0.716**	**0.677**
3.20	11.0	0.375 ± 0.017	0.035 ± 0.005	0.863	0.994	0.929
		0.376 ± 0.016	**0.033 ± 0.004**	**0.926**	**0.770**	**0.962**
1.41	20.0	0.502 ± 0.017	0.034 ± 0.004	0.366	0.997	0.605
		0.498 ± 0.017	**0.035 ± 0.005**	**0.750**	**0.818**	**0.631**
0.76	30.5	0.730 ± 0.027	-0.001 ± 0.007	1.719	0.990	1.311
		0.720 ± 0.027	**0.001 ± 0.008**	**1.797**	**0.001**	**1.341**
0.64	34.6	0.791 ± 0.034	-0.034 ± 0.009	1.765	0.984	1.329
		0.779 ± 0.035	**-0.032 ± 0.008**	**1.999**	**0.565**	**1.414**
0.57	37.8	0.690 ± 0.060	-0.023 ± 0.003	1.820	0.981	1.350
		0.701 ± 0.055	**-0.025 ± 0.011**	**2.101**	**0.420**	**1.449**
Wouters et al. (2015) [38] - V79 Cells						
160	1.10	0.122 ± 0.000	0.047 ± 0	2.1×10^{-4}	1.00	0.015
		0.122 ± 0.000	**0.047 ± 0.000**	**0.000**	**1.00**	**0.016**
160	2.06	0.112 ± 0.001	0.054 ± 0	1.1×10^{-3}	1.00	0.033
		0.112 ± 0.001	**0.054 ± 0.000**	$\mathbf{4\times10^{-4}}$	**1.00**	**0.035**
160	2.41	0.123 ± 0.000	0.053 ± 0.000	2.0×10^{-4}	1.00	0.014
		0.123 ± 0.000	**0.053 ± 0**	**0.00**	**1.00**	**0.016**
160	3.20	0.141 ± 0.000	0.054 ± 0.000	4.2×10^{-4}	1.00	0.020
		0.141 ± 0.000	**0.054 ± 0.000**	$\mathbf{4\times10^{-5}}$	**1.00**	**0.022**
160	4.74	0.154 ± 0.001	0.057 ± 0	1.9×10^{-3}	1.00	0.044
		0.154 ± 0.001	**0.057 ± 0.000**	**0.000**	**1.00**	**0.047**

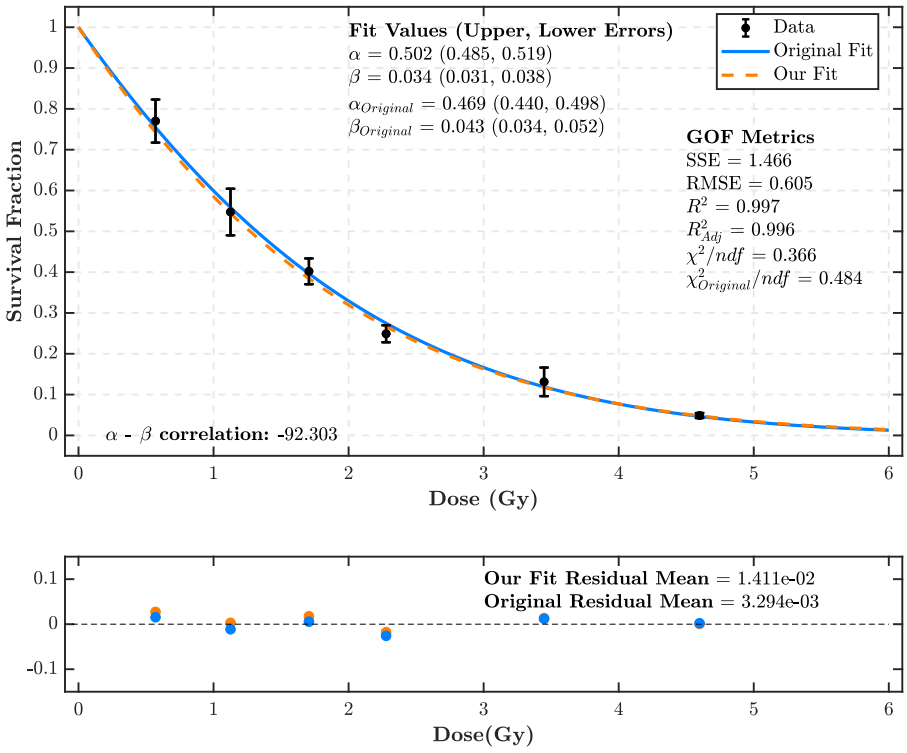


(a) 5.01MeV (7.7 keV/μm)

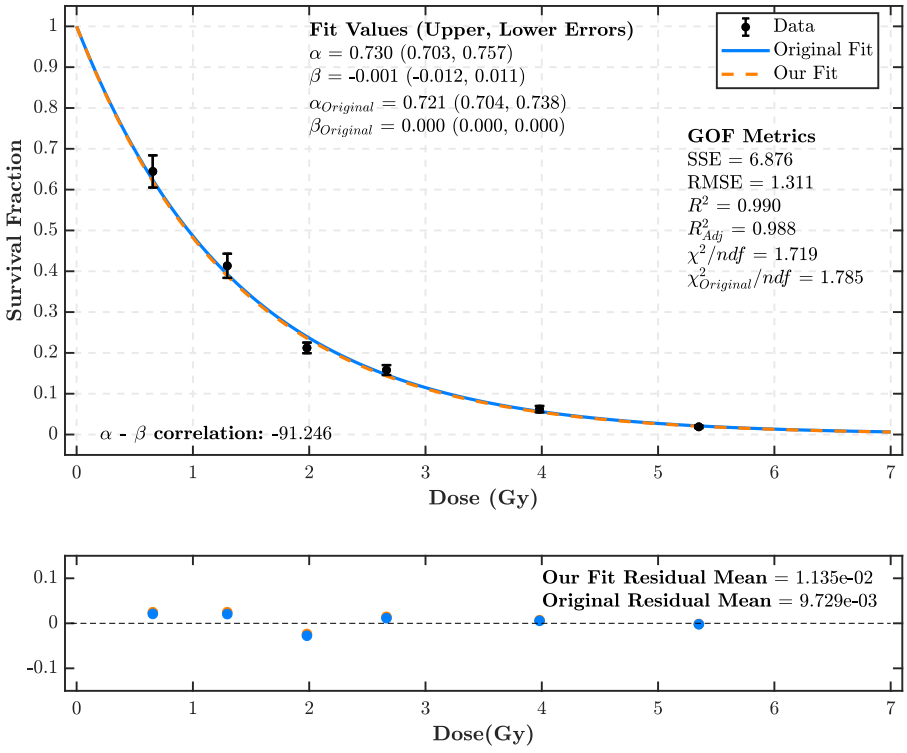


(b) 0.76MeV (30.5 keV/μm)

Figure 5.7: LQ fits on data from Belli et al. [26] with protons of varying energy incident on V79 cells. The second panel is a plot of the residuals between the observations and the linear LQ model.

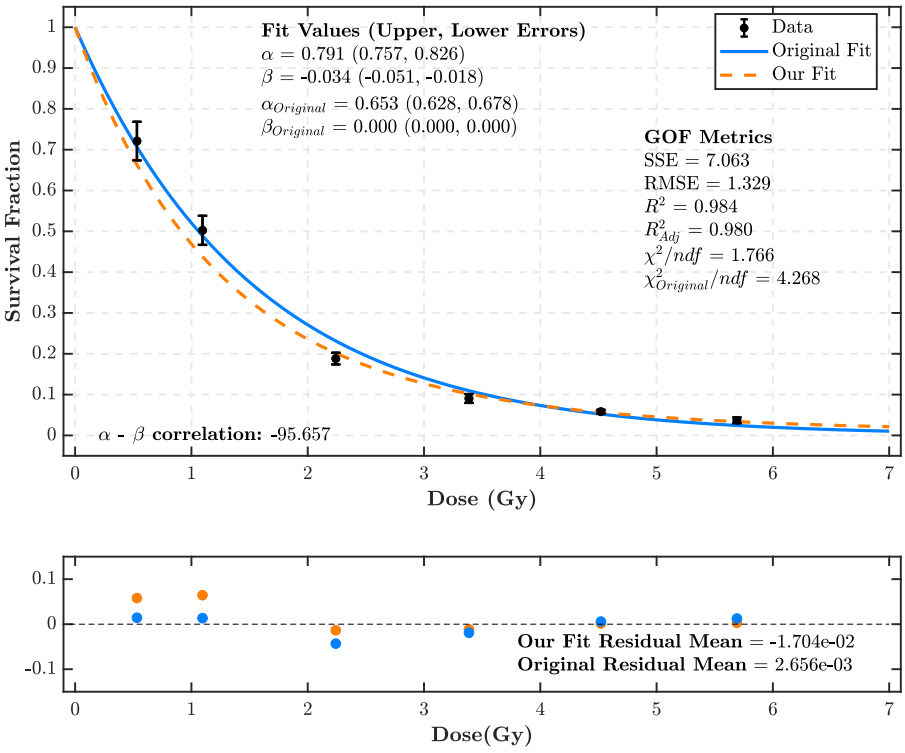


(a) 1.41MeV (20.0 keV/μm)

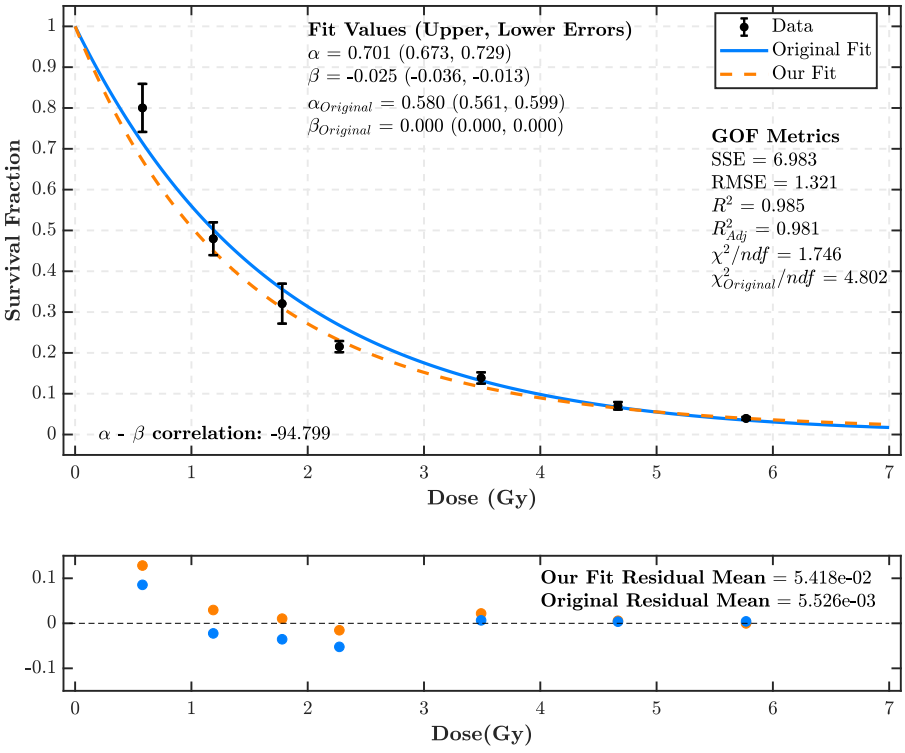


(b) 0.76MeV (30.5 keV/μm)

Figure 5.8: Linearised LQ fits on data from Belli et al. (1998) [26] with protons of varying energy incident on V79 cells. The second panel is a plot of the residuals between the observations and the linear LQ model.



(a) 0.64MeV (34.6 keV/μm)



(b) 0.57MeV (37.8 keV/μm)

Figure 5.9: LQ fits on data from Belli et al. [26] with protons of varying energy incident on V79 cells. The second panel is a plot of the residuals between the observations and the linear LQ model.

Overall, we can conclude that the residual errors for the non-linear LQ fits are not Normally distributed, which is more prominent in the high-LET region. This is not much of an improvement over the linear case where more than half of the data sets studied failed both tests. This shows that the non-linear version of the LQ model is also inadequate at describing the data under high-LET exposure conditions, based on regression analysis.

5.3.2.3 Testing the Assumption of Homoscedasticity

The assumption of homoscedasticity (constant variance of residual errors) can be assessed visually and through the use of the Breusch-Pagan (BP) and White Tests introduced in Section (E.1.1) of Appendix E. Although the BP test is designed for linear regression, we wish to observe the power of the test for detecting homoscedasticity in the non-linear LQ case, because the LQ model is inherently linear via the transformation Eq.(5.12). Of course, we recall that the BP test cannot be applied to data with non-Normally distributed errors (i.e. the tests that fail the KS and AD tests). The relevant plots are presented in the panel below Figs.(5.7) and (5.8) and in Section (H.2).

In the low-LET region we observe correlations in the residual plots in all figures except Fig.(H.5(a)), where the residuals appear to remain scattered about the mean. The residual plots of the remaining fits show the best agreement in the low survival region (high dose) with an increase in residual error for high survival fraction (low dose). A similar observation was made with the linear case. An explanation for this is that we performed a weighted fit, meaning that we give more weight to the data points with small errors. Figures (5.7) and (5.8) and in Section (H.2) of Appendix H we show that the errors are consistently small in the high dose region, meaning that our fit is more heavily weighted to this region.

The results for the BP Test at the 95% significance level paint a similar to picture to our linear LQ model fits, in that a large fraction of fits indicate homoscedasticity when small sample sizes are considered. This is despite many of the residual plots discussed above showing some degree of correlation with absorbed dose. Again, the BP test will be more powerful for large sample sizes that could not be considered in this study. The data from Belli et al. [26] passes both the BP and White tests in the low-LET region.The White test passes for all fits in the high-LET region ($>$30keV/μm), however the BP test could not be performed on this data because their residual errors are not Normally distributed. However, the residual plots show that the residuals are larger in the low dose region and gradually approach zero as dose increases. This implies that the fits are bias, which is likely due to the weighting of the least squares fits and the small errors on the high dose measurements. At an LET of 30.5keV/μm, the fit fails the BP test, implying that the residuals are correlated with dose. Wouters et al. [38] could not be considered in light of the BP test because the residual errors are not Normally distributed. Similar to the linear

case, many data sets from [38] fail the White test. Due to time constraints we could not examine this anomaly further however we wish to explore this in future work.

Overall the BP test passes in most cases with no trend observed in the pass rate. The White test again passes for nearly all the fits with the exception of Wouters et al. [38]. Similar behavior is observed in the linear case despite many of the residual plots showing a degree of correlation. It is likely the result of Type II errors due to the tests not detecting correlations in data with small sample size despite the residual plots implying otherwise. The result of correlated residual errors with dose implies that the LQ fit and its estimated parameters are not trustworthy. This effect is observed more in the mid to high-LET region.

5.3.2.4 Testing the Assumption of No Autocorrelation

Unfortunately we could not test for autocorrelation in the non-linear case because the Durbin-Watson (DW) test can only be applied to linear regression model [117]. However a model for approximating the exact distribution of the DW test statistics for first-order autocorrelation (lag value of 1) has been developed for non-linear regression [129]. Due to time limitations we have decided to leave this for future studies.

Overall, the non-linear LQ fits (the LQ model form used in the literature) are improved compared to the linear case in the mid to high-LET region. They are comparable in the low-LET region where the linearity condition of the previous section is satisfied. Despite the improved fit, the assumptions of regression analysis are still not satisfied, particularly in the mid to high-LET range. The pass rate for each test did not improve substantially between the linear and non-linear LQ fits. Therefore, we have shown that the assumptions of Normally distributed and correlated errors for the non-linear case of the LQ model are not satisfied in the high-LET region, despite the improved goodness-of-fit.

5.4 The Influence of Fitting on Relative Biological Effectiveness (RBE)

A clinical proton RBE of 1.1 has been adopted in hadron therapy centres despite many studies reporting conflicting evidence [25, 26, 123]. The literature has shown the RBE to vary with LET, depth and cell line/biological conditions [27, 60, 130]. The LET range considered in the current study is 0.99 to 37.8keV/μm, however heavy ions are capable of achieving a higher LET range and RBE. Figure (5.10) depicts the RBE at 10% survival as a function of LET using published fits from Belli et al. [26] and Furusawa et al. [131]. Here the RBE peaks for proton radiation at LET 30keV/μm whilst heavy ions peak at a much higher LET. Furthermore, the peak proton RBE for LET $>$ 30keV/μm is higher than that of heavy ions. When using high RBE radiation such as protons and heavy ions for cancer treatment, it is

critical to fit the data accurately to derive the correct RBE.

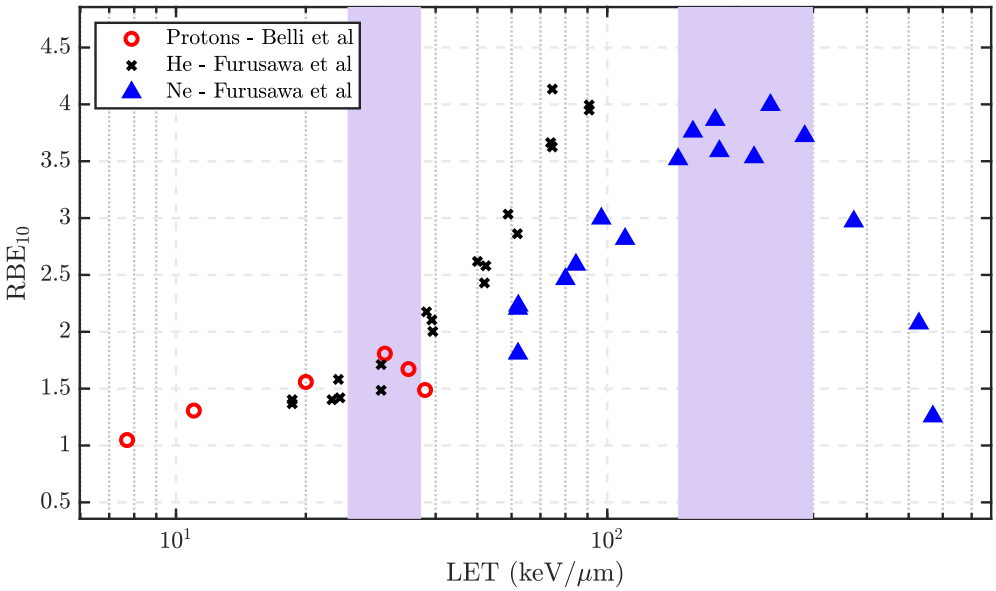


Figure 5.10: RBE as a function of particle LET at 10% survival for proton beams [26], a helium ion beam and a neon beam [131] all incident on V79 cells. The shaded regions indicate the LET at which protons and heavy ions have the highest RBE.

The current study has uncovered several deficiencies in the LQ model fits to data from the literature (Appendix D). A key observation is that the fits are not determined rigorously and the gof metrics of our fit show that those reported in the literature can be improved. We have observed that the LQ model fits to the data of Belli et al. [26] worsen with increasing LET in both the linear and non-linear forms of the LQ model and this effect is also present for other data from the literature (see Table (F.4)). Additionally we observed differing responses across cell lines under similar exposure conditions such as the data from Howard et al. [123]. The depth study also yields worse fits at larger depths in the Bragg Peak which is inherently related to LET. By comparing the RBE computed using our fits to data from Belli et al. [26], Bettega et al. [61], Howard et al. [123] and Wouters et al. [38] to those reported in the literature, we have demonstrated the effect that a poor LQ model fit has on the RBE.

The RBE can be calculated using multiple endpoints as we have discussed in Chapter 2. In the current study we calculate the RBE by taking the ratio of doses required for the test (proton/heavy ion) and reference (photon) radiation to achieve a survival fraction of 10% and 37% respectively (as per Fig.(5.11)). We can also consider the linear component of dose response only (i.e. $SF = \exp(-\alpha D)$) by taking the ratio of fit parameters α_x and α for the reference and test radiation. This expression of RBE is derived using

$$\begin{aligned} SF =& e^{-\hat{\alpha}D} = e^{-\hat{\alpha_x}D_x} \\ \Rightarrow \log(SF) =& -\hat{\alpha}D = -\hat{\alpha_x}D_x \\ \Rightarrow \frac{D_x}{D} =& \frac{\hat{\alpha}}{\hat{\alpha_x}} \quad . \end{aligned} \tag{5.15}$$

where α and α_x are the fit parameters and D and D_x are the absorbed doses of the test and reference radiation respectively. The error in RBE($\hat{\alpha}/\hat{\alpha_x}$) is calculated using

$$\frac{\hat{\alpha} \pm \sigma_p}{\hat{\alpha_x} \pm \sigma_x} = \frac{1}{\hat{\alpha_x^2}}\left[\sigma^2 + \sigma_x^2 \frac{\hat{\alpha^2}}{\hat{\alpha_x^2}}\right] \tag{5.16}$$

where σ and σ_x are the errors associated with $\hat{\alpha}$ and $\hat{\alpha_x}$, respectively.

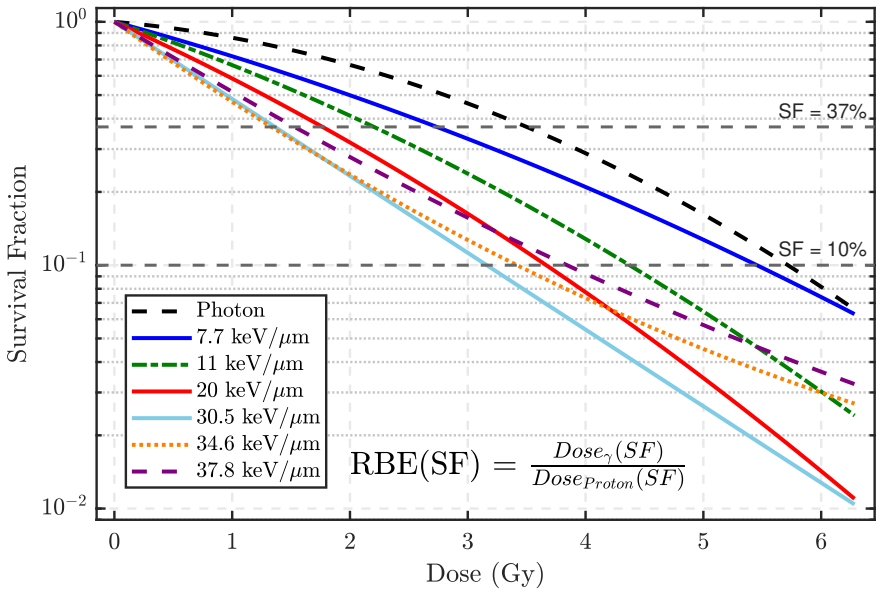


Figure 5.11: An example of how RBE is computed at 10% and 37% survival using the data from Belli et al. [26].

5.4.1 RBE as a function of LET

In this section we consider the RBE computed at 10% and 37% survival as a function of LET. We also consider the response of different cell lines and their resultant RBE. We follow this by calculating the RBE using the linear components of the LQ model, α/α_x, where α is the fit parameter to the proton survival curve and α_x is that of the reference beam.

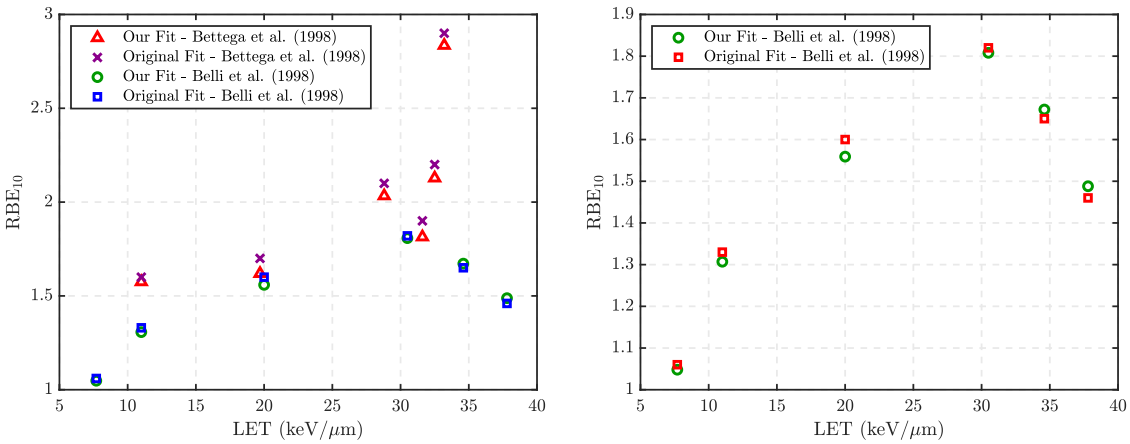
Figure (5.12(a)) depicts the RBE calculated at 10% survival using the non-linear LQ fits from the previous section to data from [26, 61] as a function of LET. The RBE values calculated in the literature are also presented for comparison. Considering our benchmark study, Belli et al. [26], the RBE increases with LET until 30keV/μm where the effectiveness peaks and begins to decrease. This effect is also observable in Fig.(5.11) where the cell killing efficiency decreases for LET 34.6 and 37.8keV/μm in the high dose region. The RBE is known to peak at a given LET depending on the radiation type [132].

The ability of a heavy ion to propagate further into tissue than protons suggests that the peak RBE will occur at a higher LET than that of a proton beam (see Fig.(5.10)). This effect is not visible in the data from Bettega et al. [61] however this may be the result of differing radiation sensitivity between the V79 and C3H10T12 cell lines. In fact, fast proliferating cell lines (i.e. with a greater repair capacity) require higher-LET radiation to induce cell death than slowly-proliferating cells, which have a much lower repair capacity [132]. This implies that the LET at which the maximum or "effective" RBE occurs is also dependent on the cell and its intrinsic biological environment parameters. It is therefore likely that C3H10T12 cells are fast-proliferating in comparison with V79 cells which explains why we do not observe the maximum RBE which will occur at a higher LET.

There are also slight differences in the resultant calculated RBE using our fits to the data compared with those reported in the literature [26, 61]. They are not significant but nonetheless demonstrate that the RBE is affected by a poor LQ model fit to the data.

Figure (5.13) depicts RBE(D_x/D) as a function of proton LET on different cell lines in the low LET range evaluated at 10% and 37% survival. First we observe that the RBE values are slightly lower at 10% survival in comparison with 37% survival due to the loss of cell killing efficiency in the high dose region. Noting that high survival occurs at low doses, later in this section it is shown that low dose RBE(D_x/D) is larger than the high dose range. Therefore, a higher RBE should be expected at 37% survival with respect to 10%. The variance in RBE appears to be slightly higher at the 37% survival rate, particularly at an LET of 7.29keV/μm. The RBE calculated from our fit in comparison to the fit reported in Howard et al. [123] also differs particularly at 37% survival. This demonstrates the importance of rigorous survival curve fitting because low doses such as these are used commonly in hadron therapy and this is the region where the RBE has a high variance.

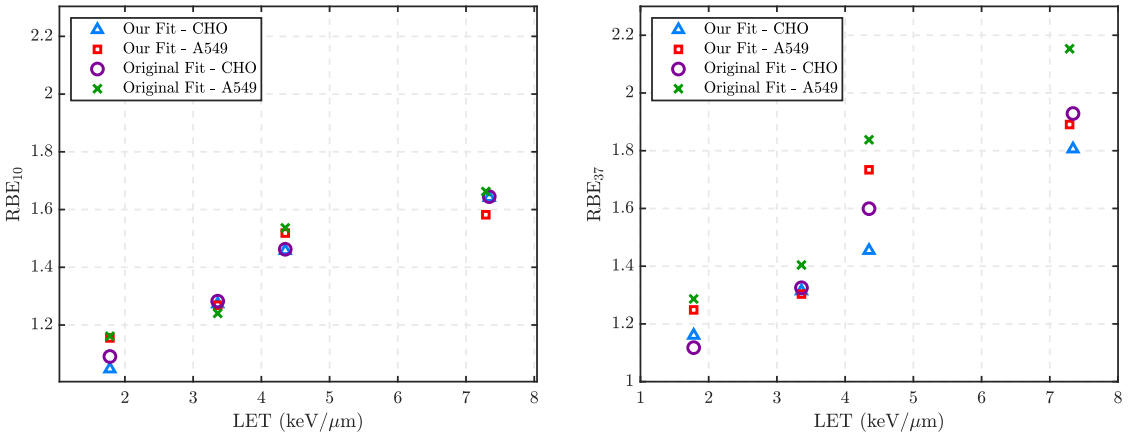
The above discussion of fast and slow-proliferating cells is difficult to implement here because the maximum RBE likely occurs at an LET much higher than those considered in this study. In this LET range the CHO and A549 cells appear to have a similar response to high doses (i.e. low survival) than in the low dose region. How-



(a) Bettega et al. [61] and Belli et al. [26] **(b)** Belli et al. [26] only

Figure 5.12: A comparison of RBE as a function of particle LET at 10% survival for proton radiation of varying LET incident on V79 cells from Belli et al. [26] and C3H10T12 cells from Bettega et al. [61] using the fits reported in the literature and in this study.

ever even in the low dose region (i.e. high survival) the RBE of each cell line differs by < 0.2. It is possible that both cell lines have similar repair capacities in the low LET region and hence have a similar response. This difference in fit shows a very different RBE prediction in the high survival fraction (low dose) region compared to the low survival fraction (high dose) region. This is because the our fits agree with the literature for high doses more than low doses (see Fig.(H.5(b)) as an example).



(a) 10% survival **(b)** 37% survival

Figure 5.13: RBE(D_x/D) as a function of particle LET for the CHO and A549 cell lines in response to a 71MeV pristine proton beam with varied depth at low and high survival rates [123].

Figure (5.14) depicts RBE(α/α_x) as a function of LET for the experimental results from [26, 133]. Similar to Fig.(5.12(a)), the RBE increases with LET to a maximum before decreasing again - only here the maximum occurs at an LET of 34.6keV/μm compared to 30.5keV/μm. The carbon ion RBE from [133] is lower than the proton RBE which is also consistent with that we observe in Fig.(5.10). This is expected because protons have a higher interaction cross section in this LET range than heavy ions because the latter typically travel faster at an LET of 30.0keV/μm. The RBE behavior of heavy ions will be a focus of future work.

Using the original fit parameters reported in the literature for this same calculation results in a maximum RBE at LET 30.5keV/μm for V79 cells in response to proton radiation. There is also a maximum difference in RBE of ≈ 3 between the RBE calculated using our fit and from the literature. The result is a significant under estimation of the RBE for all LET in Fig.(5.14) using the fits presented in the literature and the effect is most prominent for LET $>$ 30keV/μm. The heavy ion RBE was calculated using our non-linear LQ model fits however no comparison with the literature could be made because the LQ fit parameters are not reported [133]. We observe that in the LET range considered for this study, the carbon RBE is lower than that of proton radiation.

Another noteworthy observation to make from Fig.(5.14) is the increase of error in the high-LET range. At a low-LET of 7.7keV/μm the errors are relatively small indicating a good estimate with respect to the LQ model fit. An increase in error is observed as the LET increases, indicating a poor RBE estimate in the high-LET range for each fit using the LQ model. Furthermore, the error bars between our RBE predictions are those reported in the literature do not overlap. This shows the full effect of a non-rigorously determined fit on RBE calculations.

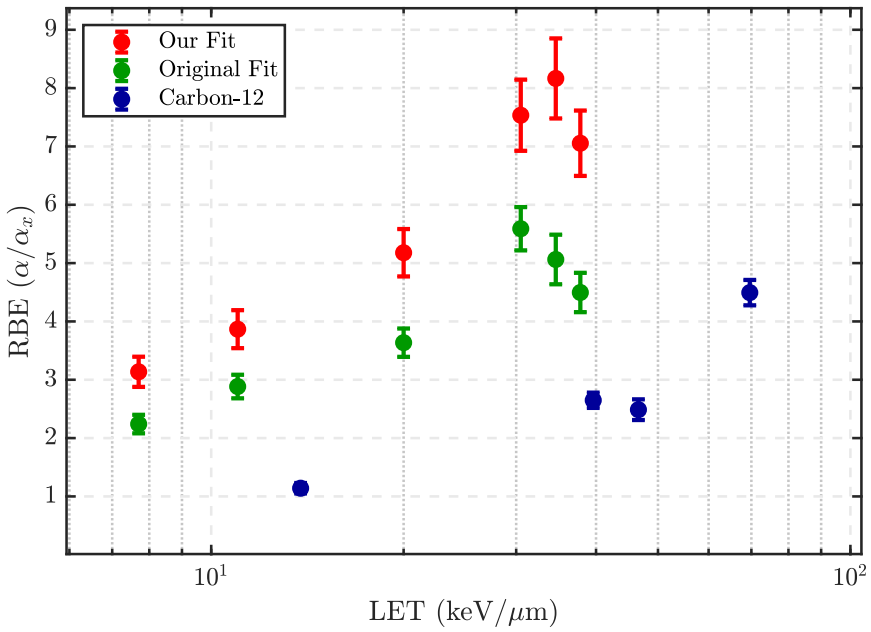


Figure 5.14: RBE(α/α_x) as a function of particle LET for proton beams of different energy incident on V79 cells [26], a carbon beam on HSG cells [133], a helium ion beam incident on V79 cells and a neon beam incident on V79 cells [131]. The errors are calculated using Eq.(5.16). Note: errors are not present on the results from [131].

To conclude, the reported fit parameters from the literature offer a substantial difference in RBE compared to the fits performed in our study. This effect is particularly prominent in the high-LET range (Fig.(5.14)) where the published fits from the literature underestimate RBE by at least 1. In the low-LET region the RBE calculated from our fits in comparison with those reported in the literature are more clustered together than in Fig.(5.14). Recalling that the fit from our study is improved in comparison with [123], the RBE values calculated from this study are more trustworthy. The error bars on Fig.(5.14) also indicate that the RBE(α/α_x) estimation is less reliable in the high-LET region.

5.4.2 RBE as a function of Depth

The RBE(D_x/D) can also be reported as a function of depth in the Bragg Peak. Figures (5.15) and (5.16) depict the RBE(D/D_x) as a function of depth in a SOBP and pristine Bragg Peak respectively. As expected from the literature [25], the RBE should increase with depth in the Bragg Peak. Again, this is exactly what in observed in the calculated RBE values for both our fits and those reported in the literature. The exact LET values were not reported in [130] however they were in [123]. The corresponding LET for each point in Fig.(5.16) ranges between 1 and 7keV/μm and we can assume a similar range for those from Fig.(5.15). In fact many depth studies from the literature result in lower LET values.

In comparison with Fig.(5.13) in the same LET range, we observe more agreement between the RBE values than in Fig.(5.14). Fig.(5.15) shows that RBE calculated from the published fits and our fits differ by approximately 0.1. This shows that small changes in the model fit, even in the low-LET region, can result in a very different RBE prediction.

Similarly in Fig.(5.16) the differences between RBE calculated using our fits and the published fits are quite small but nonetheless differ at some depths. Again, the data from [123] is in the low LET range where more agreement between the RBE values from our study and the literature is observed.

Overall, the quality of fit to the survival data influences the RBE as demonstrated by comparing those calculated using our fits and the published fits. They also demonstrate the expected trend of increasing with depth and decreasing when calculated at higher survival rates/lower doses. Despite the differences between RBE using our fits and the published fits being small, they are still nonetheless different. Since our fit performs better with respect to both experimental studies from [123,130], we can conclude that our RBE calculation is more reliable than that in the literature.

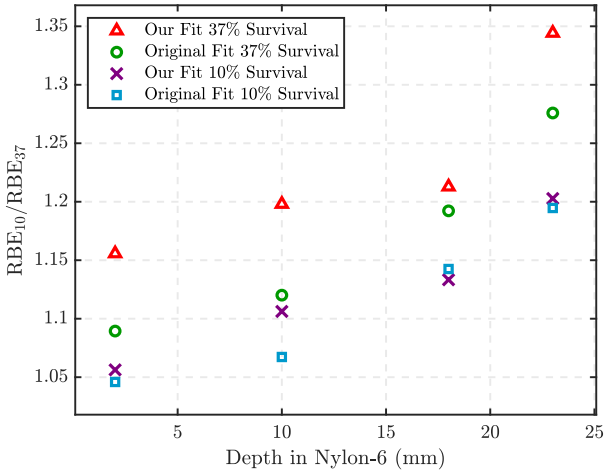


Figure 5.15: RBE(D_x/D) measured at 10% and 37% as a function of depth in nylon-6 for the CHO cells in response to a 65MeV, 10mm SOBP from Tang et al. [130].

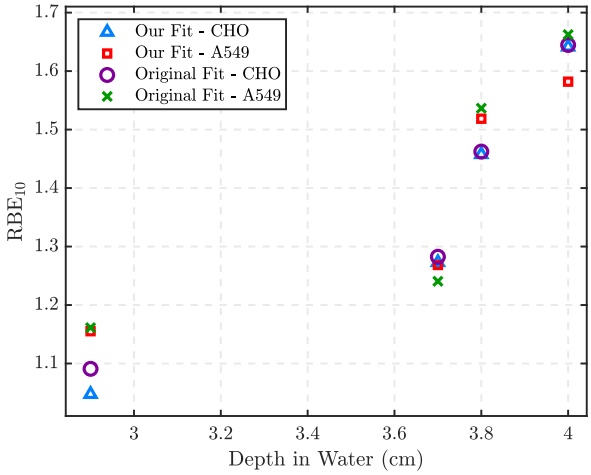


Figure 5.16: RBE(D_x/D) measured at 10% as a function of depth in a water for the CHO and A549 cells in response to a 71MeV pristine Bragg Peak [123].

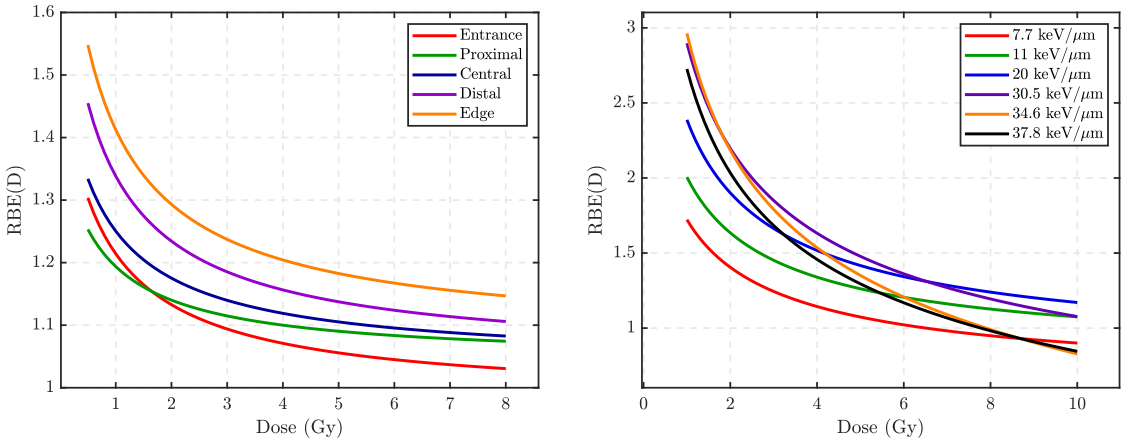
5.4.3 RBE as a function of Dose

When the LQ fit parameters α, α_x, β and β_x are known, one can determine the RBE as a function of absorbed proton dose D. It is derived by solving for the ratio D_x/D at which $S(D_x) = S(D) = e^{-\alpha D - \beta D^2}$. The expression is as follows

$$RBE(D) = \frac{D_x}{D} = -\frac{1}{2D}\frac{\alpha_x}{\beta_x} + \frac{1}{2D}\sqrt{\left(\frac{\alpha_x}{\beta_x}\right)^2 + 4\big(\alpha + \beta D\big)\frac{D}{\beta_x}} \tag{5.17}$$

where D_x is the reference beam dose, α/β are the fit parameters for the test radiation and α_x/β_x are the fit parameters for the reference beam. The resultant relationship is a decreasing RBE for increasing dose. Figures (5.17(a)) and (5.17(b)) depict the relationship in Eq.(5.17) using the fit parameters to the proton and reference dose response curves from our study. This same figure is produced in [38] as a function of reference beam dose. The same trend of RBE increasing with decreasing dose with a non-linear relationship is observed in other parts of the literature [25, 113].

It is worth noting that the experimental data used to produce Fig.(5.17(a)) considers proton LET values between 1 and 5keV/μm. The LET range is much larger in the data used to generate Fig.(5.17(b)). For this reason the RBE is consistently higher in Fig.(5.17(b)) compared to Fig.(5.17(a)). This is analogous with the behavior observed in Section (5.4.1) where the RBE grows with LET before peaking at an LET $\approx$ 30keV/μm. Recall from Fig.(5.14) that for LET $>$ 30keV/μm an "overkill" effect is observed where RBE begins to drop. This is also evident in Fig.(5.17(b)) where all LET $>$30keV/μm decrease at a higher rate with dose than lower LET.



(a) LET varied by depth in a 160MeV 10cm SOBP Wouters et al. [38].

(b) LET varied by energy Belli et al. [26]

Figure 5.17: RBE as a function of absorbed dose for V79 cells irradiated with proton beams [26, 38] using Eq.(5.17).

5.5 Conclusions

The aim of this investigation was to assess the limits of the LQ model in its linear and non-linear forms on the experimental data given in Appendix D for different biological and exposure conditions. The following conclusions are drawn from our findings.

- Many of the LQ model fits performed in the literature are not rigorously determined. By performing a least squares fit weighted by the SEM, we improved upon all the published fits.
- Regression analysis of experimental hadron therapy data shows that the LQ model does not adequately describe the data in the mid to high-LET range. Meanwhile, the LQ model performs much better in the low-LET range.
- If the survival curve is not rigorously determined, endpoints such as RBE are not computed accurately. Even small deviations in the fit can impact the RBE prediction significantly. Therefore, an accurate dose-response model is key to calculating reliable RBE values.
- There are more complex physical, chemical and biological processes involved with high-LET radiation such as protons and heavy ions. The model used to compute the RBE in this LET region should account for such processes.

6

Beyond the Linear-Quadratic Model

The Linear-Quadratic (LQ) model describes a "hit and kill" target model based on the dual radiation action theory [23] and is modelled using the Poisson process. In the previous chapter we questioned the validity of the LQ model by fitting the available experimental data through extensive statistical analysis. We concluded that the published LQ model fits contain deficiencies when goodness-of-fit statistics, error (residual) analysis and the assumptions of regression are considered. Our analysis supports the conclusions derived from the regression study performed by the JLab group [17, 18, 39].

Upon comparison observations, the LQ model appears to fit a number of dose-response curves with reasonable accuracy, however evidence from Chapters 2 and 5 have shown that it does not extend to cover full generality, for instance, in the high-LET region. The importance of stringent statistical testing of such models can be critical to the success or failure of their ability to describe a system or process.

The conclusions drawn in Chapters 2 and 5 has shown a reduction in the LQ model's effectiveness for high-LET radiation and high dose regions in particular. Furthermore, **the LQ model can fit to a dose-response curve under specific combinations of parameters. However, predictions with other parameters cannot be derived based on this fitting**. Despite this, use of the LQ model is substantially higher in clinical and research applications than alternative, more detailed models such as the TLK [33], RMR [31] and LPL [32] models discussed in Chapter 2.

In this chapter, a new model is introduced with the aim of re-evaluating the underlying statistical assumptions of the classic LQ model (see Chapter 5). It is our hope that this proposed model explains the discrepancies of goodness-of-fit with high-LET dose-response data. The testing of this model was undertaken in a series of

stages, first by testing the proposed model : (i) against experimental (details of the data are given in Appendix D) (ii) simulated data at different biological endpoints and (iii) comparing with the LQ model. The test results are presented in Chapter 8.

6.1 The Poisson Process within our Physical Picture

Modelling cell death with the highest achievable accuracy through radiation induced biological damage is the ultimate goal in the field of radiation physics. There are many attempts in the literature to achieve this goal, some of which are given in [31–33]. We will first discuss the physical problem and how it is currently handled in the literature. Afterwards we will put forward our proposal model to approach solving this problem.

Let us visualise the problem in four distinctive stages from the initial state of a system, $|i\rangle$ to the final state of a system $|f\rangle$ with a very simplified sketch :

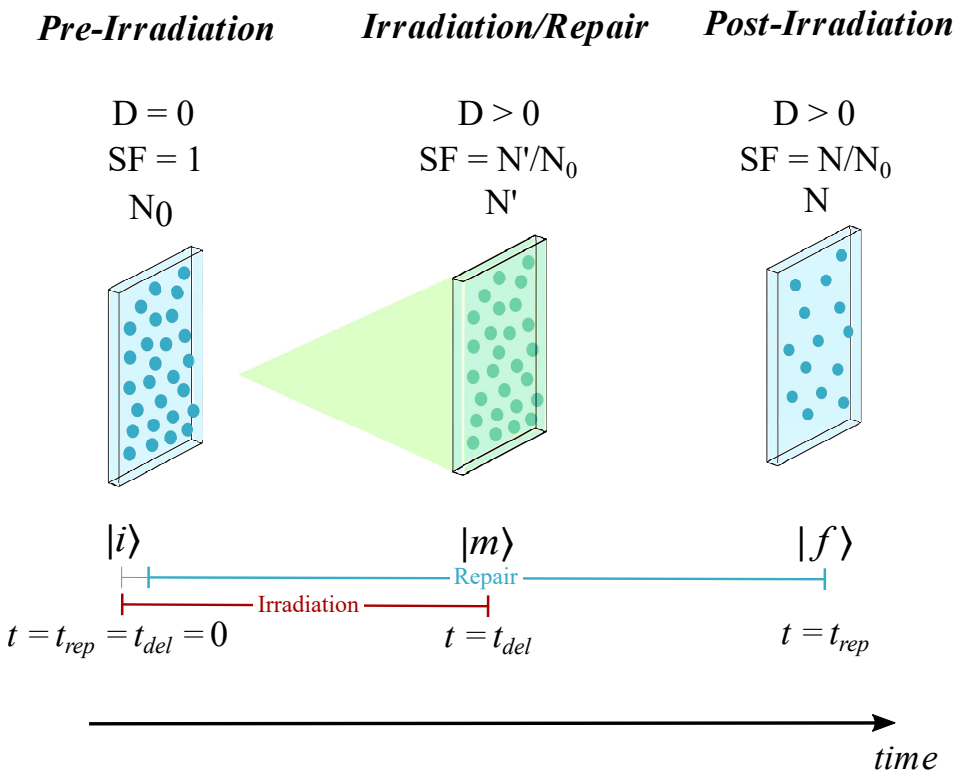


Figure 6.1: In the physical picture, the timescale of beam delivery and damage repair with cell survival are shown by t_{rep} and t_{del} representing the damage repair and beam delivery time, respectively.

1- Pre-Irradiation Stage Prior to irradiation the initial state consists of a total number of cells (target) are given by N_0 at $t_{del} = 0, t_{rep} = 0$ and $D = 0$. At this time the survival fraction is $SF = 1$, i.e. all the cells are alive.

2- Irradiation and Repair Stage Irradiation of the target (N_0 cells) commences and continues until the desired dose is delivered. The absorption of dose has been achieved by ionising particles passing through the sensitive regions (nucleus) of the target (cells), causing ionisations and excitations as they propagate through. This results in DNA damage (lesions) being induced within the sensitive regions of the cells, i.e. the regions that contain DNA. The number of ionisations (hits) increases with the dose absorbed by the target, so more hits can be found in a single cell when high doses are delivered. As discussed earlier in Section (2.1), cell death is related to the yield of lethal lesions induced on the DNA helix. As shown in Fig.(2.1), the repair process starts at $t = 10^{-6}$seconds after the damage is produced, hence the repair process is already present during the irradiation process. At a constant dose rate, the dose delivery is a function of time; while low doses are delivered in a short time, high doses typically take longer.

3- Post Irradiation - only Repair Stage When the desired dose is reached the irradiation stage ceases whilst the repair process continues until t_{rep}, which is when the survival fraction reaches equilibrium (i.e. saturation is reached). Towards the end of this repair time a large proportion of lesions which are of the form SSB, DSB and cDSB will have been repaired through biological and chemical processes inside the cell. However, if a lesion cannot be repaired it will become lethal and contribute to cell death.

4- Final Stage In this stage where $t > t_{del} + t_{rep}$, after saturation time $t > t_{rep}$, the number of cells that survive, N, are counted in the experiment/simulation.

As mentioned above, the process of radiation induced cell death is quite complicated. The damage caused by the radiation is counter-balanced by the chemical and biological repair process which affects the final cell survival fraction.

The theoretical models proposed in the literature that govern the mechanism for radiation induced cell death are probability models. More specifically they are a Poisson counting process. As an analogy, consider an experiment in which every outcome is described as either a "hit" or "failure", this is depicted in Fig.(6.2). The probability of a "hit" is denoted by p $(0 < p < 1)$ and the probability of a failure is denoted by $q = 1 - p$. Suppose the experiment is repeated n times producing a sequence of "hits", H and "failures", F. For example the outcome of four experiments might be $HFHH$. The order may represent discrete time, in which case the sequence can be interpreted as a discrete time stochastic process. The outcomes of each experiment at different points in time are independent and the probability p

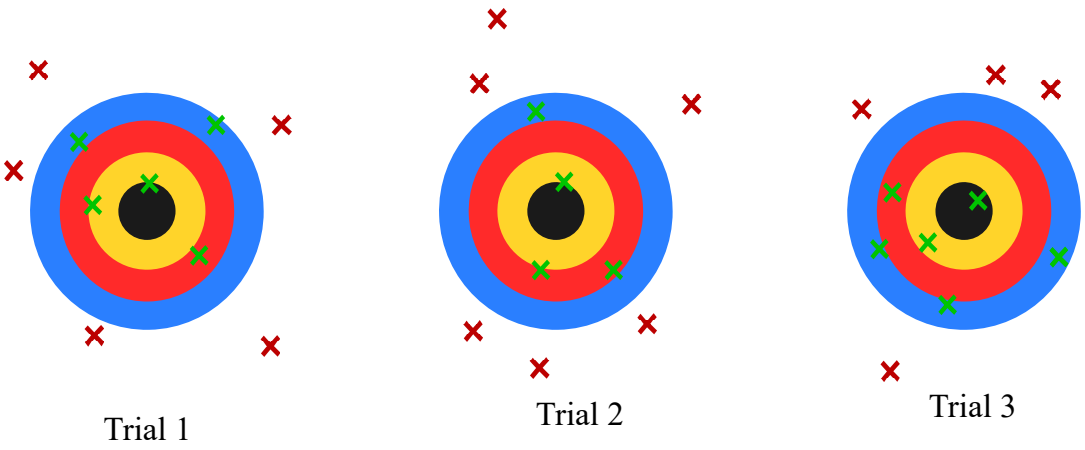


Figure 6.2: A visualisation of N arrows being shot at a target with k successes (green) and $N-k$ failures (red) in three independent experiments.

of a "hit" is constant over time.

Let X_i denote the outcome of the ith experiment, with $X_i = 0$ denoting a "failure" and $X_i = 1$ a "hit". The sum of n repetitions of such trials describes the Bernoulli experiment :

$$S_n^{(p)} = X_1^{(p)} + X_2^{(p)} + ... + X_n^{(p)} = \sum_{i=1}^{n} X_i^{(p)} \quad . \tag{6.1}$$

$S_n^{(p)}$ is an integer random variable taking on values between 0 and n. If k $(0 \leq k \leq N)$ is the sum of X_i after n experiments then it can be shown that the probability of k has a binomial distribution :

$$P(k) = \text{prob}(X_1^{(p)} + X_2^{(p)} + ... + X_n^{(p)} = k) \; = \; \binom{n}{k} p^k q^{n-k} \quad . \tag{6.2}$$

Let us now consider a model in which the random arrivals of events ("hits") occur at a fixed rate λ per unit time. At $t = 0$ there are zero successes (no arrival of events yet) so $N(0) = 0$. We divide the total time t into n intervals of equal length $\delta = t/n$. In this case p is the probability of "hits" is proportional to the length of the interval, δ. As the number of intervals increases to infinity, λ is kept constant so that

$$p = \lambda\delta = \lambda t/n \quad . \tag{6.3}$$

As the intervals get smaller, the probability of success in each interval also gets smaller. A success or a "hit" in an interval δ is typically preceded by a series of failures as is shown in Fig.(6.3).

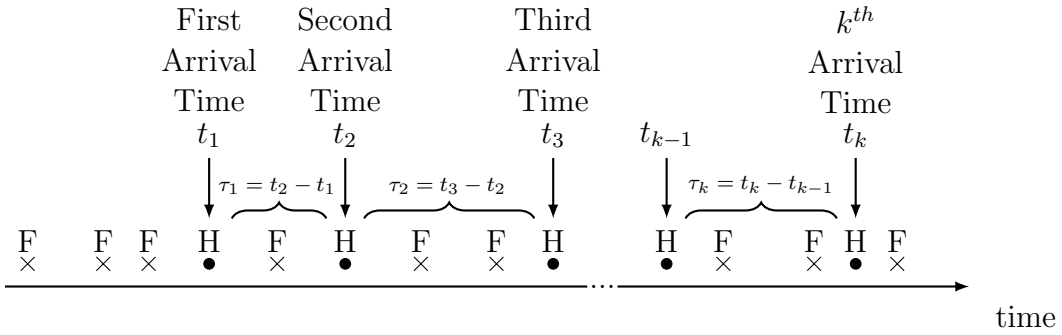


Figure 6.3: The event inter-arrival times of the Poisson process.

Let $N(t)$ denote the number of successes in the time from 0 to t. As there are n intervals and p is the probability of success per interval it follows that

$$np = \lambda t \quad . \tag{6.4}$$

In the limit $\delta \to 0$ as $n \to \infty$ with the realisation of Eq.(6.4), the distribution of $N(t)$ is Binomial with parameters n and p [134] :

$$P(S = k) \;=\; \left(\begin{array}{c} n \\ k \end{array} \right) p^k q^{n-k} \qquad \text{with } p = \lambda t/n \quad . \tag{6.5}$$

So

$$\begin{aligned} P(S = k) \;&=\; \frac{n!}{(n-k)!k!} \left(\frac{\lambda t}{n}\right)^k \left(1 - \frac{\lambda t}{n}\right)^{n-k} \\ &=\; \frac{(\lambda t)^k}{k!} \frac{1}{n^k} \left(1 - \frac{\lambda t}{n}\right)^n \left(1 - \frac{\lambda t}{n}\right)^{-k} \frac{n!}{(n-k)!} \end{aligned} \tag{6.6}$$

Making use of $\frac{n!}{(n-k)!}\frac{1}{n^k} = \frac{n(n-1)\cdots(n-k+1)(n-k)!}{n^k(n-k)!} \quad \approx 1$ as $n \to \infty$,
$\lim_{n\to\infty} \left(1 - \frac{\lambda t}{n}\right)^n = e^{-\lambda t}$ and $\lim_{n\to\infty} \left(1 - \frac{\lambda t}{n}\right)^{-k} = 1$, Eq.(6.6) becomes

$$P(S = k) = \left(\begin{array}{c} n \\ k \end{array} \right) p^k (1-p)^{n-k} \Rightarrow \frac{(\lambda t)^k}{k!} e^{-\lambda t} \quad . \tag{6.7}$$

This is the probability distribution for a Poisson process which is found as the limiting case of the Binomial distribution. The value $k = 0$ corresponds to the survival probability (i.e. no hits). In a Poisson count process the distribution of the number of events (arrival times - hits) in any interval depends only on the length of the interval and not on the exact location of the interval on the time axis, hence the Poisson process has stationary increments.

6.1.1 What is "n" in the Current Picture?

In the current context of a proton beam incident on a series of cells (target) there are many factors that influence the relative probabilities. If we think about this as a Bernoulli counting experiment, the number of protons incident on the cells can be defined as n. Bearing in mind that the aim in a dose-response experiment is to expose the cells to increasing absorbed doses and we must increase n to achieve this. Due to the variance in ionisation density along a particle track with changing LET, larger n is required for a low-LET radiation to achieve the same dose range as for a high-LET radiation.

With increasing n, the probability of $k > 0$ successes (i.e. the induction of k DNA damages) increases. This trivially results in a decrease of the survival probability ($k = 0$) of the target/cells on which the n protons are incident.

In our physical picture, Fig.(6.1), we may think of an experiment such as in Fig.(6.3) repeated infinitely often and adjusting the success probability p to the limit $n \to \infty$ by keeping $n \cdot p = \Lambda$ fixed. Then we finally look at the outcome that S takes on the integer value k. If we wish to determine the survival probability in particular, we observe the probability that $k = 0$, i.e. the fraction of targets that are untouched (the cells are alive). Once we determine the only parameter Λ from this survival probability, all the other probabilites are given by Eq.(6.7).

The probability functions $P(k)$ for $k = 0, 1, 2, \ldots$, satisfy the Feller-Kolmogorov [135] forward equations

$$\frac{d}{dt}\, p_0 = -\lambda p_0 \tag{6.8}$$

$$\frac{d}{dt}\, p_k = \lambda p_{k-1} - \lambda p_k \qquad \text{for} \quad k = 1, 2, \ldots \tag{6.9}$$

with the initial conditions $p_0(0) = 1$ and $p_k(0) = 0$ for $k = 1, 2, \ldots$ because at $t = 0$ all the cells are alive and no damage has been induced.

Now let $N(t)$ be a Poisson count process with rate λ and τ_1 be the time passed until the first success then

$$\begin{aligned} P(\tau_1 > t) &= P(\text{no arrival in } (0,t]) \\ &= e^{-\lambda t}, \end{aligned} \tag{6.10}$$

hence the survival probability at time t is given by

$$F_{\tau_1}(t) = \begin{cases} 1 - e^{-\lambda t} & t > 0 \\ 0 \quad \text{otherwise}, \end{cases} \tag{6.11}$$

where $F_{\tau_1}(t)$ is the CDF and τ_1 is distributed exponentially. Let τ_1 denote the time between the first and second arrival, τ_2 is also distributed exponentially where both τ_1 and τ_2 are independent. $\tau_1, \tau_2, \ldots$ are called the inter-arrival (waiting) times (or waiting times) of the count process $N(t)$. All the τ's are independent and exponentially distributed; and this process suggests no memory [136].

After showing that a Poisson count model having an exponential distribution on the inter-arrival times we can now find the distribution of the arrival times $T_1 = \tau_1, T_2 = \tau_1 + \tau_2, \ldots, T_n = \tau_1 + \tau_2 + \ldots \tau_n$. The τ_i's are independent exponential random variables and the summation of exponential distributions add up to a Gamma distribution $T_n \sim \Gamma(n, \lambda), \quad n \in \mathbb{N}$. The $\Gamma(n, \lambda)$ is also called the Erlang distribution [134].

$$T_n \sim Erlang(n, \lambda) = Gamma(n, \lambda) \quad \text{for} \quad n = 1, 2, 3, \cdots \tag{6.12}$$

The PDF of T_n is

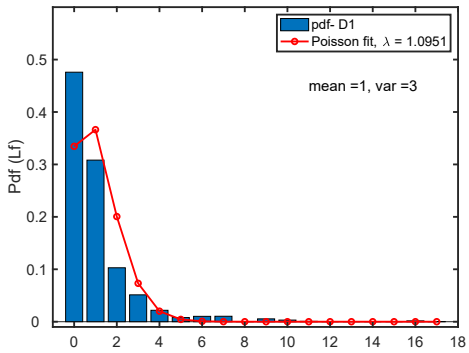
$$f_{T_n}(t) = \frac{\lambda^n t^{n-1} e^{-\lambda t}}{(n-1)!} \quad \text{for} \quad t > 0 \tag{6.13}$$

with $E[T_n] = n/\lambda, \quad \text{var}(T_n) = n/\lambda^2$. Note here that the arrival times are not independent as $T_1 \leq T_2 \leq \cdots$.

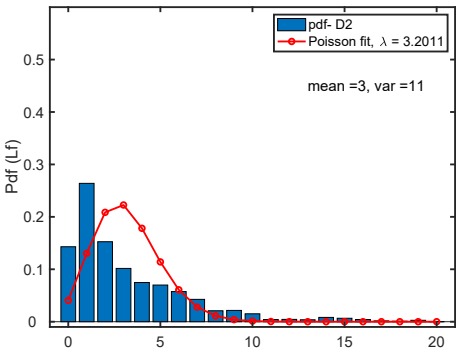
Recall that most dose-response models in the literature are based on a Poisson process where the mean and variance are equal $E[x] = \text{VAR}[x] = \lambda$, which suggests an *equidispersion* condition. The exponential inter-arrival times (waiting times) for the pure Poisson process have a PDF, $f(t) = \lambda e^{-\lambda t}$, CDF, $F(t) = 1 - e^{-\lambda t}$ and the survival function $\Psi(t) = 1 - F(t) = e^{-\lambda t}$ gives the hazard rate $h(t) = \lambda$ which is constant, indicating no duration dependence. The hazard rate defines the underlying time dependence of the process.

We can utilize here some of our simulated dose-response results, Figs.(6.4) and (6.5) from Chapter 8 to check whether the equal mean and variance condition of the Poisson process is satisfied in distributions of lethal lesions in the system after exposure to high and low-LET radiation with increasing dose. From these figures we observe a fatter tail in the high-LET case compared with the low-LET case, which indicates an overdispersed data due to the variance being considerably greater than the mean.

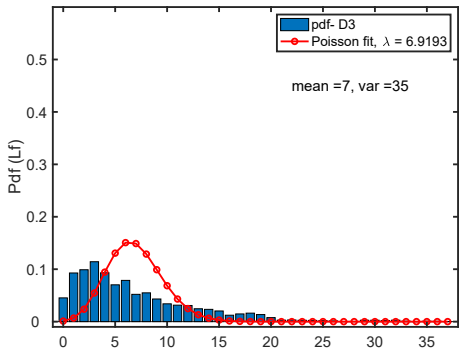
In the high-LET case, Fig.(6.4), increasing dose the difference between the mean and variance becomes greater, the variance exceeds the mean (overdispersion) and the data does not satisfy the equal mean and variance (equidisperison) condition of the Poisson process. On the other hand Fig.(6.5(h)) shows the probability of lethal damages with the Poisson distribution fit seems to represent the data much better, especially in the intermediate to high dose region. Whereas for low doses the



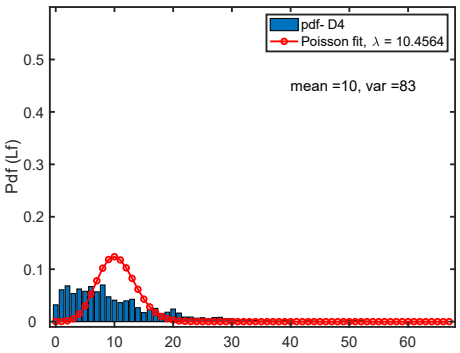
(a) LET = 35keV/μm, D = 1.18Gy



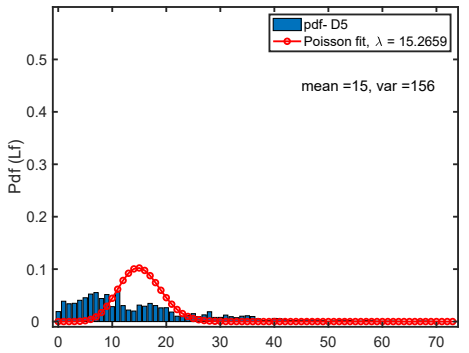
(b) LET = 35keV/μm, D = 2.36Gy



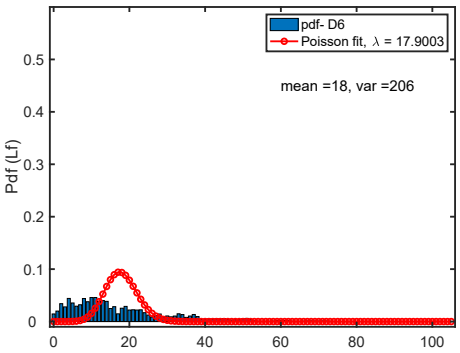
(c) LET = 35/μm, D = 3.54Gy



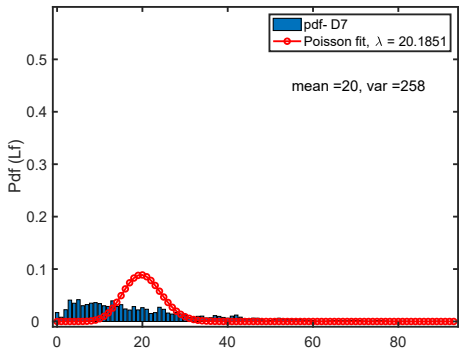
(d) LET = 35keV/μm, D = 4.73Gy



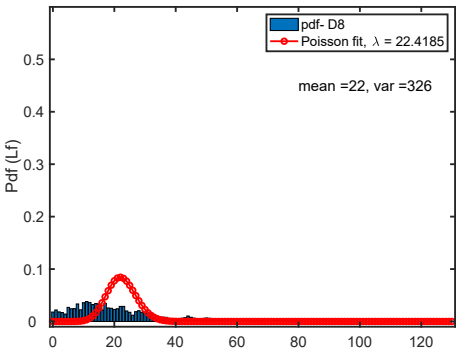
(e) LET = 35keV/μm, D = 5.67Gy



(f) LET = 35keV/μm, D = 6.39Gy

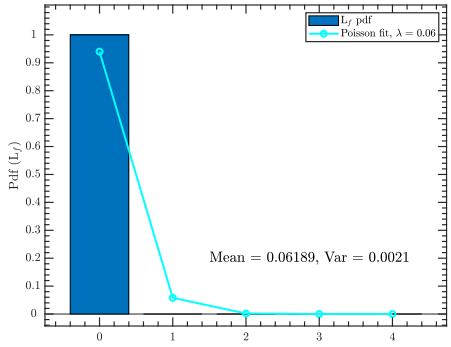


(g) LET = 35keV/μm, D = 6.85Gy

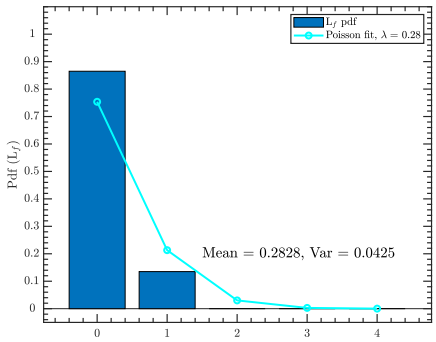


(h) LET = 35keV/μm, D = 7.34Gy

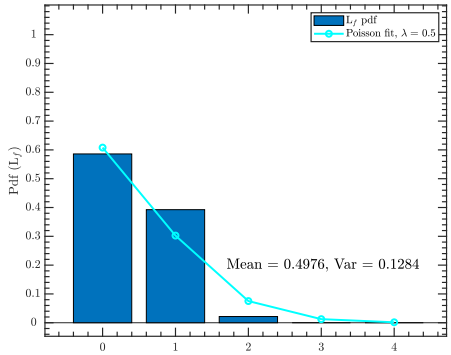
Figure 6.4: Normalised histograms of lethal lesions induced in our system with increasing doses, D, of high-LET radiation with the fitted Poisson PDF.



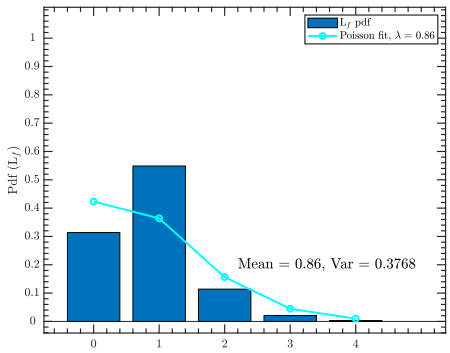
(a) LET = 1.20keV/μm, D = 0.96Gy



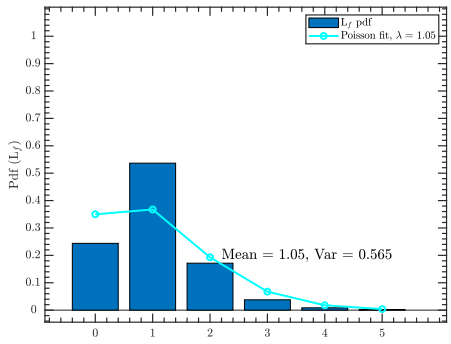
(b) LET = 1.20keV/μm, D = 2.14Gy



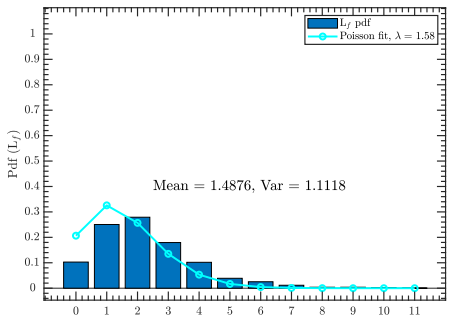
(c) LET = 1.20keV/μm, D = 2.89Gy



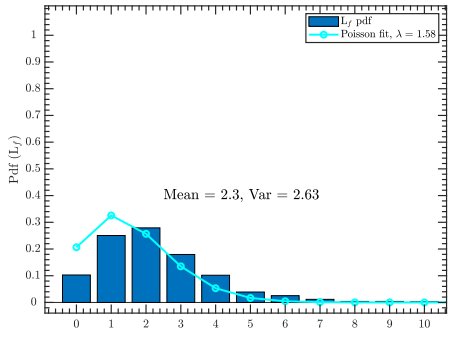
(d) LET = 1.20keV/μm, D = 3.85Gy



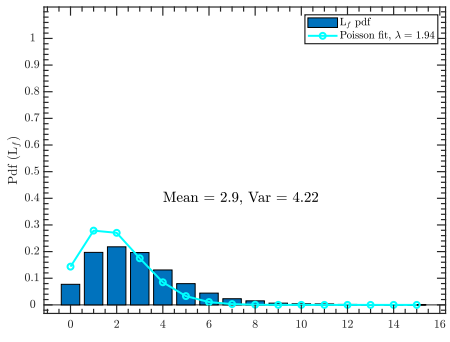
(e) LET = 1.20keV/μm, D = 4.28Gy



(f) LET = 1.20keV/μm, D = 5.13Gy



(g) LET = 1.20keV/μm, D = 6.40Gy



(h) LET = 1.20keV/μm, D = 7.30Gy

Figure 6.5: Normalised histograms of lethal lesions induced in our system with increasing doses, D, of low-LET radiation with the fitted Poisson PDF.

mean exceeds the variance suggesting under-dispersion (see Figs.(6.5(a)) to (6.5(c))).

This explains why the LQ model, which is based on the Poisson process, is problematic in describing survival data in the high-LET region in particular, Fig.(5.8). We can therefore argue that the low-LET region is best suited to be explained with the Poisson process but not the high-LET region.

The presence of over/under-dispersion is not uncommon in nature [137]. In fact this effect is observed in many Poisson counting models for complex systems such as that considered here [138]. In such instances the parameters estimated from the distribution are biased. In radiation induced cell death accounting for overdispersion effects is critical to radiation therapy treatments, because it results in a different relationship between the mean lethal lesion yield per cell and the survival probability ($k = 0$). The result of this is an inaccurate prediction of cell survival. Therefore an non-Poissonian alternative is needed to ensure an accurate estimation of the survival fraction with a given mean lethal lesion yield per cell when overdispersion is present.

Recalling the differences in the induction of DNA damage under high and low-LET radiation, the relative complexities of DNA damage will vary to a large degree due to a more densely ionising particle track for high-LET. Furthermore, a damaged cell can undergo a series of biological and chemical processes to repair radiation induced damages proceeding via a number of pathways (see Chapter 2). Each process can influence the lethal lesion yield in the final state $|f\rangle$ and it is therefore plausible to suggest over/underdispersion effects with a deviation from the Poisson process [68, 69].

In Chapter 5 we concluded that the Linear-Quadratic model based on a Poisson process has difficulty in representing high-LET experimental data, Fig.(6.5(h)) and therefore must be improved. However, to our knowledge, no model has been proposed to resolve this issue. At a microscopic level, by observing all other probabilities of the Poisson process rather than just the zeroth order (lowest order) namely the survival probability, we deduced that our simulated data supports this conclusion. As we show in Figs.(6.4) and (6.5) the equidispersed condition of the Poisson process is violated for high LET and dose.

As a remedy for an overdispersed/underdispersed Poisson process in the literature of complex systems [137], the exponential waiting times are replaced by the Weibull, Gamma, Lognormal or power law [139]. These distributions have a non-constant hazard rate and are therefore less restrictive. In a complex system deviations from the classical exponential waiting times can be achieved by introducing the stretched exponential distribution instead, which introduces "*Anomalous relaxation models*".

6.2 Proposal of a New Dose Response Model

The model proposed here was developed in close collaboration with Dr. Markus Kreer of Johann Wolfgang Goethe-Universität, Frankfurt, Germany. In order to improve on the problems discussed in the previous section and make an attempt to resolve them, we propose a new dose-response model where the inter-arrival times (waiting times) between the events are independent but NOT exponential. We modify the classical exponential inter-arrival time distribution by a **Mittag-Leffler** distribution [140,141] which is a generalisation of the exponential distribution. Moreover the stochastic process dictates that the Mittag-Leffler distribution is called the **Fractional Poisson Process (fPp)**. The main idea of the fPp process is to make the standard Poisson process more flexible by permitting a non-exponential, heavy tailed distribution for the inter-arrival times with long memory.

6.2.1 Fractional Poisson Process

The introduction of Mittag-Leffler distributed waiting times with index γ turns out to be equivalent to substituting the 1st order time derivative on the left-hand side in the Feller-Kolmogorov forward equations, Eqs.(6.8) and (6.15), by a fractional derivative $\frac{d^\gamma}{dt^\gamma}$ where $0 < \gamma \leq 1$, which is to be interpreted in the sense of Dzerbayshan-Caputo (e.g. Podlubny (1999, p. 78) [142]).

The probability functions p_k for $k = 0, 1, ...$, satisfy the Feller-Kolmogorov forward equations

$$\frac{d^\gamma}{dt^\gamma}\, p_0^{(\gamma)} = -\lambda p_0^{(\gamma)} \quad , \tag{6.14}$$

$$\frac{d^\gamma}{dt^\gamma}\, p_k^{(\gamma)} = \lambda p_k^{(\gamma)} - \lambda p_{k-1}^{(\gamma)} \qquad \text{for} \quad k = 1, 2, ... \tag{6.15}$$

with initial conditions $p_0^{(\gamma)}(0) = 1$ and $p_k^{(\gamma)}(0) = 0$ for $k = 1, 2, ...$ [143–145]. Our new survival probability for the fractional Poisson process (fPp) is given by

$$p_0^{(\gamma)}(t) = 1 - F(t) = 1 - (1 - E_\gamma(\lambda t^\gamma)) \tag{6.16}$$

$$= E_\gamma(-\lambda t^\gamma) \quad , \tag{6.17}$$

where $E_\gamma(-\lambda t^\gamma)$ is the Mittag-Leffler (ML) function with parameters γ and λ. The details of the ML function can be found in Appendix I. The other theoretical probabilities, according to Eq.(25) in Ref. [143] are given by the following equation

$$p_k^{(\gamma)}(t) = \frac{(\lambda t^\gamma)^k}{k!} \sum_{j=0}^{\infty} \frac{(j+k)!}{j!} \frac{(-\lambda t^\gamma)^j}{\Gamma(\gamma(j+k)+1)}, k = 1, 2, \cdots \quad . \tag{6.18}$$

Our series expansions for $p_k^{(\gamma)}(t)$ as given in Eq.(6.18) will have a very limited range of convergence for practical purposes whilst being numerically unstable in other ranges (even though the radius of convergence is infinite). Therefore we need to use a different approach to evaluate the function. We use the following integral representation of Eq.(6.18) from [146] Corollary 1 :

$$p_k^{(\gamma)}(t) = \frac{1}{k!}\,\lambda^k t^{k\gamma} \int_0^\infty dz\; z^k e^{-z\lambda t^\gamma} M_\gamma(z), \tag{6.19}$$

where the M-Wright function or Mainardi function is defined by

$$M_\gamma(z) = \frac{1}{\pi}\sum_{j=1}^{\infty} \frac{(-z)^{j-1}}{(j-1)!}\Gamma(\gamma j)\sin(\pi\gamma j), \tag{6.20}$$

which is convergent [142]. It can be shown [147] that we have the following asymptotic behavior as $z \to \infty$

$$M_\gamma(z/\gamma) \sim a(\gamma) z^{(\gamma-1/2)/(1-\gamma)} \cdot \exp\left[-b(\gamma) z^{1/(1-\gamma)}\right], \tag{6.21}$$

where $a(\gamma)$ and $b(\gamma)$ are positive coefficients with

$$a(\gamma) = \frac{1}{\sqrt{2\pi(1-\gamma)}}, \quad b(\gamma) = \frac{1-\gamma}{\gamma} \quad .$$

Thus, for any $\gamma \in (0,1)$ the Mainardi function decays faster than the exponential function. This allows the numerical evaluation of Eq.(6.19) by approximating the integral numerically by a simple trapezoidal rule. We can easily evaluate the series expansion in Eq.(6.20) with the asymptotics Eq.(6.21) to provide the Mainardi function for the entire positive x-axis needed in our case. For $\gamma = 1/2$ and $\gamma = 1/3$ there exists an analytic expression of the Mainardi function :

$$\begin{aligned} M_{1/2}(z) &= \frac{1}{\sqrt{\pi}} e^{-z^2/4} \\ M_{1/3}(z) &= 3^{2/3}\mathrm{Ai}(z/3^{1/3}), \end{aligned}$$

where Ai is the Airy function. In Fig.(6.6) we have depicted the Mainardi function for various values of γ. For $\gamma = 1$ the Mainardi function will be the delta-distribution, because $M_1(z) = \delta(z-1)$. Therefore the integral representation of the fractional Poisson process probabilities Eq.(6.19) exactly goes to the Poisson distribution $p_k(t) = \frac{t^k}{k!}e^{-t}$. We describe our first algorithm (I) to calculate the Mainardi function for a given $0 < \gamma < 1$ in Appendix I. Our second algorithm (II) computes the fractional Poisson probabilities $p_k^{(\gamma)}(t)$ for the fPp having encountered at time $t > 0$ for k events, also given in Appendix I. As an example, using these two algorithms we compare the Poisson and fractional Poisson distributions for $\gamma = 0.8$ with increasing time in Fig.(6.7).

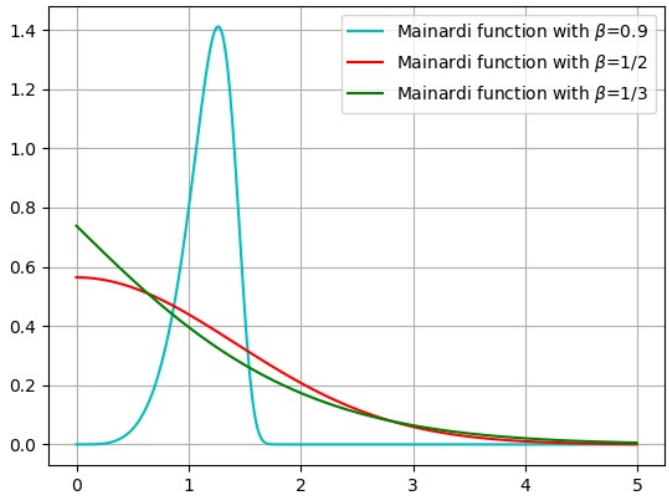


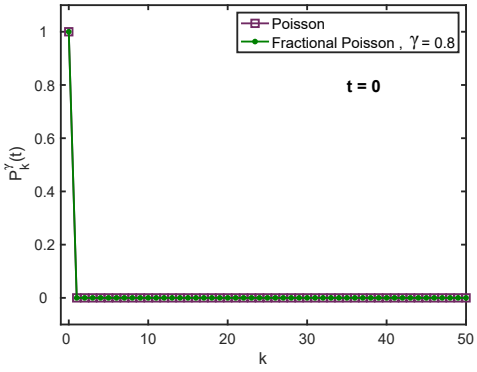
Figure 6.6: Mainardi function for different values of γ.

If in our physical picture we visualise the PDF evolution with time by considering the expected number of lethal lesions per cell as irradiation time t_{del} increases, the evolution can be explained in terms of cell response. At $t = 0$ we assume no hits have occurred because the beam has not been "switched on", therefore it is impossible for $k > 0$ hits to occur and the $k = 0$ probability is 1 under the fPp and Pp.

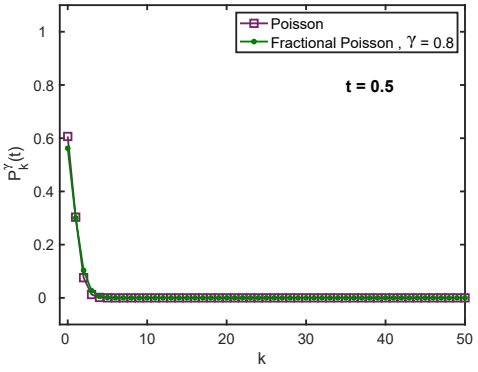
When irradiation commences but t is small the Pp and fPp are very similar (Figs.(6.7(b)) and (6.7(c))) because only relatively simple DNA damages are likely to exist. As we have described above, when irradiation continues for large t we begin to observe the overdispersion effect as DNA damage complexities increase and the repair processes begin to affect the number of lethal lesions per cell. Therefore the Pp and fPp PDFs will only continue to diverge as irradiation continues and the resultant survival probability will differ by larger amounts.

Finally at $t = 50.0$ irradiation has continued for a long period of time and the system begins to "oversaturate" with damage resulting in the extremely flat PDF shown in Fig.(6.7(h)) for the Pp and fPp. At large t the effect of overdispersion on the estimate of the mean number of lethal lesions per cell is clearly visible. Under a standard Poisson process the mean will be severely under estimated along with the survival probability.

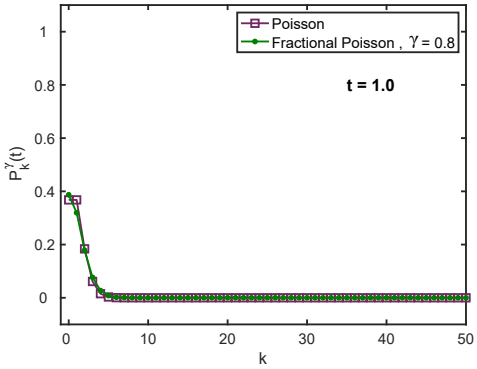
Figure (6.8) depicts the evolution of survival probability with time for the standard and fractional Poisson process for $\gamma = 0.8$. Similar to Fig.(6.7) the survival fraction under the Pp and fPp begin at 1 at $t = 0$. As t increases the survival



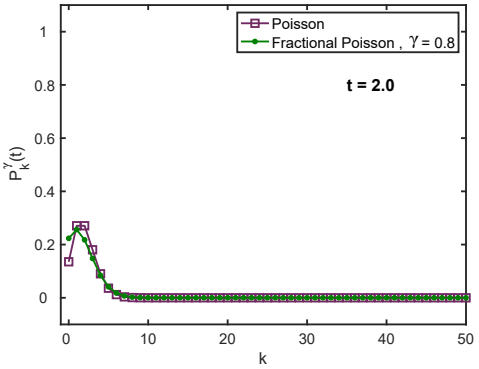
(a) FPp and Pp at $t = 0$



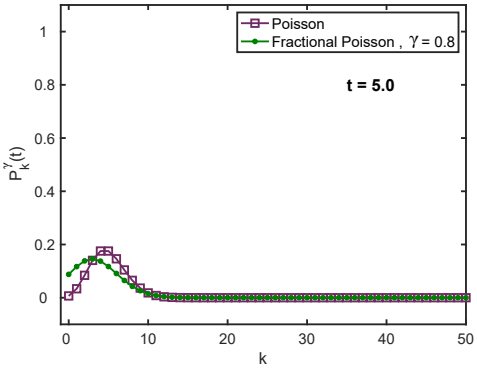
(b) FPp and Pp at $t = 0.5$



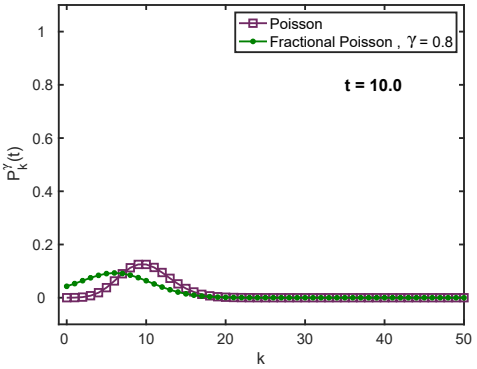
(c) FPp and Pp at $t = 1$



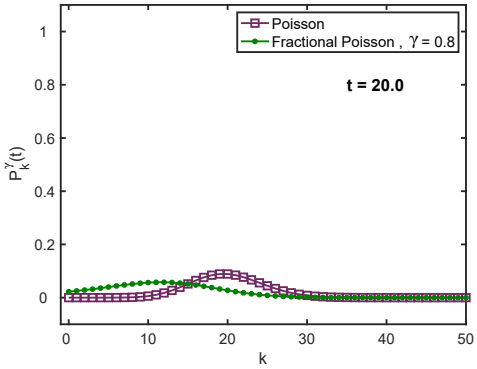
(d) FPp and Pp at $t = 2$



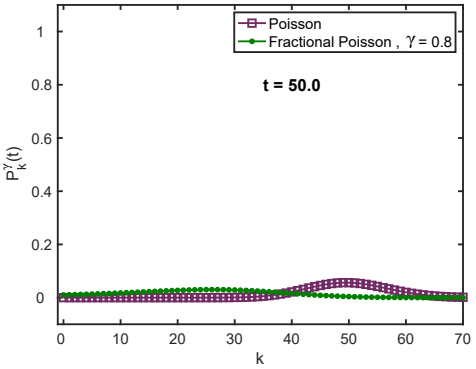
(e) FPp and Pp at $t = 5$



(f) FPp and Pp at $t = 10$



(g) FPp and Pp at $t = 20$



(h) FPp and Pp at $t = 50$

Figure 6.7: Evolution of the Pp and fPp PDFs as a function of k (hits) with increasing time.

probability in the classical and fractional limits diverge from each other. It is in the same time region of Fig.(6.7) that we observe the largest difference between the mean number of lethal lesions per cell as expected. For large t we again observe the system to be saturated with lethal lesions rendering the survival fraction effectively zero for large t despite the difference in the mean lethal lesion count per cell in Figs.(6.7(g)) and (6.7(h)).

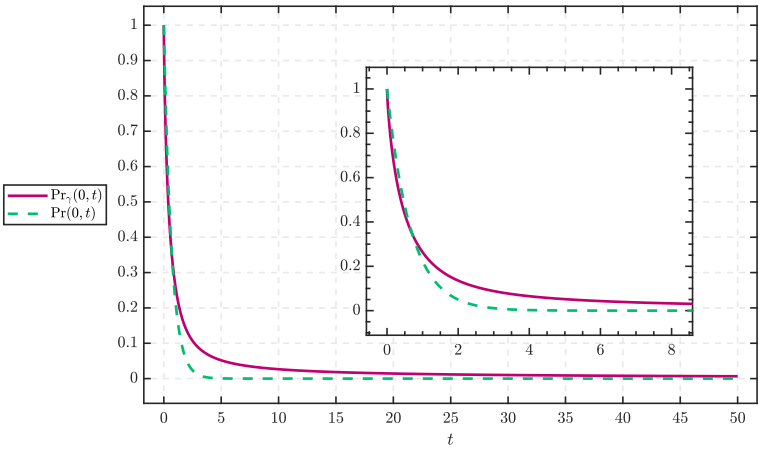


Figure 6.8: The survival probability functions of the classical ($\gamma = 1$, $\lambda = 1.5$) and fractional ($\gamma = 0.8$, $\lambda = 1.5$) Poisson processes using the Mittag-Leffler function.

We give our third algorithm (III) in Appendix I to fit the distribution of lethal damages, L_f, to the fractional Poisson probabilities , p_k^γ, for $k = 1, 2, \cdots$ using the integral representation Eq.(6.19) after rescaling it by $\Lambda \equiv \lambda t^\gamma$:

$$p_k^{(\gamma)}(\Lambda) = \frac{1}{k!}\,\Lambda^k \int_0^\infty dz\; z^k e^{-z\Lambda^\gamma} M_\gamma(z). \tag{6.22}$$

6.3 Our Continuous Random Walk Model as a Realisation of the Fractional Poisson Process

It is constructive to apply the concepts described above to the current physical picture. Let us consider a beam of protons incident on a group of cells. The cells contain DNA targets or more generally, molecules that come into contact with protons to cause lethal damages (leading to cell death). If the time difference (waiting times) between these lethal damages are described by the Mittag-Leffler distribution, then the foundation of this physical picture is a fractional Poisson process instead of the standard Poisson process.

For the sake of completeness let us sketch how we can arrive constructively at a fractional Poisson Process. We wish make a simple dose-response model based on the Fractional Poisson process using the Mittag-Leffler (ML) distribution as an inter-arrival time. In order to achieve this we adopted the argument given by Meerschaert et al. [148] Consider a random waiting time J_i described by a fat-tailed distribution such as ML. The total time elapsed for n events to occur is

$$T_n = J_1 + J_2 + ... + J_n$$

and the corresponding renewal process[1] can be described as

$$R(t) = \max\{n \geq 0 : T_n < t\} \quad . \tag{6.23}$$

We consider our physical picture to be a renewal process the times at which a cell is classified as "alive" or "dead" is i.i.d. and random. Reconsider the Bernoulli experiment for the case of a continuous time random walk (CTRW) related to this renewal process :

$$S^{(p)}(R(t)) = S_n^{(p)} = X_{J_1}^{(p)} + X_{J_2}^{(p)} + ... + X_{J_n}^{(p)} \quad . \tag{6.24}$$

As in the previous section, we take the limit $n \to \infty$ as $p \to 0$ but in such a way that it is described by Meerschaert et al. [148]. They made this limit behavior exactly available for the CTRW $S^{(p)}(R(t))$ in their Theorem 2.5 and for the Mittag-Leffler waiting times the result is stated in their remark 2.6 as :

$$S^{(p)}\left(\lambda R(p^{-1/\gamma}t)\right) \Rightarrow N_1(E(t)) \qquad \text{as} \quad p \to 0 \quad ,$$

where $N_1(E(t))$ is a so-called fractional time Poisson process (ftPp) and by using this form of re-scaling corresponds to a special thinning procedure. The convergence is "in distribution". In Chapter 2 of [148] they demonstrate that the fractional time Poisson process and fractional Poisson process are actually the same. Thus, our heuristic approach in the previous subsection by replacing the ordinary derivative with a fractional derivative did indeed lead to a generalisation of the approach.

The continuous time random walk (CTRW) as a physical picture means that single protons of our beam "hit" the target with inter-arrival times distributed as a "stretched" Mittag-Leffler with critical exponent $0 < \gamma \leq 1$. In the case $\gamma = 1$ we recover the standard Poisson process. This allows us to formulate a theory that will be tested in the coming chapters of this thesis and is stated as follows.

[1] A renewal process investigates the sum of usually non-exponential waiting times as well as the corresponding counts occuring in these time intervals. Thus is it a generalisation to the *Erlang* process for the sum of waiting times.

When the time interval (waiting time) between the occurrences of the final lethal lesion damages is dictated by the Mittag-Leffler function with degree $0 < \gamma \leq 1$, then the underlying process is not the standard Poisson but a fractional Poisson process. Therefore, the probability of a given cell surviving after irradiation at time t is denoted by

$$Pr(0,t)^{(\gamma)} = E_\gamma(-\Lambda t^\gamma), \tag{6.25}$$

where $E_\gamma(-\Lambda t^\gamma)$ is the Mittag-Leffler function. Note that for $\gamma < 1$ the Mittag-Leffler function has power-law asymptotic behavior namely $\Pr(0,t) \sim t^{-(1+\gamma)}$. Therefore the waiting time is much "longer" as compared to the exponential waiting time ($\gamma = 1$) and this means a "thinning procedure", i.e. the protons induce damage less frequently.

However, we cannot stop here. Recall that dose-response data is typically not presented as a function of time, but as a function of absorbed dose. To present the fPp model in terms of absorbed dose, we must define a link to time.

6.3.1 Dose versus Time

In the picture of irradiated cells, neither the time of protons hitting the target nor the inter-arrival time differences between the "hits" are available but only the knowledge of the dose deposited at the time following irradiation is accessible. Since the dose deposited into the target or dose absorbed by the target will be proportional to the number of damages produced, absorbed dose and time are also related to each other. The number of lethal damages produced increases with dose and, at a constant dose rate, delivering higher doses takes a longer time. By increasing LET more complex damages are produced and therefore the repair process commences very quickly.

Let us put forward a heuristic argument by relating dose to time. Here we argue that $D = D(t)$ is a (strictly) monotonically increasing function in its argument t (see Fig.(6.9)), suggesting that $D(.)$ being linear in its argument t is quite a good approximation (see Fig.(6.10)). The relationship can be easily tested using TOPAS. A simple $100 \times 100 \times 100\mu$m liquid water cube is irradiated with 200,000 primary protons of mean energy 1, 10 and 100MeV over 200 seconds (i.e. 55 protons per minute)[2]. A "DoseToMedium" scorer is implemented to measure the absorbed dose every 10 seconds of delivery time. In all three cases the absorbed dose appears to follow a linear relationship with time.

Furthermore, the assumption that $D(.)$ is a concave function of some argument is also a plausible. For example, when the cells are exposed to radiation, there occurs some saturation after a certain time (repair time) and the effective dose will grow

[2]Ample coverage of clinical protons energies considered in the current study.

sublinearly in t. We can argue that, due to the repair process which already begins during the irradiation, some proportion of damage effects will be compromised (repaired) by the biological responses. These repair processes will change the nature of damages caused by the ionisation (absorbed dose) in some stretched time (repair time).This can be interpreted as the effective dose leading to a sublinear relationship in effective time, $\tilde{D} = \tilde{D}(\tilde{(t)})$, which we argue to be a strictly monotonically increasing and concave function. Therefore an inverse function $t = t(D)$ exists which is also strictly monotonically but convex in its argument D, Fig.(6.9).

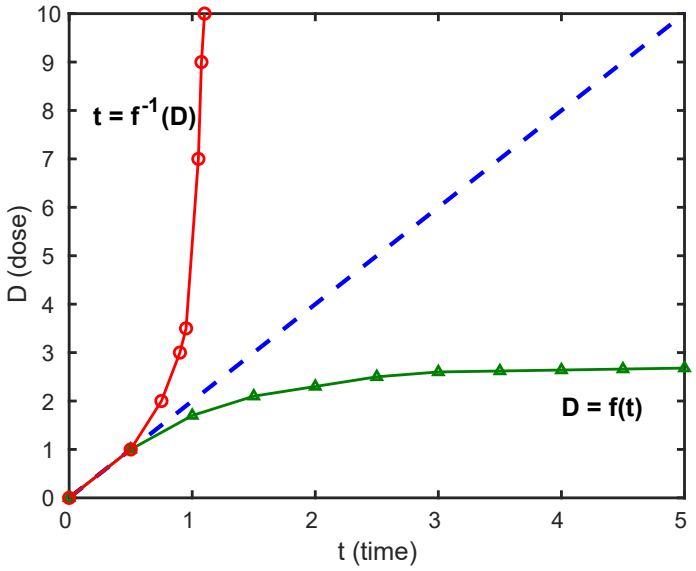


Figure 6.9: Concave monotonic increasing function and its inverse function.

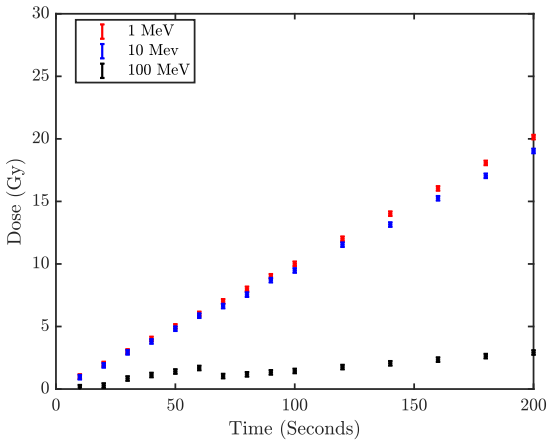


Figure 6.10: Dose as a function of beam delivery time for a 1, 10 and 100 MeV proton beam incident on a 100 × 100 × 100 μm liquid water medium.

As examples, we choose three possible candidates for convex argument t

$$t = aD + bD^2, \tag{6.26}$$

$$t = \left(\frac{D}{a}\right)^b \qquad \text{for } b > 1 \tag{6.27}$$

$$t = \left(\frac{D^2}{2a}\right). \tag{6.28}$$

Due to time restrictions we will only explore Eq.(6.26) in this thesis. The analysis of Eqs.(6.27) and (6.28) will be presented elsewhere. By substituting the dose-time argument into the survival probability function for the classical and fractional Poisson processes we recover the dose-response curve as a function of absorbed dose D under each process using the probabilities that S takes on $k = 0$, i.e. there are no hits (see Figs.(6.11) and (6.12)).

6.3.2 Bringing Everything Together

We can utilise Eq.(6.26) of $t = t(D)$ listed to map the survival probability expression for the fPp process to absorbed dose, allowing this model to be tested on experimental data. This gives rise to :

$$Pr^{(\gamma)}(t) = E_\gamma\big[-\lambda(\overline{a}D + \overline{b}D^2)^\gamma\big]. \tag{6.29}$$

We can make simplifications to the expression to decrease the number of fitting parameters, hence decreasing the degrees of freedom.

$$Pr^{(\gamma)}(t) = E_\gamma\big[-(aD + bD^2)\big] \tag{6.30}$$

We have now adapted the generalised fPp form to fit dose-response curves. In Chapter 8 we will test our model (i) against experimental data described in Appendix D and (ii) against our simulated data. The testing will proceed via the workflow provided in Fig.(8.9). **Note:** the computation of the Mittag-Leffler function is not trivial due to numerical problems in the series expansion. We will make use of its integral representation Eq.(6.19) to an arbitrary accuracy [149].

6.4 Summary

In this thesis, we discuss the physics of proton radiation-induced lethal damages in a cell (target) described with increasing LET converging to a fractional Poisson process. This leads to the survival fraction to follow a Mittag-Leffler (ML) distribution. The ML survival fraction will be tested against experimental and simulated dose-response data, where its performance will be compared with the LQ model. A rigorous statistical evaluation will be performed to assess the performance of the

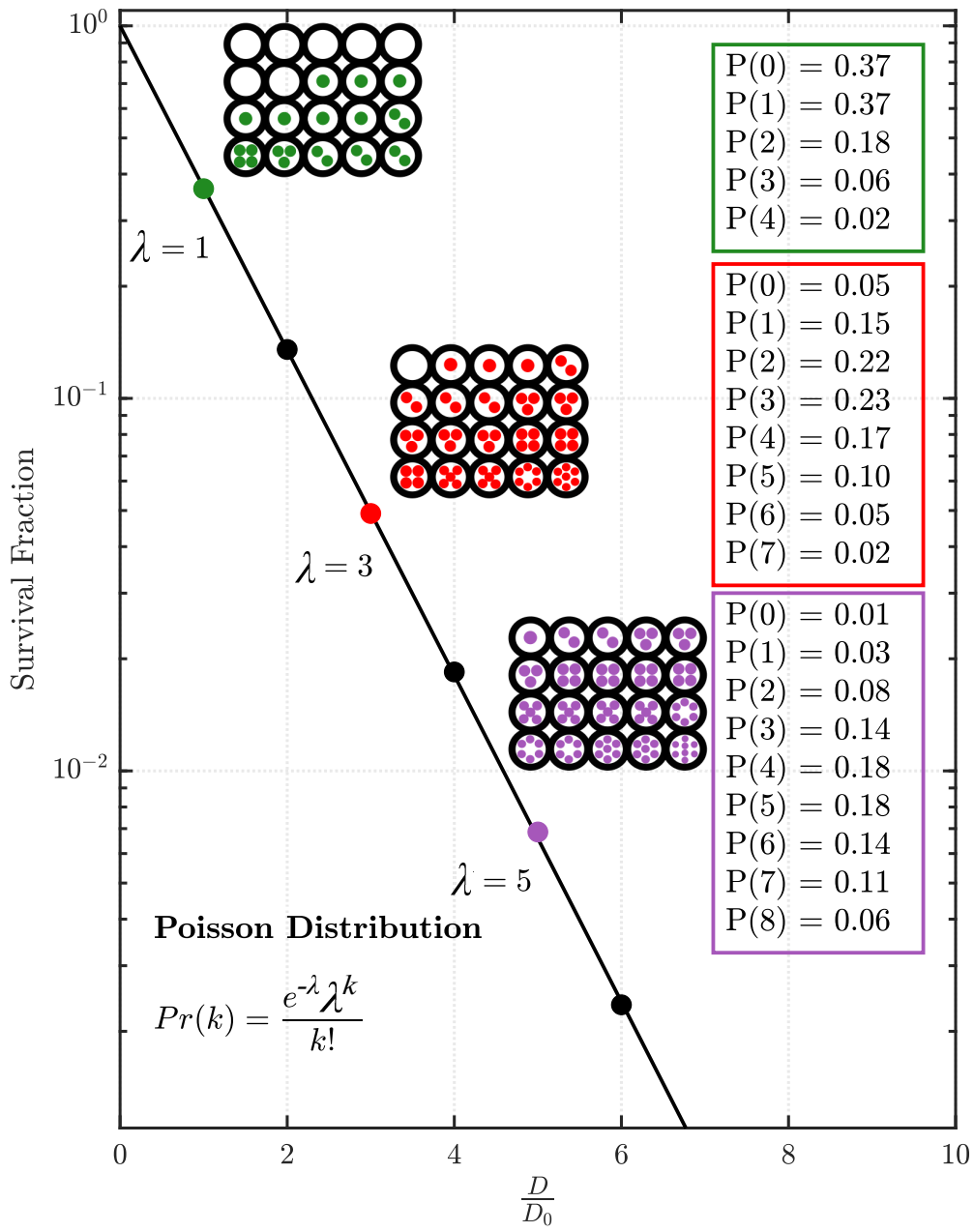


Figure 6.11: A survival fraction constructed from Poisson distribution probabilities, $1 - \exp(-\lambda)$, for the mean number of lethal lesions produced in the system, λ. Graphical depictions of the cell damage probabilities are shown for select SF. Credit to [111] for the original figure idea. Note: the total probabilities should sum to 1, however the probabilities < 0.01 are not shown.

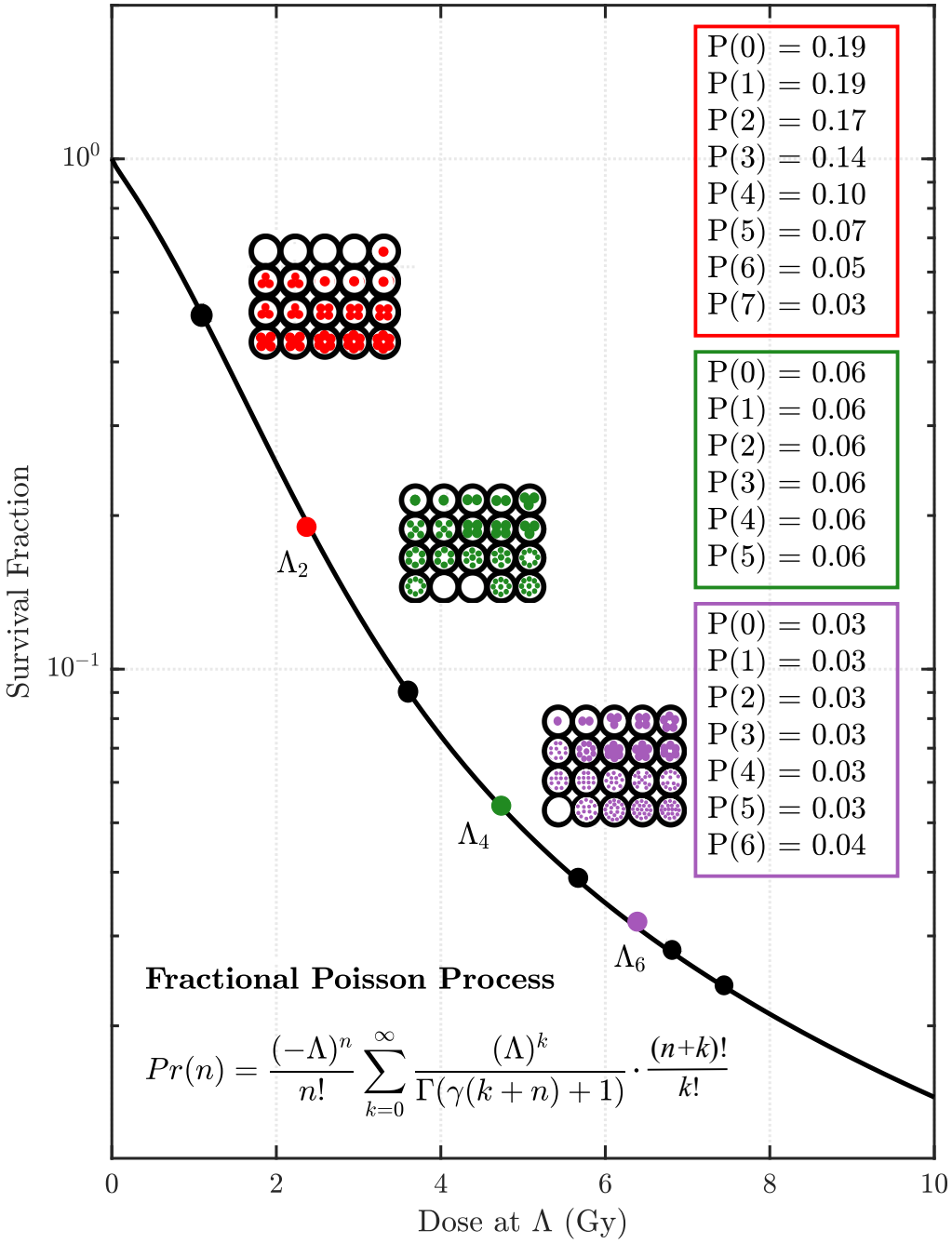


Figure 6.12: Survival probabilities as a function of the dose required to produce a given number of lethal lesions Λ assuming a fractional Poisson process with $\gamma = 0.60$ with a graphical depiction. Note: due to the large distribution of lethal lesion counts, not all probabilities are shown and the graphical depictions are approximations of the numerical probabilities. Note: the total probabilities should sum to 1, however the probabilities < 0.01 are not shown.

fPp model in comparison to the classical Poisson model, from which the LQ model is derived.

In summary, the two models to be compared are defined as follows.

The Standard Poisson Process
A beam of n protons incident on a group of cells containing a series of molecules to interact with is exposed to dose D on the time interval t, where $D \propto n$. After exposure to dose D, the probability that k molecules inside of the cell have experienced an interaction resulting in a DNA lesion is used to determine its probability of survival at that instant using,

$$\mathrm{SF(D)} = \exp\big(-\alpha D - \beta D^2\big) \quad .$$

The dose D is defined in terms of t and hence it is assumed that the proton beam is delivered to the cells at constant contact flux $\dot{D}(t)$.

The Fractional Poisson Process
A beam of n protons incident on a group of cells containing a series of molecules to interact with is exposed to dose D on the time interval t, where $D \propto n$. After exposure to dose D, the probability that k molecules inside of the cell have experienced an interaction resulting in a DNA lesion is used to determine it's probability of survival at that instant using,

$$\mathrm{Pr}^{(\gamma)}(t) = E_\gamma\big[-(aD + bD^2)^\gamma\big] \quad ,$$

where $E_\gamma(f(D))$ is the Mittag-Leffler function and parameters a and b are constrained to be positive. The dose D (a concave function of t) is defined in terms of t and hence it is assumed that the proton beam is delivered to the cells at constant rate $\dot{D}(t)$. The Mittag-Leffler index γ determines the degree of fractionality in the system and approaches the classical Poisson process as $\gamma \to 1$.

7

Testing the Validity of Monte Carlo Simulation Results

In Chapter 4 a multiple stage Monte Carlo simulation is introduced to predict radiation damage and response in cells under specific exposure and biological conditions[1]. The simulation workflow was developed with the aim of accurately predicting multiple endpoints of cell response following exposure to proton radiation of varying mean energy whilst optimising simulation performance. An effective Monte Carlo simulation will always give results comparable with experiment and theory at multiple stages[2].

In this chapter we focus on the results of our simulations of V79 Chinese Hamster cell irradiation using protons by performing comparisons with experiment in its various stages including

- geometry and cellular conditions,
- physical processes undergone by primary and secondary particles within the cells,
- predicting DNA damage and repair
- and cellular survival.

The Geant4 and TOPAS MC toolkits have been extensively tested against experiment [76–78], however for the sake of completeness we ensure that the choice of physics models, clustering methods and post-processing used in the simulation workflow is also comparable with experiment and theory. The parameters chosen in Chapter 4 will be reiterated and shown to agree with ionisation and DNA damage yield, interaction cross section and cell survival data from the literature in addition to cell response theory. The parameter calibration analysis performed when using

[1] For example, the cell line and oxygen concentrations.
[2] A point also highlighted in Chapter 2.

the TLK model from [33, 48] to predict the survival fraction is also discussed in this chapter.

The mean proton energies considered in our simulations are 1.6, 1.7, 1.8, 2.0, 2.1, 2.5, 3.5, 5.0, 20 and 50MeV which corresponds to LETs 35, 32, 29, 24, 22, 16, 11, 8, 2.6 and 1.2keV/μm. The results from high-LET radiation simulations are found to fluctuate more than low-LET radiation and could possibly be attributed to the larger number of particles required to achieve a given absorbed dose at low-LET compared to high-LET which is most likely due to the uncertainty being proportional to $1/\sqrt{N}$. It has also been highlighted in the IRCU guidelines[3] that small volumes[4] exposed to radiation are more likely to experience statistical fluctuations in metrics such as dose and particle LET [150]. The target volumes considered in our simulations are $\sim 9\mu$m in diameter however we still except to see some degree of dose and LET fluctuations.

The simulations were repeated such that the uncertainty in proton LET delivered to the cells is <2% whilst the absorbed dose was not found to fluctuate much in most situations with some exceptions. To control the uncertainties, every measurement taken at a given absorbed dose and proton LET was repeated 30 times at high-LET and reached as low as 5 repetitions at low-LET due to time restrictions.

7.1 Geometry and Cellular Compositions

The initial testing on cell geometry and composition performed in Chapter 4 aimed to find the optimal cell model whilst giving reliable results. The cell line of interest, the V79 Chinese Hamster cell, was chosen due to its consistent use in radiobiological research. Table (7.1) provides an outline of the parameters chosen to represent the cell line in simulation.

7.1.1 Methods

Geometry testing exclusive to V79 cells is not reported extensively in the literature however there are tests performed on cell geometries with arbitrary organelle dimensions. In addition to other cellular damage metrics, the tests involve scoring of ionisation yields and the resultant fraction of absorbed dose in each organelle [151–153].

To test the yield of ionisations in each organelle (cytoplasm, nucleus and nucleolus), a single V79 cell as defined in our simulations (i.e. as per Table (7.1)) is

[3]The International Commission on Radiation Units & Measurements tasked with defining international radiation calculation standards.
[4]In the order of <1μm in diameter.

exposed to proton radiation of varying LET using TOPAS. It is exposed to 10 proton tracks ranging from 1 to 50MeV[5] (0.25MeV FWHM). All Geant4 physics models (i.e. Livermore and Geant4-DNA physics lists) that were chosen in Chapter 4 are also implemented in the simulations.

Organelle	Parameter Name	Parameter
Cytoplasm[a]	Material	Liquid Water
	Density	1.0g/cm^3
	Inner Radius	7.0μm
	Outer Radius	8.5 μm
Nucleus[b]	Material	Liquid Water
	Density	1.0g/cm^3
	Inner Radius	5.0μm
	Outer Radius	7.0μm
Nucleolus[c]	Material	Liquid Water
	Density	1.1g/cm^3
	Radius	5 μm

Table 7.1: V79 Chinese Hamster cell properties used in our simulations.
[a]Green region in Fig.(7.1).
[b]Green → Pink region in Fig.(7.1).
[c]Pink → yellow region in Fig.(7.1).

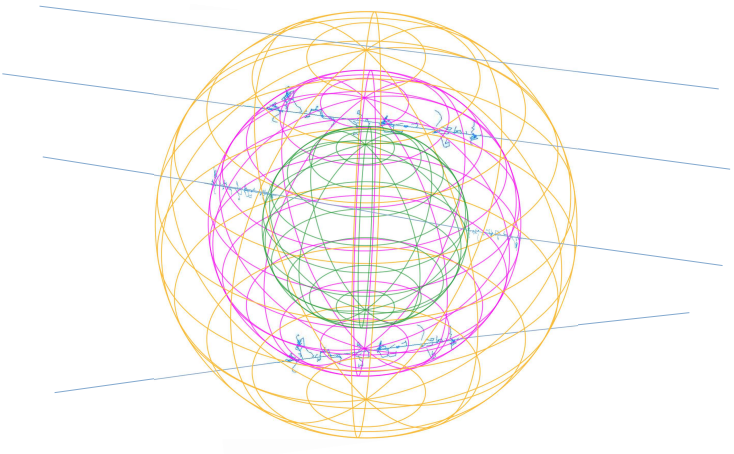


Figure 7.1: A single V79 cell with 4 low energy proton tracks traversing through it.

[5]The range of clinical proton energies considered in our study.

A "pTuple" scorer was implemented in the simulation to record the relevant particle track information including the energy deposited, the name of the interaction and organelle for each event (see Chapters 3 and 4 for a description of a "pTuple" file in TOPAS) allowing the ionisation yield (defined as energy depositions $\geq$10.79eV in our simulations) in the cytoplasm, nucleus and nucleolus can be calculated respectively. This simulation was repeated 30 times at each energy using different random seeds.

7.1.2 Summary of Findings

The results of these simulations are outlined in Figs.(7.2) and (7.3). We expect the ionisation yield to increase with particle LET due to the increasingly dense ionisation track structure. The ionisation yield in each cell organelle is in the range of 10^6 ionisations for high-LET radiation and $\sim 10^5$ for low-LET radiation. This is around the range reported in the literature [152, 153] where ionisation yield and DNA damage is measured in a nucleus of radius 8μm. Inferences can also be made on the cell's probability of survival based on these measurements.

By observing the ionisation yield in the cytoplasm, nucleus and nucleolus as a function of LET, one can see in Fig.(7.2) a steady increase between 0 and 10keV/μm followed by a more drastic increase for high-range LET (>10keV/μm). As discussed above, we expected this behavior due to the densely ionising particle track of high-LET beams thus implying that the DNA damage count will increase with LET, resulting in more efficient cell killing. Despite having the largest volume, the ionisation yield is the lowest in the cytoplasm due to the Livermore physics list being used in place of the Geant4-DNA model as a simulation optimisation method. As discussed in Chapter 3, the Livermore physics model simulates a less detailed track structure compared to the Geant4-DNA physics model, which is more suited to nanoscale simulations necessary for DNA damage counts.

Fig.(7.3) shows that the absorbed dose delivered to the cell per particle decreases exponentially with increasing mean proton energy. From 1 to 10MeV, the absorbed dose per particle decreases rapidly before levelling off for energies >10MeV.

7.2 Implementation of Physical Processes

It would be remiss to consider ionisation yield alone when testing the accuracy of our simulations. We know from Chapter 2 that there are numerous processes undergone by primary and secondary particles propagating through biological media. The relative probabilities of each particle undergoing a given interaction varies with energy and the material through which it is propagating (liquid water in the current investigation). We know that interactions do not occur via collisions alone and a

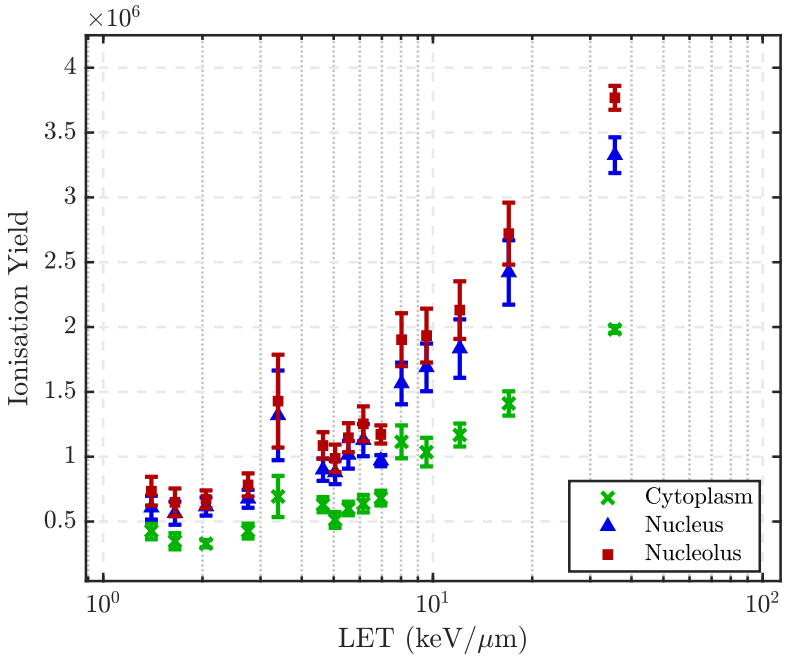


Figure 7.2: Ionisation yield within a single V79 cell as a function of LET from our simulations.

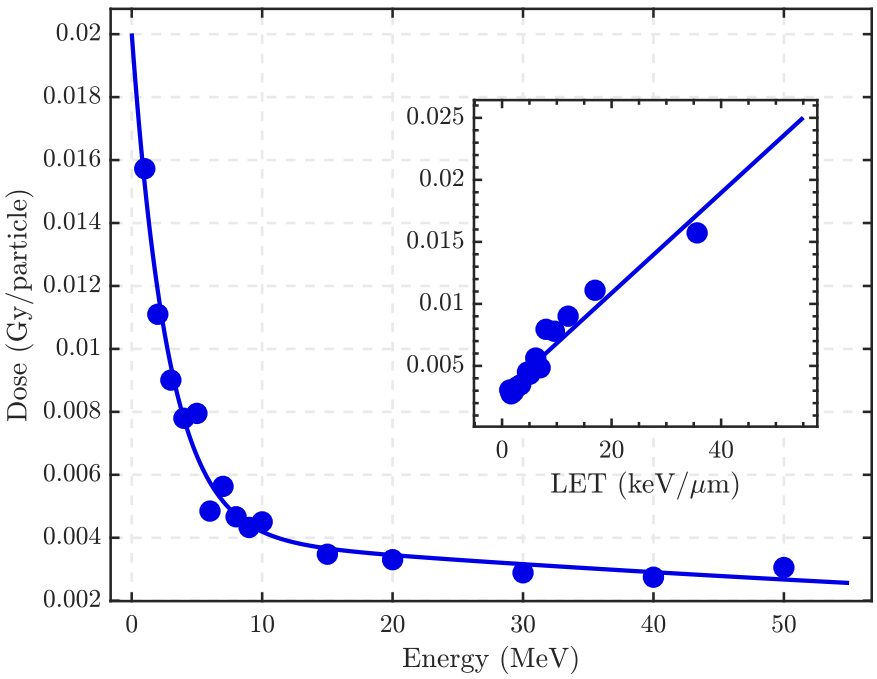


Figure 7.3: Absorbed dose per particle delivered to a single V79 cell as a function of proton energy and LET from our simulations.

particle can interact after propagating some distance, making it difficult to predict the exact fate of a particle. The interaction cross section is used to describe physical interaction yields by providing a range of probabilities that a particle of a given energy will undergo a specific interaction.

The large output file generated from our TOPAS simulations describes the distribution of processes that occur as every particle propagates and loses energy. In this section we observe the relative number of occurrences of each physical process for comparison with experimental and MC simulated interaction cross sections in liquid water available in the literature.

7.2.1 Methods

The simulation output files generated in the TOPAS stage contain every event that occurs inside a cell in the simulation from beginning to end of radiation exposure. Each output file is in the order of 50 to 100GB in size prior to compression depending on the dose delivered and LET. The results required from the output file for the current study are the names of all processes that occur during exposure at least once and the frequency. Given the correlation between absorbed dose and number of primary particles generated, the resultant yield of each process is expected to grow with dose. Therefore, the yield of each process is reported in units of $\text{cell}^{-1}\text{Gy}^{-1}$, leaving the results dependent on particle energy/LET.

Each output file is too large to be imported directly into MATLAB for processing, therefore the Bash Scripting command, "`grep`" (see Appendix C for a description of the command), is used to determine the processes and number of occurrences in each simulation. The experimental cross section data [97, 103, 154] presented below is measured with respect to the kinetic energy of the incident particle. We also provide the cross sections generated from the Geant4 physics lists [155] for comparison. To compare the experimental and Geant4 cross sections with our data we determine the mean yield of each interaction type per Gy of absorbed dose per cell as a function of mean kinetic proton energy. We compare the trend in our interaction yields with what is expected from the corresponding interaction cross section.

To compare our secondary electron interaction yields with experiment we generate histograms of the electron events binned by their kinetic energy allowing us to observe the distribution of electron energies that undergo a specific interaction. We then compare our results with the corresponding interaction cross section - this investigation is performed for electron ionisations and excitations. The experimental cross sections for electron ionisation and excitation interactions were obtained from the literature [156–160] in addition to the Geant4-DNA physics model cross sections from [155].

7.2.2 Summary of Findings

7.2.2.1 Proton Interaction Cross Sections

The total interaction cross sections for proton ionisations, excitations and charge transfer from the literature are shown in Fig.(7.4) as a function of mean proton kinetic energy. The cross section for each interaction is maximised at approximately 10^4-10^5eV. Recalling that the energies considered in this study range between 1.6 and 50MeV, the cross sections for each interaction are slightly smaller which will result in a lower yield.

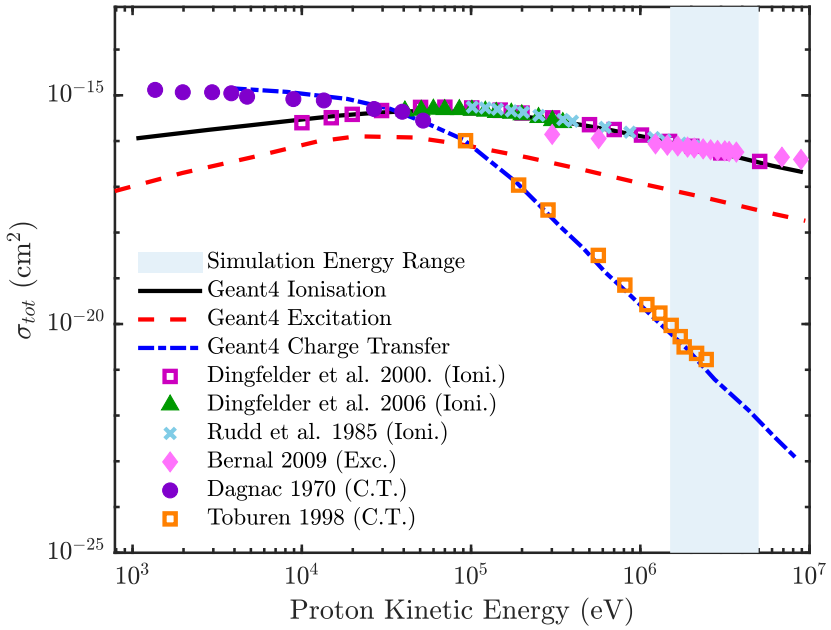


Figure 7.4: Total interaction cross section of proton ionisation, excitation and charge transfer between proton kinetic energies 10 eV to 10 MeV inside of a liquid water medium. The cross sections used in the Geant4-DNA physics model obtained from [155] are shown (red, blue and black lines) with experimental ionisation and charge transfer data for comparison [97, 103, 154, 161, 162]. The shaded area represents the energy range considered in our simulations.

Figure (7.5) depicts the yield ($Gy^{-1}cell^{-1}$) of each process in our simulations as a function of mean kinetic proton energy. By comparing our simulated yields with their corresponding cross sections in Fig.(7.4) we see that our yields are consistent with what is observed in experiment. The shaded region in Fig.(7.4) represents the energy range considered in our simulations. Here the ionisation and excitation cross sections are equidistant throughout the entire range. The same effect is observed in our simulated ionisation and excitation yields from Fig.(7.5). Additionally we note the much smaller cross section of charge transfer interactions in relation to ionisations and excitations which is reflected in our results from the rapidly decreasing

yield with proton kinetic energy.

From theory we expect the ionisation and excitation yield to decrease with increasing proton energy. The inverse relationship between kinetic energy and particle LET suggests that a higher interaction yield is expected at low proton energies. Again this is exactly what occurs in our simulated results.

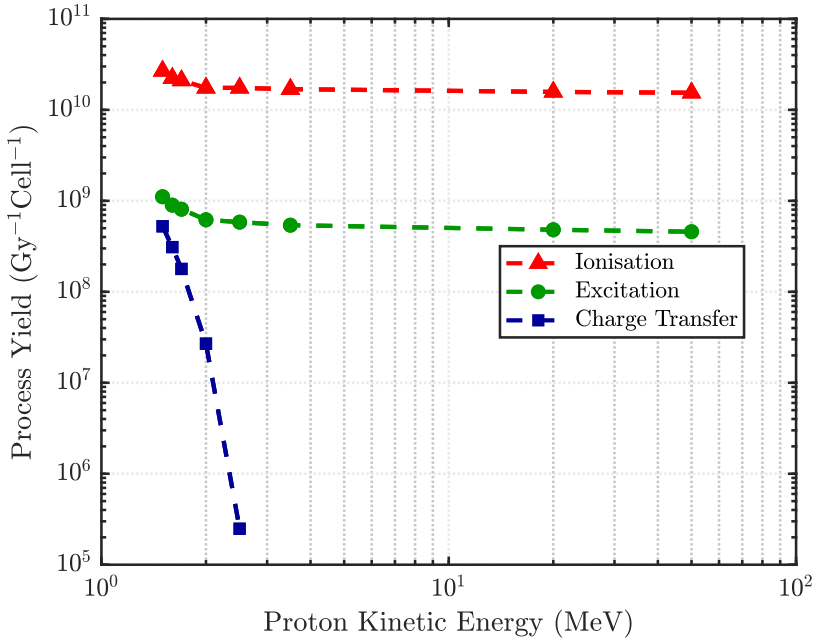


Figure 7.5: The interaction yield Gy^{-1}cell^{-1} for proton interactions as a function of incident proton energy from our simulations. **Note:** The error bars on the data points are very small and hence are not visible. Here the proton kinetic energy range is the same as the blue shaded area of Fig.(7.4).

7.2.2.2 Secondary e$^-$ Interaction Cross Sections

It is known from Chapter 2 that primary protons can induce DNA damage via the production of secondary electrons which can also go on to induce excitations and ionisations. The distribution of electron kinetic energies that undergo each interaction type for three mean kinetic proton energies from our simulations is shown in Fig.(7.6)[6]. These distributions can be compared with the electron ionisation and excitation cross sections as a function of kinetic electron energy shown in Fig.(7.8). The results of [163] depict the secondary electron energy distributions produced from primary protons in PMMA material. The mean energy for each interaction does not vary much between proton kinetic energies which is consistent with our results in Fig.(7.6) where mean of each distribution is in the same region for different mean

[6] 1.6, 2.5 and 50MeV corresponding to an LET of 35, 16 and 1.2keV/μm respectively.

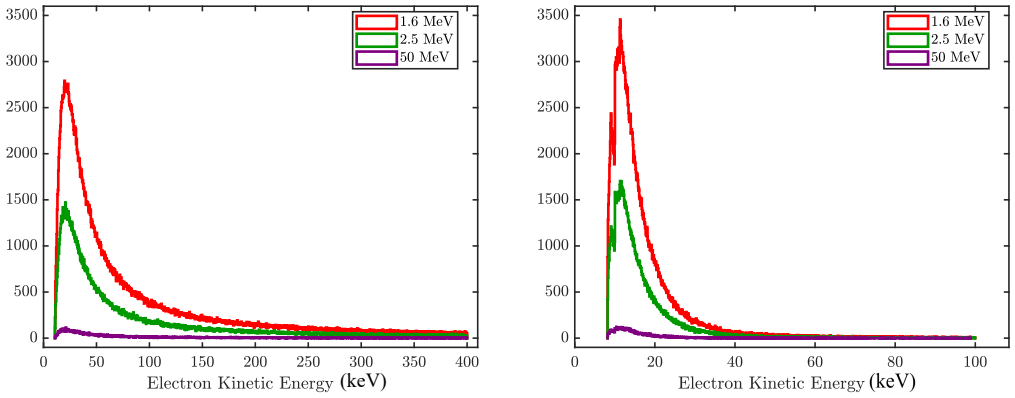
proton kinetic energies.

Figure (7.6(a)) shows that the mean electron kinetic energy at which ionisations occur is in the range <50eV for all three mean kinetic proton energies. In the same energy range of Fig.(7.8), the electron excitation cross section is higher than that of ionisation interactions according to the experimental cross section data of Munoz et al. [160] and simulated cross sections of the NOREC Monte Carlo code [159]. However the Geant4-DNA model suggests an even lower cross section in comparison to [160] and [159]. In Fig.(7.6(a)) the number of excitations are much higher than the ionisation yield for electron kinetic energies ranging from 0 to 20keV meaning that the cross section data is consistent with what is observed in our simulated results because the cross sections suggest that the excitation yield should be higher than that of the ionisations. According to Fig.(7.8), the cross section is at its highest at 20eV and 100eV for electron excitations and ionisations, respectively, which contradicts our observations in Fig.(7.6). This can be explained using Fig.(7.7) which depicts the energy distribution of all electrons involved in interactions which are originally produced from primary proton kinetic energies of 1.6, 2.5 and 50MeV. The yield of electrons that exist with kinetic energy >20eV is very small and hence only a small fraction are available to interact.

Furthermore, we observe in Figs.(7.6(a)) and (7.7) that the production of secondary electrons increases at smaller kinetic proton energies. We expect a higher yield of interactions at smaller proton kinetic energies due to the increasing ionisation density of high-LET radiation and is hence consistent with what we except from radiation theory.

Overall it is difficult to draw conclusions due to the disagreement of cross section data in Fig.(7.8) from the literature, however comparing the relative ionisation and excitation yields for electron kinetic energies < 50MeV shows that more excitations occur in this small energy range compared to ionisations, which is consistent with the cross sections in Fig.(7.8) for the same energy range.

The relative yield of electron excitations and ionisations per Gy per cell as a function of primary proton energy is shown in Fig.(7.9). This plot was generated using the full simulation described in Chapter 4 for mean proton energies between 1.6 and 50MeV. From this plot we can see that the ionisation yield is larger than the excitation yield. The trend shows that the overall yield of each interaction decreases with mean primary proton energy. This can be attributed to the smaller probability of secondary products being produced for high energy protons. The discrepancy between the experimentally measured electron excitation cross section and that used in Geant4-DNA should be highlighted. The Geant4-DNA cross section [155] appears to be underestimated when compared with experiment. This could potentially cause the electron excitation yield to be underestimated in our simulations. However,



(a) e^- Ionisations **(b)** e^- Excitations

Figure 7.6: Distribution of electron kinetic energies that undergo interactions for which the cross sections are depicted in Fig.(7.8) for proton kinetic energies of 1.6, 2.5 and 50MeV.

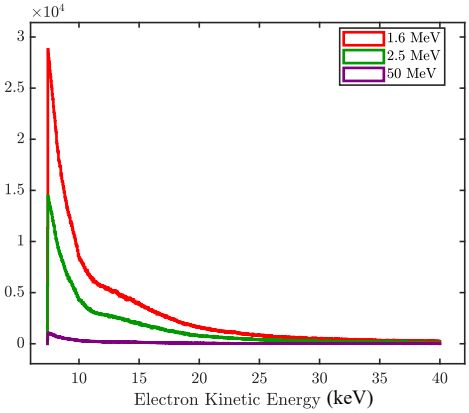


Figure 7.7: Our simulation kinetic energy distribution for all secondary electrons produced in liquid water from kinetic proton energies 1.6, 2.5 and 50MeV.

this should not have a significant impact on the results because ionisations alone are considered to be the primary pathway to DNA damage as a result of ionising radiation.

7.2.2.3 Hydrogen-Proton Interaction Cross Section

The interaction between hydrogen and an incident proton in liquid water is possible, but with a smaller probability than proton and electron interactions, Fig.(7.10(a)). Ionisations formed as a result of proton-hydrogen interactions contribute to cellular damage, therefore it is constructive to discuss in this section. The ionisation yield is an order of magnitude smaller than that of proton ionisations (see Figs.(7.5) and (7.10(b))), however it is still significant enough to influence DNA damage yields. As is the case for proton and electron interactions, ionisations have the largest cross

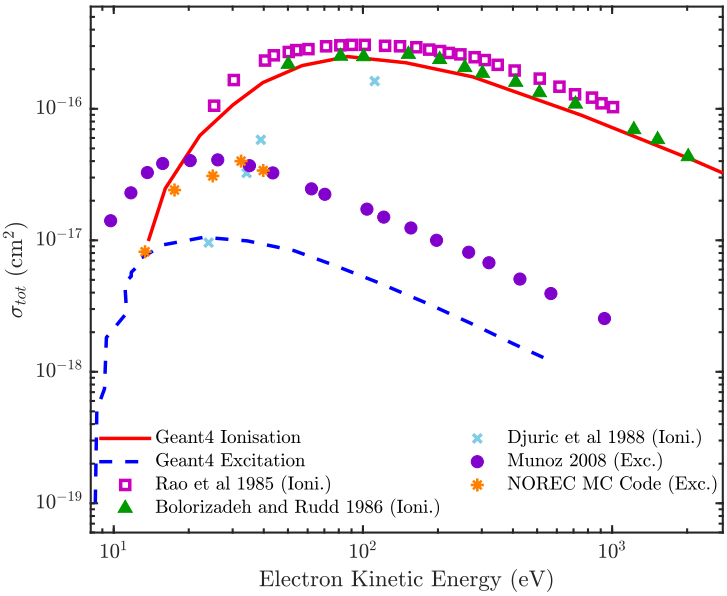


Figure 7.8: The total interaction cross section for electron ionisations and excitations as a function of kinetic energy between 10eV and 10keV in liquid water. The Geant4-DNA cross section is plotted with experimental data for comparison [156–160].

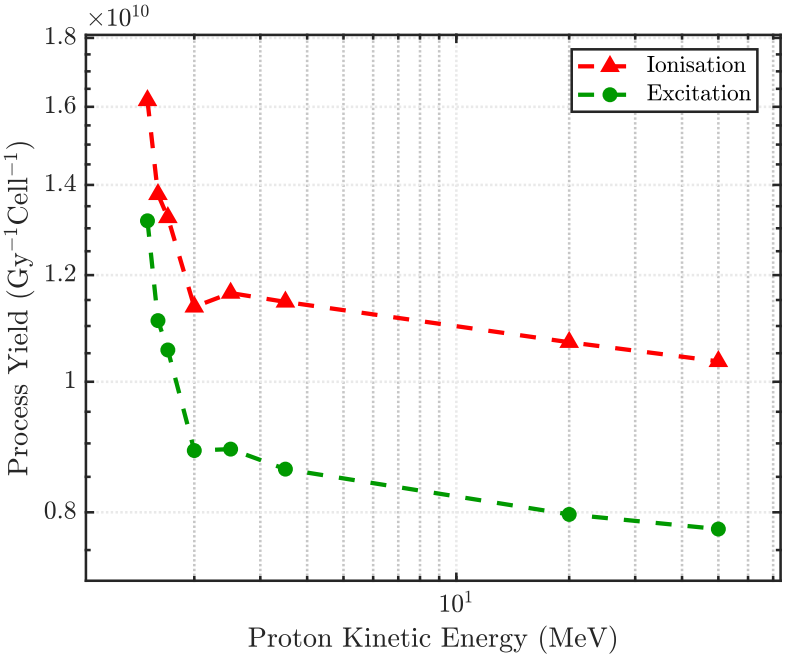
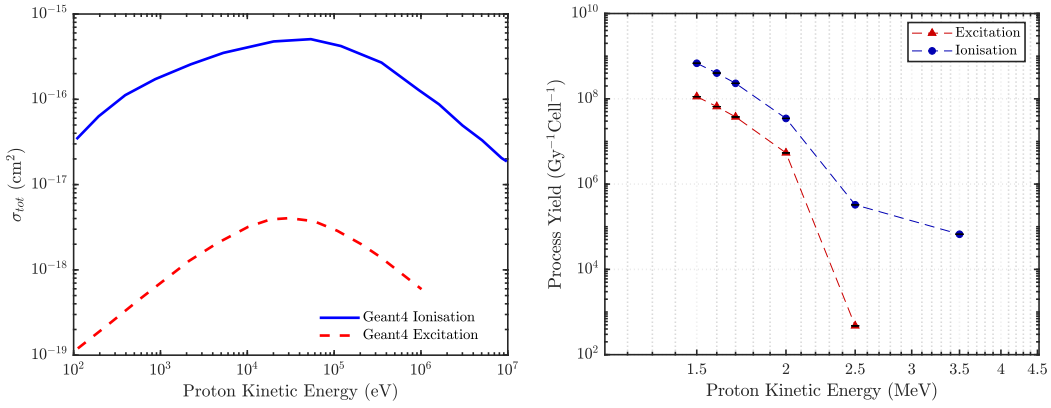


Figure 7.9: The interaction yield per Gy^{-1}cell^{-1} of electron excitation and ionisation as a function of incident proton energy.

section compared with excitation and charge transfer interactions.

The relative yield of excitation and ionisation events are presented in Fig.(7.10(b)). The charge transfer yield was not presented because an extremely small yield was achieved for low proton energies before diminishing completely as the energy was increased. Overall the interaction yields considered appear to be consistent with what was observed in proton and electron interactions. This being that high energy protons are less likely to interact with the medium through which it is propagating due to its larger penetrative range. The ionisation cross sections shows that this interaction will occur with a higher probability than an excitation. This is again consistent with the interaction yields achieved using the current model.



(a) Cross section **(b)** Our simulation yield

Figure 7.10: Hydrogen-proton interaction cross section from [155] (a) and corresponding yield (b) as a function of proton kinetic energy.

7.3 Prognosis of DNA Damage in Cells for Forecasting

Accurate prediction of cell damage induced by ionising radiation is at the forefront of radiobiological research, because this is a major pathway to predicting a given cell's probability of survival. In the literature, MC modelling is used extensively to predict these endpoints. Here we compare the spectrum of cellular damage for proton radiation of different LET predicted by the simulation model developed in Chapter 4 with the results currently available in the literature.

So far in this chapter we have discussed the irradiation stages of the cell response process which is shown in Fig.(7.11), where physical interactions such as ionisations and excitations produce DNA stand breaks inside the sensitive region of the cell (nucleus). Recall from Chapter 6 our physical picture of a cell under irradiation where (i) the cell is in its initial state $|i\rangle$ prior to irradiation, (ii) when irradiation ceases the cell is in an intermediate state $|m\rangle$ where DNA damage has been induced

and (iii) when the cell goes to state $|f\rangle$ by attempting to equilibrate via biological damage repair mechanisms. In our simulation the repair mechanisms result in repair and mis-repair of DNA lesions which contribute to cell death.

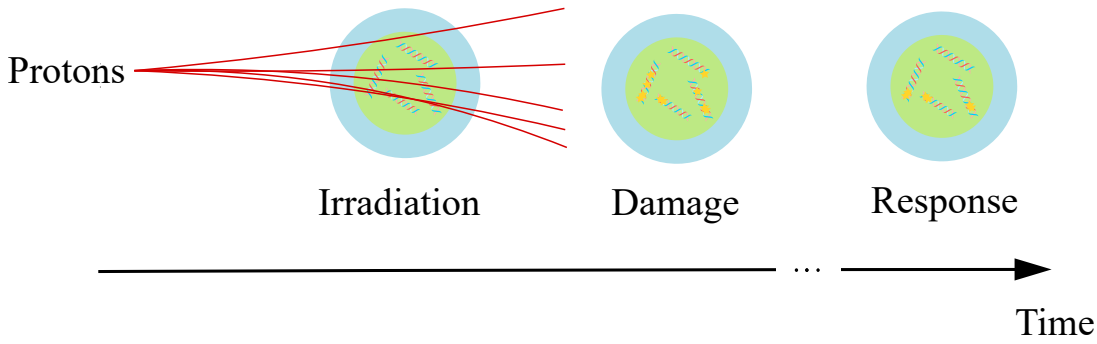


Figure 7.11: The induction of DNA damage within the nucleus of a cell via ionising radiation and its biological response.

Next we will compare the DNA damage yields immediately after irradiation from our simulations with experimental results from [152,164,165], i.e. when the cells are in the end of state $|m\rangle$. We then compare the DNA damage response phase of our simulations to experiment and theory before repair has commenced, i.e. when the system moves from its non-equilibrium to its equilibrium state. DNA repair/mis-repair mechanisms are critical to the prediction of cell survival fraction because the long-term response comes after the repair processes have repaired/mis-repaired the DNA damages and the cells containing lethal lesions have died. The repair processes in our simulations are achieved using an adaption of the TLK model [33] from [48,87] which includes first-order repair (where broken segments from the same DNA strand are repaired) and secondary-order repair (where more complex clusters of damage interact and can lead to mis-repaired and lethal damages).

7.3.1 Methods for Testing DNA Damage Data

The characterisation and prediction of DNA damage in our simulations are explained in greater detail in Chapter 4. To recall, three types of DNA damages are considered in our simulations : (i) single strand breaks (SSB), (ii) double strand breaks (DSB) and (iii) complex double strand breaks (cDSB). In the full model simulations described in Chapter 4 the total yield of each damage type is reported for all of the cells. To compare with the literature a single damage yield of each complexity is reported in units of $\mathrm{Gy}^{-1}\mathrm{cell}^{-1}$ for all energy and depth-varied proton LET simulated in this work as discussed in Section (7.2).

The experimental data that our measurements will be compared with come from a number of studies [152, 164, 165] where DNA damage yields are measured experi-

mentally and via MC methods. These studies report the damage yield in units[7] of $Gy^{-1}Da^{-1}$ at different values of the particle LET. The classification of damages in the above studies consists of several types :

- ionisations inducing no strand break (SB) - the DNA bases are damaged but the strands are still intact,
- SSB - a single break on a DNA strand,
- complex SSB - two breakages within 10 bp on the same strand or on opposite strands at a distance apart of >10bp,
- DSB - two strand breaks within 10 base pairs (bp) of each other,
- and complex DSB (cDSB) - more than two breakages within 10bp.

Due to the DNA not being physically represented in the simulations of this study, the standard 3 types of SBs[8] are considered. These results must be converted from $Gy^{-1}Da^{-1}$ to $Gy^{-1}cell^{-1}$ to make them comparable. The literature values used for comparison are presented in Table (1) of [164] and Table (5) of [151]. To make the values comparable to the data generated using the current study, the genome molecular weight inside of a V79 cell is required[9]. This is determined by taking the product of the genome length of a V79 cell[10] and the molecular weight of a single DNA base pair[11]. This unit conversion makes the experimental data comparable to our results.

$$\boxed{650\text{Da bp}^{-1}\text{cell}^{-1} \times 3.8 \times 10^{9}\text{bp} = 2.47 \times 10^{12}\text{Da cell}^{-1}}$$

It is also worth noting that the experimental damage yields are measured using slightly different methods compared to that adopted in this study. For the sake of completeness we compare our simulation results with what is expected from theories of radiation induced cell death. Therefore the data collected across all energy- and depth-varied LET will be compared to ensure the behavior is expected based on the concepts of cellular response discussed in Chapter 2.

7.3.2 Summary of Findings: DNA Damage

The mean SB yields measured as a function of energy-varied and depth-varied LET in the current thesis are presented in Tables (7.2) and (7.3), respectively with their energy/depth and corresponding LET. The uncertainties reported are the standard

[7] Da stands for Dalton. 1Da = 1.66×10^{-27}kg.
[8] SSB, sDSB and cDSB.
[9] The molecular weight of DNA present inside of a single V79 cell.
[10] In the current study this is assumed to be a mean value of 3.8Gbp.
[11] Equal to 650 Da.

error of the mean, $\sigma/\sqrt{N}$ where σ is the standard deviation of the yield measurements and N is the number of samples taken.

Energy (MeV)	LET (keV/μm)	SSB ($Gy^{-1}Cell^{-1}$)		Simple DSB ($Gy^{-1}Cell^{-1}$)		Complex DSB ($Gy^{-1}Cell^{-1}$)	
		Mean	SEM[a]	Mean	SEM[a]	Mean	SEM[a]
50	1.24	169.6	1.6	29.5	5.8	11.7	3.9
20	2.6	160.6	0.8	29.8	0.1	11.2	0.1
5	8.26	143.3	7.8	38.8	2.5	16.7	1.3
3.5	11.39	130.7	3	41.3	0.5	19.6	0.8
2.5	16.45	113.1	20	43.3	9.3	25.2	8.7
2.1	21.58	94.7	10	40.9	4.3	29.6	16.6
2	23.67	88.8	0.7	40.1	0.3	30.5	0.2
1.8	28.1	80.4	9.5	39.2	4.7	36.5	4.6
1.7	31.9	73.1	3	36.9	2.4	38.5	3.2
1.6	35.04	69.4	7.7	36.4	3.9	43	4.4

Table 7.2: The yield of DNA damages with energy-varied LET.
[a]Standard Error of the Mean.

Depth (mm)	LET (keV/μm)	SSB ($Gy^{-1}Cell^{-1}$)		Simple DSB ($Gy^{-1}Cell^{-1}$)		Complex DSB ($Gy^{-1}Cell^{-1}$)	
		Mean	SEM[a]	Mean	SEM[a]	Mean	SEM[a]
1.2	3.01	160.7	0.3	31.1	0.1	11.7	0.03
2.6	3.97	147.1	0.7	35	0.2	14.9	0.1
3.9	8.29	124.3	0.5	41.2	0.2	21.4	0.1
4.2	20.58	76.1	1.0	35.1	0.5	34.4	0.4
4.1	13.87	98.1	0.7	39.3	0.3	27.7	0.2
4.3	25.39	58.6	0.9	29.5	0.5	37.4	2.3
4.4	28.12	51.0	3.8	26.8	0.2	41.3	11.6

Table 7.3: The yield of DNA damage with depth-varied LET in a proton beam of mean energy 20 MeV in the current simulation study.
[a]Standard Error of the Mean.

Figure (7.12) depicts the yield of each damage type as a function of proton LET. The DSB yields from experimental studies [152, 164, 165] are also added to the plot for comparison. Here we note that each study characterises DNA damage complexity differently. In [164] DSB damages are characterised similar to our study, however "distant"/complex DSBs are defined as clustered damages with in 10kbp of each other, whilst SSBs are not discussed. In our study a complex DSB is defined as a cluster of damages that exist within 20bp of each other. Therefore the definition of a complex DSB in [164] involves a much larger distance than is considered in our simulations. In [152] the definition of an SSB is equivalent to our definition however the

yields provided in Table 2 of [152] account for all breaks (including DSBs), thus we will not consider this in our comparison. Finally [152] classifies DSB^{+} and DSB^{++} as damage clusters at a distance of >10bp apart. Similar to [164] the complexities DSB^{+} and DSB^{++} are equivalent to a complex DSB in the current study but over a larger distance. The differences between the classification of damage across different studies may affect the agreement between the literature and our results. A comparison of DNA damage classifications across different studies is summarised in Table (7.4).

The Current Study	**Nikjoo et al. 2001 [152]**	**Friedland et al. 2002 [164]**
SSB	SSB	-
sDSB	DSB,SSB^{+}	DSB
cDSB	DSB^{+}, DSB^{++}	"Distant" DSB

Table 7.4: DNA damage definitions in the current study and [152, 164]. **Note:** SSB^{+} = two damages on the same DNA strand, DSB^{+},DSB^{++} = more than two damages at a distance >10bp and "distant" DSB = damage clusters of more than 10kbp apart.

The general conclusion drawn here is that the DSB and cDSB yield of our simulated results are in the experimental range. We can see that in our simulated LET range the DSB yield appears to range between 10 and 50 damages per Gy of absorbed dose per V79 cell. The steady increase in DSB complexity with particle LET explains why high-LET radiation demonstrates a higher cell killing efficiency than low-LET radiation. The SSB yield appears to decrease rapidly with increasing LET which is expected because damages are more likely to exist in closer spatial proximity as LET increases resulting in fewer SSBs. The rate of change in the DSB and cDSB yield is flatter in the low-LET region before increasing in the high-LET region[12], both of which are in agreement with the experimental data.

As discussed in Chapter 2 the damages induced by low-LET radiation (such as photon and very low-LET proton radiation) are more often indirect[13] in nature. The chemical stage of cellular irradiation was not implemented into our simulations due to time restrictions, therefore it is likely that the damage yield in this region has been underestimated in our results. Regardless of this our yields appear to follow the trend within a reasonable range based on the literature.

The DSB and cDSB yields from our results contain a "crossover" region where the complex DSB damage yield becomes larger than that of simple DSB damages

[12]For LET greater than 10keV/μm

[13]Damage as a result of events occurring within the vicinity of a DNA segment.

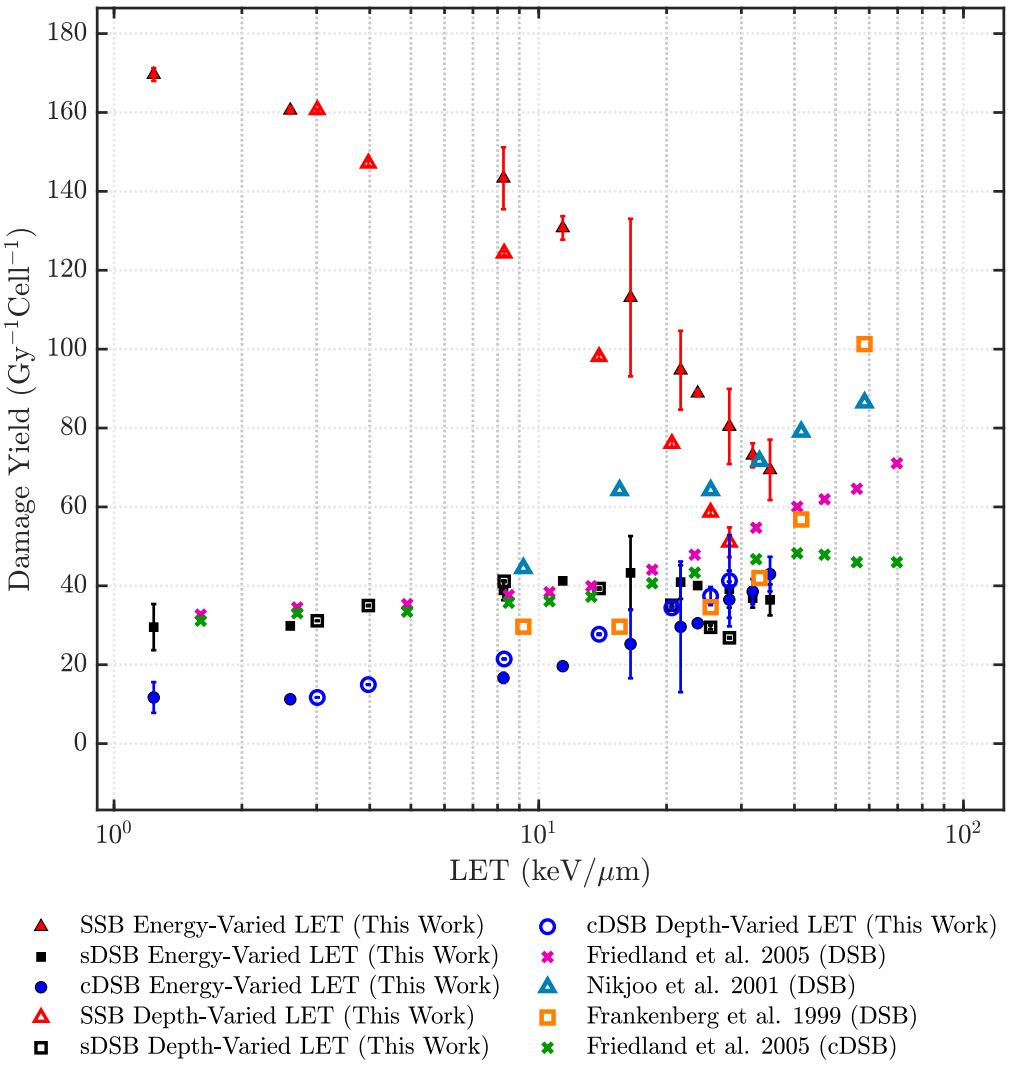


Figure 7.12: The yield of each damage type per Gy of absorbed dose per cell as a function of LET from the current study in comparison with the experimental results of [152, 164, 165].

due to the saturation effects of high-LET radiation. The experimental data consistently increases for all LET with the exception of Friedland et al. [164] where the cDSB yield flattens out. The simple DSB yield from our simulated results does not reflect the trend observed in the experimental data for a similar damage complexity, however there have been similar trends observed in other Monte Carlo studies from the literature [49, 166]. On the other hand, the simulated cDSB yield continues to increase with LET, which is more consistent with the trend between damage yield and LET observed in the experimental results. The complex DSB yield from [164] shows similar behavior to our simple DSB yield in that it begins to decrease for LET >20keV/μm and strangely the cDSB yield is almost equal to the DSB yield. The reason for the discrepancies between our results and experiment could be due to different classification of damage complexities as outlined in Table (7.4) or different strand break induction parameters such as the ionisation energy (in our simulations an energy deposition of 10.79eV is required to produce a strand break). Furthermore, the measurements made in the results derived from experiment [165] are performed with the full physical picture where some DNA damage has likely repaired before counting. The discrepancy between Monte Carlo simulated and experimental DSB yields is a topic still under investigation in the literature [164].

Figure (7.12) also demonstrates consistencies in the damage yield when LET is varied by both depth and energy. The simple and complex DSB yield are very consistent between the two methods, while the SSB yield appears to be slightly larger in the high-LET region when varied by energy. Quantitatively this is in the order of ∼20 SSB larger than the yield measured with depth-varied LET.

The variance of yields in the three studies from the literature is quite large, however this can be attributed to the different characterisation of damage types and methods with which the measurements were made. For the purpose of this investigation the reported simple DSB yields are assumed to account for all types discussed in each of the literature studies, i.e. DSB^{+} and DSB^{++}. Considering this, the yields achieved by the model used in the current study are in good agreement with the experimental data. The experimental spectrum of complex damages occur within the same range as our results.

Figure (7.13) depicts the relative yields of each damage type using the current study with a comparison to similar results from [152] for SSBs and cDSBs which are defined as per Table (7.4). Recalling that in [152] the complex DSB is classified as a damage cluster over a distance of >10bp. In our study we do not classify damages of more than 20bp apart to be part of the same cluster. This is a possible explanation for the smaller percentage of complex damages in [152], because a single cluster is classified over a larger number of base pairs than we do here. The expected percentage of damages classified as an SSB is in good agreement with our results. The most agreement is observed in the low-LET region with a discrepancy of ∼10% in

the high-LET region. Furthermore, [33] predicts that 20% of SB damages are considered to be complex for low-LET radiation. Our results are also in good agreement with this prediction.

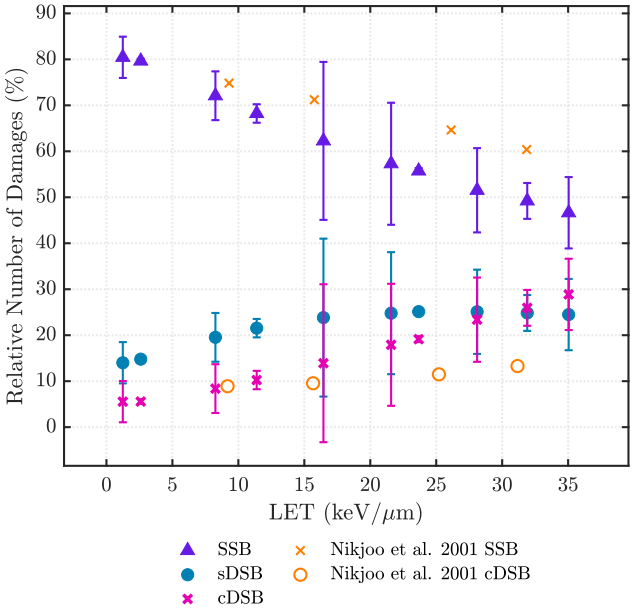


Figure 7.13: The percentage of each damage type produced as a function of LET for our simulated results compared with that of SSB and cDSB from [152].

Similar to Fig.(7.12), a "crossover" region is present in Fig.(7.13) where the proportion of damages classified as a cDSB exceeds that of DSB. We have not observed this effect in the literature possibly because, as discussed above, damage complexities are classified differently across studies. If we consider this in terms of radiation theory we should expect a "crossover" region to occur at high-LET, where the increased spatial proximity of damage results in more damage clusters with ≥3 SBs than with 2.

Figure (7.14) depicts the relationship between damage complexity and absorbed dose for radiation of four constant values of LET from our simulations. Firstly, the damage yield of each complexity type increases linearly with dose,

$$SB_{yield} = m_{yield} D \quad , \tag{7.1}$$

where m_{yield} is the slope of the line and D is the absorbed dose. This same property is discussed in the literature [57, 152, 165]. The slope of the line is clearly correlated with complexity and LET where sparsely-ionising proton beams produce the highest number of SSB damages before slowly decreasing with increasing LET. The highest

yield and hence the largest slope of complex DSB yields unsurprisingly corresponds to the highest LET of ∼35keV/μm. The yield of simple DSB is where things get interesting. The yield slope is extremely close for the three highest LET. This is because the "crossover" region from Fig.(7.12) occurs between 8 and 28keV/μm and for high-LET values cDSB becomes the dominant damage type of the two.

7.3.3 The Role of Repair Time in our Dose-Response Data

Until now this section has focused on the multi-cellular system at the end of the intermediate state $|m\rangle$, immediately after irradiation has ceased and a number of radiation induced DNA damages are present. Here we discuss the processes involved when the system moves from state $|m\rangle$ to the final state $|f\rangle$ where the damages repair/mis-repair/do not repair to form lethal lesions and ultimately lead to cell death (see Fig.(7.11)) in our simulations. Here we note that in our computer simulations cell response progresses on a time scale that is slightly different from what happens in experiment (real life), as shown in Fig.(7.15). In experiment the repair processes are already active during the irradiation stage, this in return causes disagreement in the counts of simple DSB when they are compared with the literature immediately after irradiation has stopped. This further explains the quantitative discrepancy between experiment and our simulations. When a cell is exposed to radiation, the biological response takes place attempting to repair the damage soon after it is induced, whereas our simulations consider the delivery and repair time to be two separate time scales, t_{del} and t_{rep}, respectively. If in our simulation we wish to irradiate for an arbitrary time t, the beam delivery time t_{del} will reach its maximum and simultaneously the repair time will commence such that $t = t_{del}$ and $t_{del} = 0$.

The amount of repair time given to the system to equilibrate can strongly influence the survival probability.

Figure (7.16) depicts the kinetics of cell response for decreasing LET including the sDSB and cDSB yields as a function of repair time (middle column), survival fraction as a function of repair time (right column) and dose (left column) for our simulation. In experiment it is reasonable to assume that the cells are given ample time to repair before the survival fraction is measured. Depending on the complexity of the damage induced, the repair processes result in a number of repaired, mis-repaired or un-repaired damages where the latter is considered a lethal lesion. As the cell progresses through the four phases of the cell cycle, it will detect the presence of a lethal lesion via the p53 protein or the "tumour suppressor" (see Chapter 2). It will actively prevent a cell from proliferating (undergoing mitosis) when lethal lesions are present by forcing it to undergo apoptosis (programmed cell death). The result is that a cell containing lethal lesions will most likely die before it can divide - thanks to the p53 protein. The response of the V79 cell line to ionising radiation

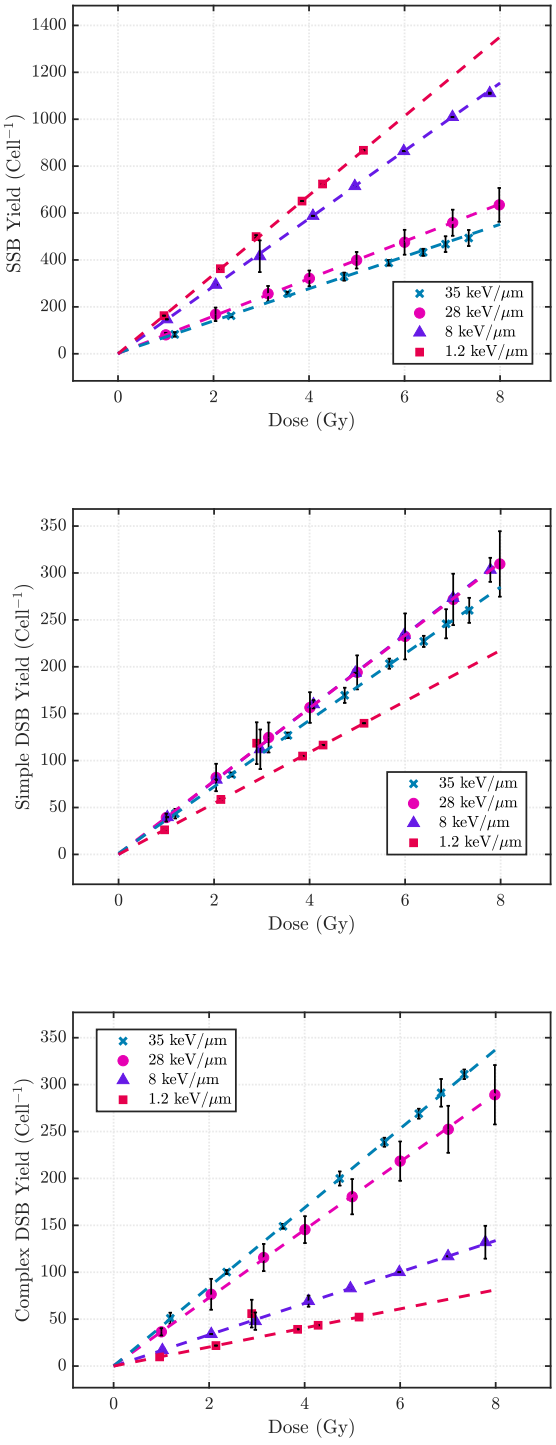


Figure 7.14: DNA damage yield as a function of absorbed dose for increasing proton LET, each with a weighted linear regression fit of slope m_{yield} and y-intercept of zero (it is assumed there is no damage prior to irradiation).

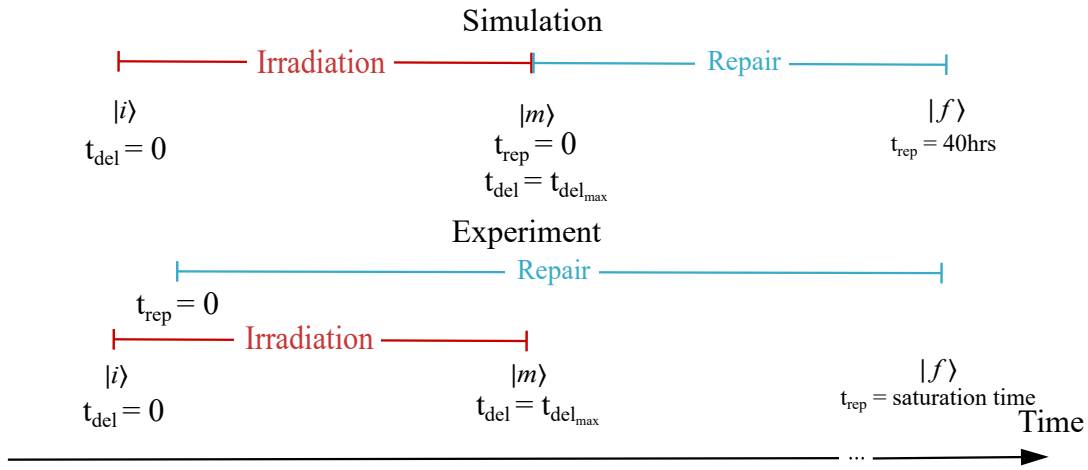


Figure 7.15: The timescale of radiation induced cell death in experiment (real life) and our simulations.

is not of particular relevance to the V79 cell line due to its lack of functionality at repairing DNA damage [167], nevertheless it is worth discussing because it plays a critical role in DNA damage repair of human cell lines.

Cell response theory suggests that a lesion induced inside a cell can repair quickly or slowly depending on its complexity [33]. We know that as LET increases independent damages are in closer spatial proximity where they interact, resulting in more complex damages. Some simple damages are able to repair in the first 15 minutes [33] while some take much longer. Recall that the repair time is a separate time scale from irradiation (t_{del}) in our simulations (which is not the case in experiment). Therefore at $t_{rep} = 0$ all of the damage induced during the irradiation process are present. Then as t_{rep} increases the low complexity damages will repair very quickly as depicted in Figs.(7.16(b)), (7.16(e)) and (7.16(h)) (dashed lines) whilst the complex damages take longer to repair (solid lines). This is just a matter of where to set the repair time counter to zero. In our simulations the counter is set to zero when the irradiation stage has ceased ($L_f = 0$). As the repair process progresses, the amount of lethal damage increases as a result of the repair process. Whereas in real life when the counter is set to zero, all the sDSB and cDSB damages are considered lethal and as repair progresses some of the lethal damages are repaired resulting in a decrease from the initial yield. Regardless of how one performs the counting, the end result is unchanged.

Figs.(7.16(b)), (7.16(e)) and (7.16(h)) describe the sDSB and cDSB yield as a function of repair time, i.e. the repair stage, in our simulations for different doses. The repair half-time for the V79 cell in particular can vary a lot in the literature and is dependent on the dose-rate at which the beam is delivered to the cells and LET [33, 112] because high complexity DNA damages take longer to repair. Generally, the sDSB repair half-time for the V79 cell is ∼4 minutes and the cDSB

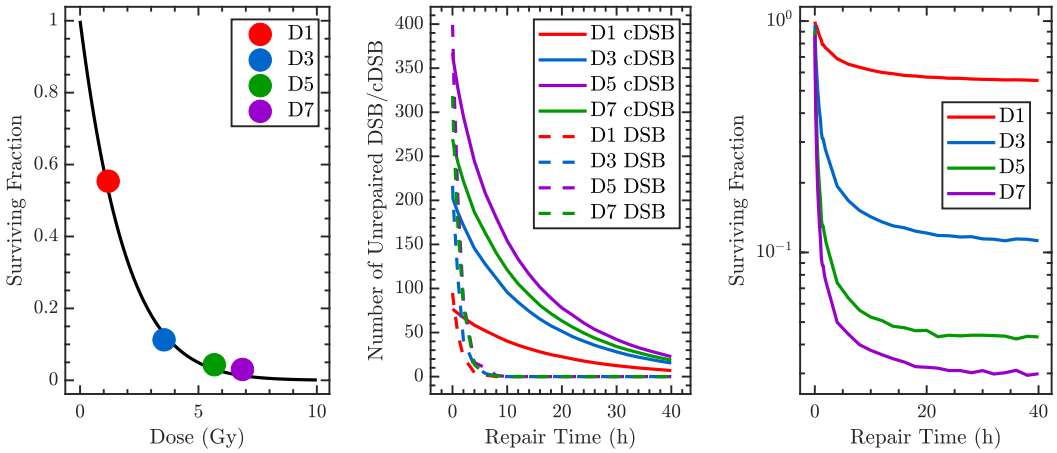
repair half-time is ∼12 hours. Simple stand breaks repair much faster than sDSB and cDSB due to the availability of many repair processes to repair simple lesions such as HR or NHEJ [11] and for this reason do not contribute to cell death. In Figs.(7.16(b)), (7.16(e)) and (7.16(h)) we show the repair rate of sDSB (dashed) and cDSB (solid) damages for increasing doses of high-LET (35keV/μm), mid-LET (16keV/μm) and low-LET (1.2keV/μm) radiation, respectively. In each of the cases considered the sDSBs are repaired within the first 10 minutes. This is consistent with the expected repair half-time of ∼4 minutes and variations from this value can be attributed to different dose-rates, cell environment conditions and LET. We observe that the total yield of sDSB increases with LET as expected, hence the sDSBs inflicted by high-LET radiation will take longer to repair. The cDSBs in the system take a longer time to repair which is reflected in Figs.(7.16(b)), (7.16(e)) and (7.16(h)) (solid lines). All the cDSB yields in the figure reach half their initial yield between 12 and 20 hours which is also consistent with the experimental V79 cDSB repair half time of ∼15 hours.

The closer spatial proximity of damages due to high-LET radiation also results in a higher yield of initial simple and complex DSBs as was discussed in the previous section and is consistent with experiment. Furthermore the complex DSB yield increases with LET resulting in a lower yield of simple DSBs which is also depicted in Figs.(7.16(b)), (7.16(e)) and (7.16(h)) by observing the relative initial simple and complex DSB yield at $t_{rep} = 0$ for each LET.

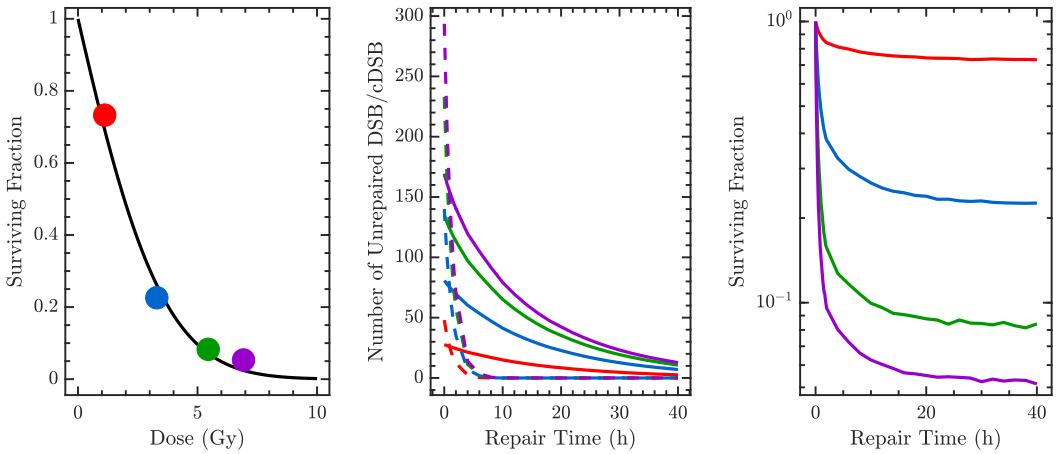
In the current study we have assumed a V79 cell cycle phase transition time of 15 hours and an asynchronous distribution of phase[14]. Taking these initial conditions into account, it is reasonable to assume that each cell will have at least progressed through all four phases at $t_{rep} \approx 16$ hours. When the final equilibrium state $|f\rangle$ is reached (given a sufficiently long repair time) there are no longer damages or mutations present in the cells that require repair and hence will not continue to die as a result of radiation, they are either classified as “dead” or “alive”. Depicted in Figs.(7.16(c)), (7.16(f)) and (7.16(i)) are the survival fractions as a function of repair time for increasing LET from our simulated results. Here we observe the survival fraction to at least begin approaching equilibrium after a repair time of 16 hours - just after each cell is assumed to have progressed through all four cell cycle phases.

The survival fraction in Figs.(7.16(a)), (7.16(d)) and (7.16(g)) show that survival decreases with increasing LET, as is expected with the more complex track ionisation density described in Chapter 2 and an increasing cDSB yield across Figs.(7.16(b)), (7.16(e)) and (7.16(h)). The dose-response curves shown in Fig.(7.16) are recorded after 40 hours of repair time. When the measurements are made at smaller repair times, the survival fraction is higher due to the classification of lethal lesions within

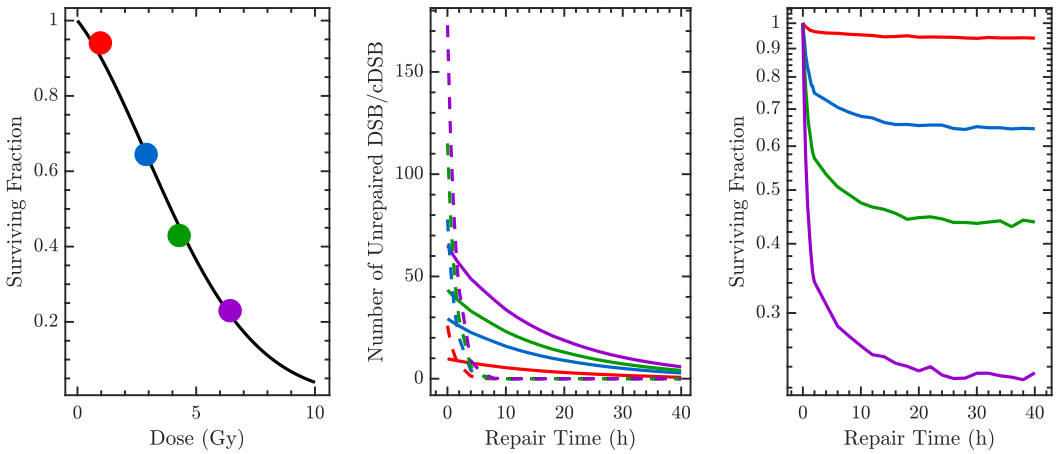
[14]The cells are in different phases of the cell cycle at the time of irradiation.



(a) 35 keV/μm (t_{rep} = 40hrs) **(b)** 35 keV/μm **(c)** 35 keV/μm



(d) 16 keV/μm (t_{rep} = 40hrs) **(e)** 16 keV/μm **(f)** 16 keV/μm



(g) 1.2 keV/μm (t_{rep} = 40hrs) **(h)** 1.2 keV/μm **(i)** 1.2 keV/μm

Figure 7.16: Cell survival kinetics from our simulations as a function of dose. The increasing doses D1≈1Gy, D3≈3Gy, D5≈5Gy and D7≈7Gy were chosen from our simulated data to show the mean survival fraction with varying dose and repair time for decreasing LET (top to bottom).

the framework of our simulation and the Two-Lesion Kinetic (TLK) Model described above. We will discuss the survival fraction more in the next section.

A visualisation of cell behavior after exposure to 1Gy of radiation with increasing LET is provided in Fig.(7.17) using our simulated results and further reflects the behavior of the system outlined in Fig.(7.16). We observe the density of cells that are alive after a given repair time decreases due to the accumulation of lethal damages and resultant decrease in the survival fraction. We observe a more defined decrease in the density of survived cells for higher LET radiation due to the decrease in cDSB yields for the first dose point (D1) in Figs.(7.16(b)), (7.16(e)) and (7.16(h)), respectively.

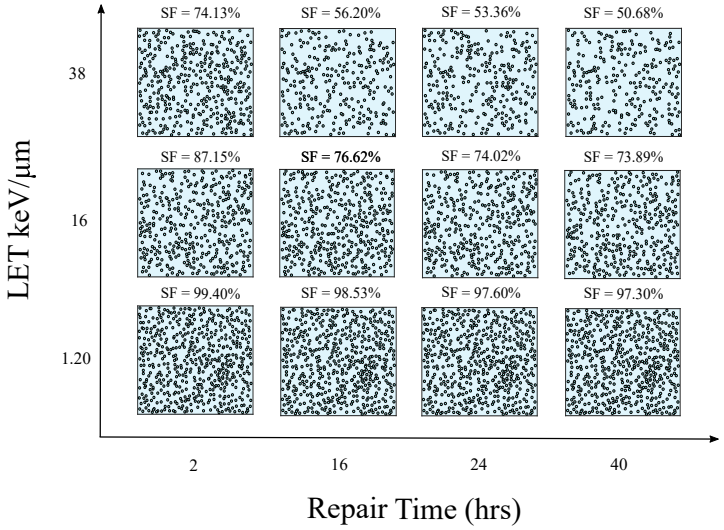


Figure 7.17: Cells that are alive with increasing repair time and LET. The cell coordinates were generated using our simulated results.

Additionally, more time (>15hrs) is required for the survival fraction to approach equilibrium at high doses which is expected because there are more damages present that are more complex, i.e. slowly repairing. Then as the cell progresses through each cell cycle phase, it encounters damage and mutation checks where they are more likely to die as a result.

Consider a thought experiment where a cell has undergone substantial damage after irradiation. At the time that radiation ceases, the cell is in the S phase where it is the most radiation resistant. Simultaneously a second cell sustains a similar amount of damage but is in the radiation sensitive M phase. Despite having a similar degree of damage the probability of the first cell dying is lower than the second cell because in the S phase the damages are more likely to be repaired. This introduces a degree of randomness into cell response and survival fraction which is implemented into our simulations using a random number generator to determine if a given cell

survives (see Chapter 4).

Within the TLK model (Chapter 4) the repair processes are simulated using two differential equations, Eqs.(4.11) and (4.12) describing the yield of simple and complex DSBs as a function of time. A third equation Eq.(4.17) determines the number of damages that are lethal at repair time t_{rep} using the simple and complex DSBs present at the same time. The number of lethal damages present will contribute to cell death at t_{rep}. The results show the yield of simple and complex DSBs decreasing with repair time, as depicted in Figs.(7.16(b)), (7.16(e)) and (7.16(h)) which is consistent with experiment because the damages will continue to repair after irradiation has ceased. The sDSB and cDSB repair half times of the V79 cell line are ∼4 minutes and ∼12 hours, respectively, depending on the dose-rate at which the beam is delivered, the cell environment conditions and LET. As previously discussed, the repair half-times of the V79 cell agreement with the literature because half the sDSB and cDSB lesions have repaired within 5 minutes and 15 hours, respectively. In our simulations, the full picture is not used due to the large computational power and processing times required. However, the agreement of our data with experiment, where the full physical picture is present, gives us reasonable grounds on which to test our stochastic model in Chapter 8

As more damages are repaired in t_{rep}, the TLK model classifies a fraction of the remaining unrepaired damages as lethal which accumulates with time as depicted in Fig.(7.18). In experiment one would expect the lethal lesion count to decrease with repair time t_{rep} and here the TLK is inconsistent with experiment because the lethal lesion yield increases with t_{rep}. The increase of lethal damage in the TLK model results in a decrease of the survival fraction with t_{rep} for our simulated results as expected (see Figs.(7.16(c)), (7.16(f)) and (7.16(i))). In experiment we expect the survival fraction to increase as the yield of potentially lethal damage decreases with repair time and hence is not consistent with our results using the TLK model.

To summarise, the cell response phase of our simulations give results that are consistent with experiment and cell response theory. The repair processes implemented in our simulations via the TLK model allow us to observe the long-term behavior of each cell after exposure to radiation which is important in testing our model in Chapter 8. Within the confines of the TLK model the survival fraction decreases with time due to the lethal lesion count accumulating as more simple and complex DSB clusters are classified as such with repair time. In experiment this does not occur because a lesion can be classified as lethal regardless of whether it is given time to repair or not whereas in the TLK model some repair time must be allowed before a lesion is classified as lethal.

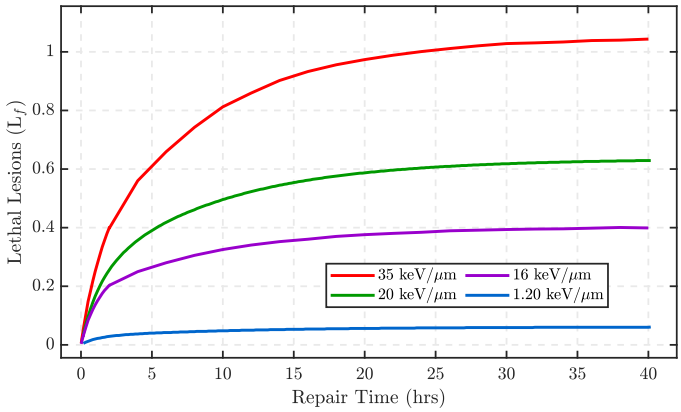


Figure 7.18: Lethal damage yield (L_f) as a function of repair time t_{rep} after exposure to ~1Gy for increasing LET from our simulated results.

7.4 Cell Survival

In this simulation the cell survival fraction is calculated using the Two-Lesion Kinetic (TLK) model developed by Stewart [33] and is described in Chapter 4. This model determines the survival probability of each cell at a given detail in repair time and reduced to contain 5 parameters (the most general form of the TLK model contains 16) that describe the radiation response of the cell. In this section a TLK parameter sensitivity analysis is performed for a high, mid-range and low-LET cases of our simulated results. After performing this calibration we can determine the optimal set of parameters used to generate the simulated dose-response curves for this study. Therefore, the results are calibrated to particular experimental results measured under similar conditions to our simulation study.

In the previous section we discussed our cell survival measurements as a function of repair time, absorbed dose and LET where we concluded that our results are consistent with what is observed in experiment aside from repair time where the trend is a result of lethal lesion accumulation in the TLK model. To further establish the validity of our cell survival measurements, we compare our simulated dose response curves against the world experimental data outlined in Appendix D. It is worth noting here that cell response is not always the same under similar biological and exposure conditions. As discussed in Chapter 2 there are many parameters (physical, chemical and biological) that influence cellular response to ionising radiation and often not all parameters are reported in the literature. This makes it difficult to directly compare results across different experiments because it is likely some parameters differ such as the cell oxygen content or cell cycle phase. Therefore we expect to see some deviation between our simulated survival measurements and experiment.

7.4.1 The Two-Lesion Kinetic (TLK) Model

In Section (4.4.1) we introduced the TLK model [33] and described how it will be implemented in our simulations. To recall, the TLK model describes cell response to ionising radiation using a system of four differential equations which describe the yield of simple and complex DSBs, mis-repaired and lethal lesions with respect to repair time. It involves several fitting parameters that are specific to the cell line of interest which is the V79 Chinese Hamster cell in our simulations. These equations, Eqs.(4.11)-(4.17) are numerically solved for increasing time increments to solve L_1, L_2 and finally L_f, the number of lethal lesions. The model assumes the Poisson process by computing the survival probability using $\exp(-L_f)$ where L_f is the lethal lesions yield at a given repair time. However in our simulations we also use a random number to determine whether each cell "survives" or "dies".

The TLK model parameter sensitivity analysis is undertaken by considering results using the TLK model at LET 35, 16.45 and 2.60keV/μm. For each data set, the TLK parameters η, β_1, β_2, λ_1 and λ_2 are changed one at a time within the allowed bounds defined by [33]. A dose-response curve is generated for a extreme, intermediate and low parameter values for each case[15]. The optimal combination of parameters that demonstrate the best agreement with the experimental data will be used in the final cell survival measurements for all of the simulations performed.

7.4.1.1 TLK Model Parameter Sensitivity Study

The method of TLK model calibration discussed in [33] involves minimising the χ^2 statistic when compared with the experimental data to calibrate the model parameters. In our study, to calibrate the model parameters we use the experimental data for V79 cells in [26,29,96,113] and unfortunately the papers do not contain information of the exact time at which the survival measurements were made which makes it difficult for us to decide to which repair time the model should be calibrated. Instead, we adopted a similar approach to [48] whereby each parameter is varied to observe how it influences the survival fraction. Their results were calibrated with a post-irradiation time of 16 hours due to the 15 hour cell cycle progression time of V79 cells, therefore this is the time at which the system should at least be approaching stability and all lethally damaged cells have died.

We selected a range for each parameter and varied them whilst keeping all other parameters constant. The range of each parameter is selected so that it is within the allowed upper and lower bounds defined in Section (4.4.1) and [33]. Recalling that the survival fraction changes with repair time, we calibrated our model at 40 hours as our system reaches equilibrium.

[15] 5 parameters $\times$ 3 data sets $\times$ 3 values = 45 survival curves to calibrate against experiment.

For the specific LET values selected, our simulated survival fraction data are compared with the experimental data for the model calibration which are outlined in Table (7.5). According to [33] the only parameter that should have an LET dependence is the binary mis-repair probability parameter η, due to the closer spatial proximity and hence increasing likelihood of interaction between multiple DSB damages, leading to a higher degree of mis-repaired and lethal lesions. In comparison, the remaining parameters do not have LET dependence and are only cell line specific. However at high-LET we have observed the fraction of simple and complex DSB to change, therefore it is likely that the parameters corresponding to simple DSB repair will have less sensitivity than for low-LET.

Our Results	**Calibrated Against :**
35keV/μm	Belli et al. 1998 - V79 - 37.8keV/μm [26] Belli et al. 1998 - V79 - 34.6keV/μm [26]
16.45keV/μm	Folkard et al. 1989 - V79 - 17keV/μm [96] Prise et al. 1990 - V79 - 17keV/μm [29]
2.60keV/μm	Wouters et al. 2015 - V79 - 2.41keV/μm [38] Wouters et al. 1996 - V79 - 70MeV [113]

Table 7.5: The data sets (for various LET) selected from our results to calibrate the TLK model parameters using selected experimental studies.

Figures (7.19) – (7.21) depict dose-response measurements (survival fraction) by varying each parameter individually whilst keeping the rest fixed. Each dose response curve corresponds to the upper, lower and optimised parameter values.

Our observations are :

- The binary mis-repair probability, η, has the largest influence on the survival fraction. Increasing this probability suggests a higher number of mis-repaired lesions resulting in more efficient cell killing. When η is assigned a value of zero the damages in close spatial proximity will not interact, resulting in a higher yield of easily repairable damages. Figures (7.19(e)), (7.20(e)) and (7.21(e)) show that nearly all DSB will repair when η is equal to zero, meaning that nearly every cell survives for all absorbed doses.

- There is a clear correlation between the binary interaction probability η and particle LET. This is expected behavior due to the increase in DNA complexity, as discussed above.

- The repair and mis-repair rates λ_2, β_1 and β_2 display the lowest sensitivity of all the parameters with β_1 only differing at low-LET. Such low sensitivity parameters should have the lowest weighting in the calibration.

Parameter	This Work	From [48]
η	1×10^{-4}	5×10^{-5}
β_1	0.00 h^{-1}	0.00 h^{-1}
β_2	9×10^{-4} h^{-1}	9×10^{-4} h^{-1}
λ_1	0.671 h^{-1}	10 h^{-1}
λ_2	5.65×10^{-2} h^{-1}	5.65×10^{-2} h^{-1}

Table 7.6: The calibration of TLK model parameters used in our simulations versus those selected in [48].

- The sensitivity of the DSB re-joining rate parameter λ_1 appears to show some correlation with LET when it is at its maximum value (i.e. a slower repair half-time). There the survival fraction increases more at low-LET whereas the optimised and lower values of λ_1 do not yield a substantial difference. From [33], the re-joining rate parameter $\lambda_1, 2$ is defined in terms of the repair half-time[16] $\lambda_{DSB} = \ln(2)/\tau_{DSB}$, i.e. the repair half-time is inversely-proportional to λ_1.

  A smaller DSB repair half time (large λ_1, i.e. faster repair) results in a larger difference in the survival fraction at low-LET compared to the lower and optimised parameter values because there is a larger fraction of simple DSBs compared to high-LET (as shown in the previous section). In the mid to high-LET range a much larger fraction of damages are complex in nature, therefore the λ_1 parameter is not as sensitive here.

Ultimately, the parameters that are used in this simulation study were chosen based on those that provide the best agreement with the experimental data in each case. Table (7.6) shows the parameters chosen to generate our results and includes those from [48] for comparison. The parameters are kept consistent for all LET in this thesis due to the agreement with all data sets considered in the calibration. It is stated in [33] that η should vary with LET, increasing with LET. However, decreasing η at low-LET causes worse agreement with the data. Therefore it is concluded that this parameter will be kept constant.

[16] This is the time in which half of the DSB damages will be repaired.

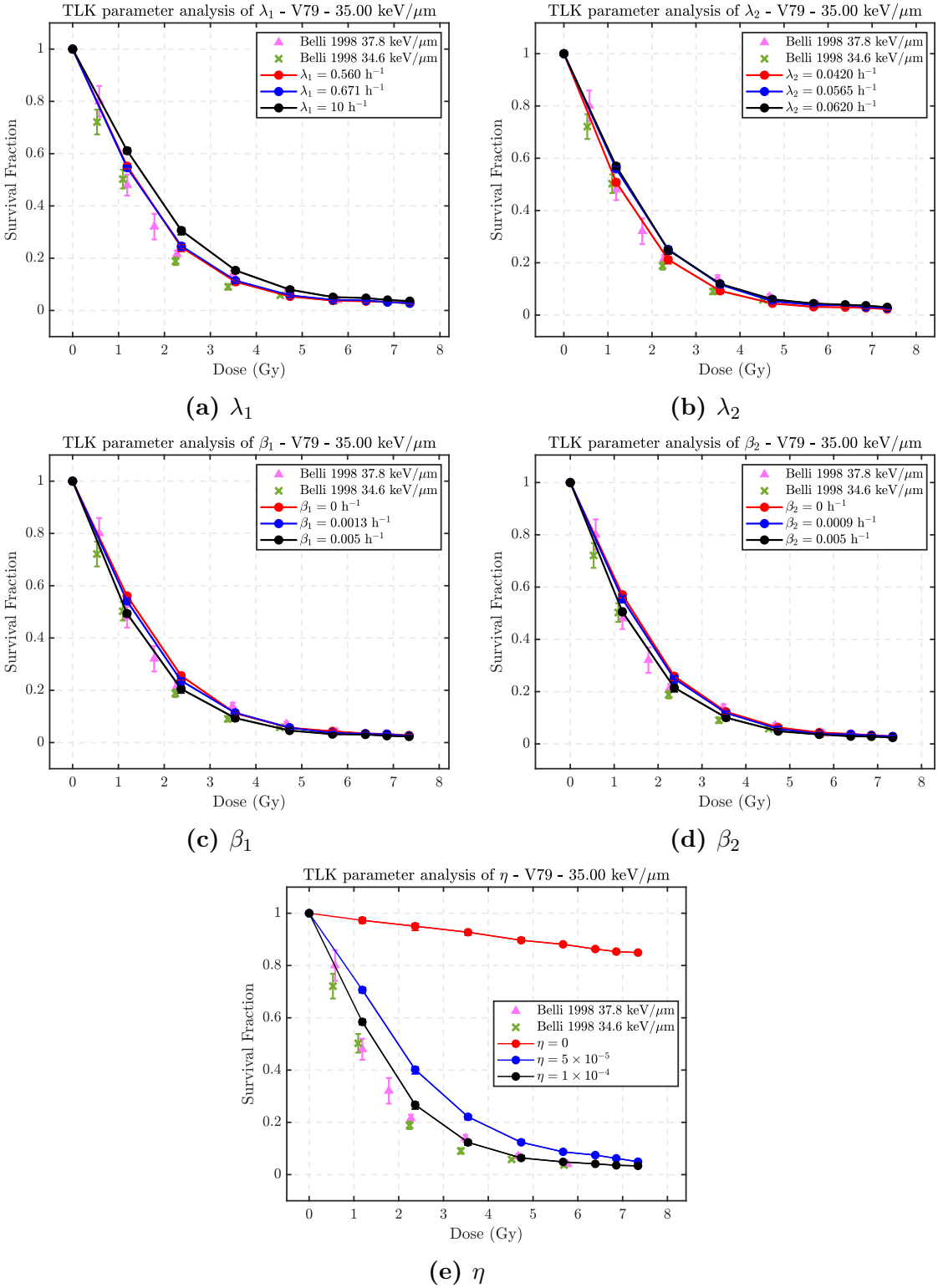


Figure 7.19: A TLK parameter sensitivity analysis of high LET (35keV/μm) proton radiation.

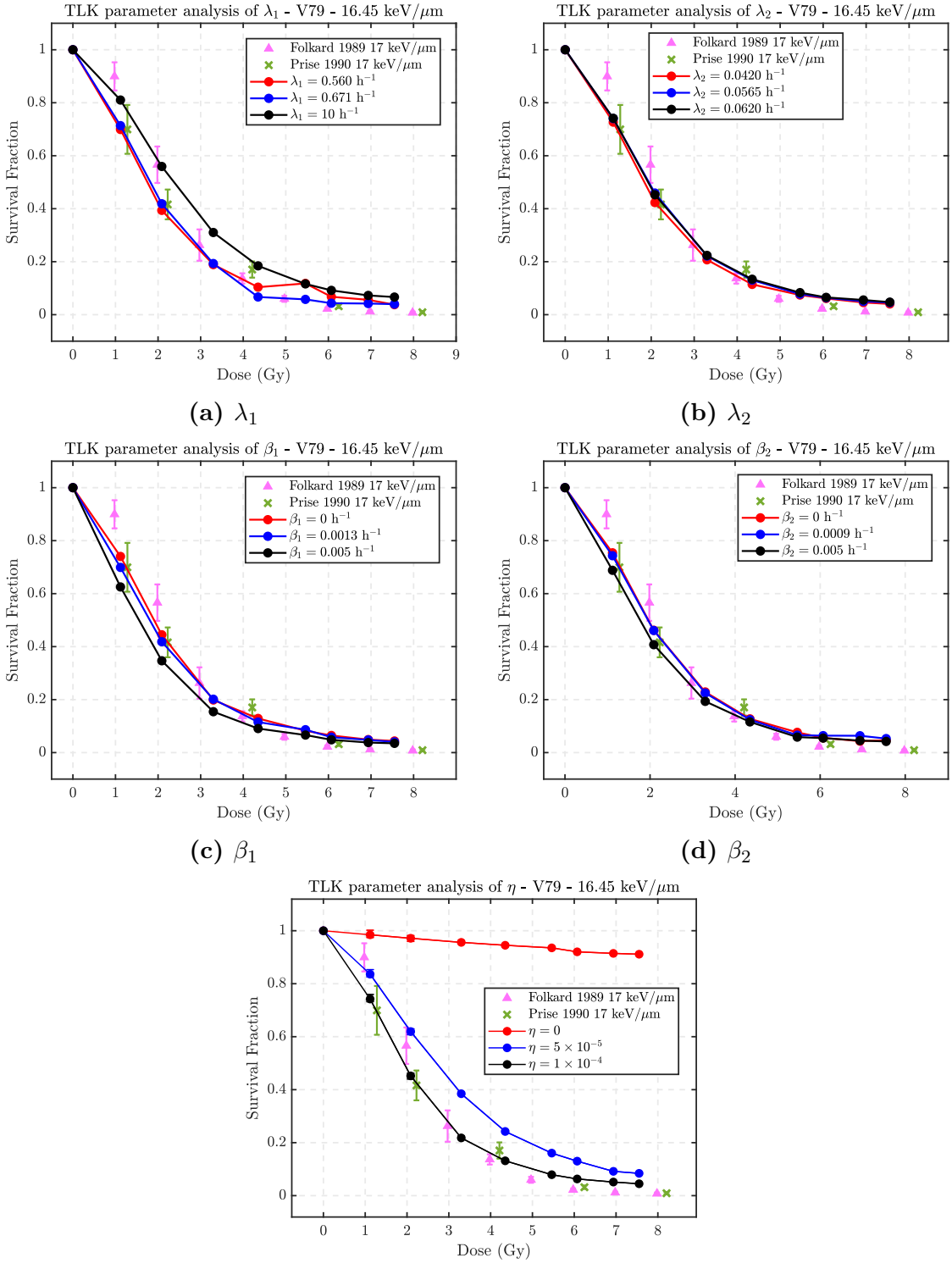


Figure 7.20: A TLK parameter sensitivity analysis of mid-range LET (16.45keV/μm) proton radiation. Note: the error bars are too small to be visible.

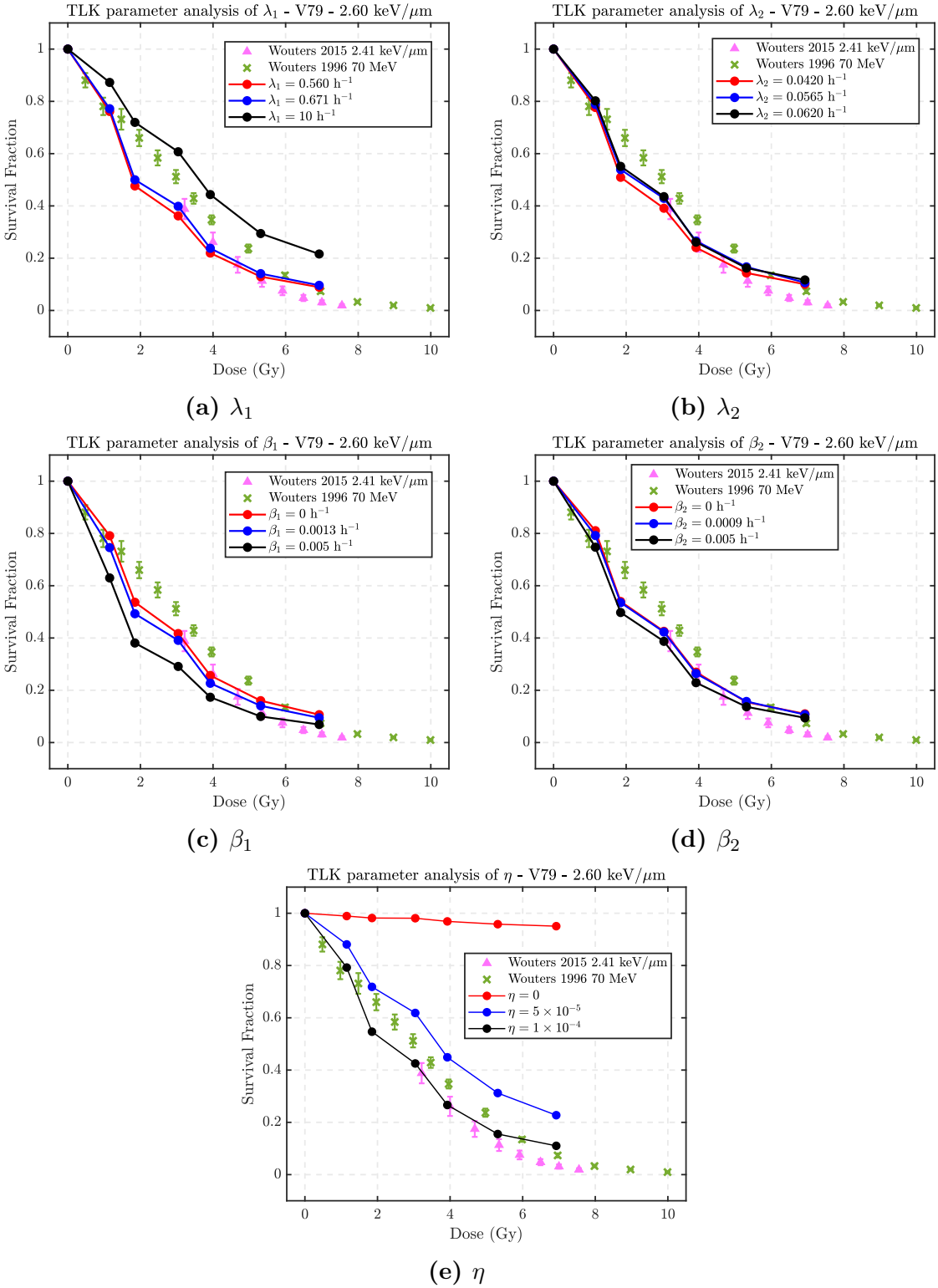


Figure 7.21: A TLK parameter sensitivity analysis of low LET (2.60keV/μm) proton radiation. Note: the error bars are too small to be visible.

7.4.2 Comparing Cell Survival Data with Experiment

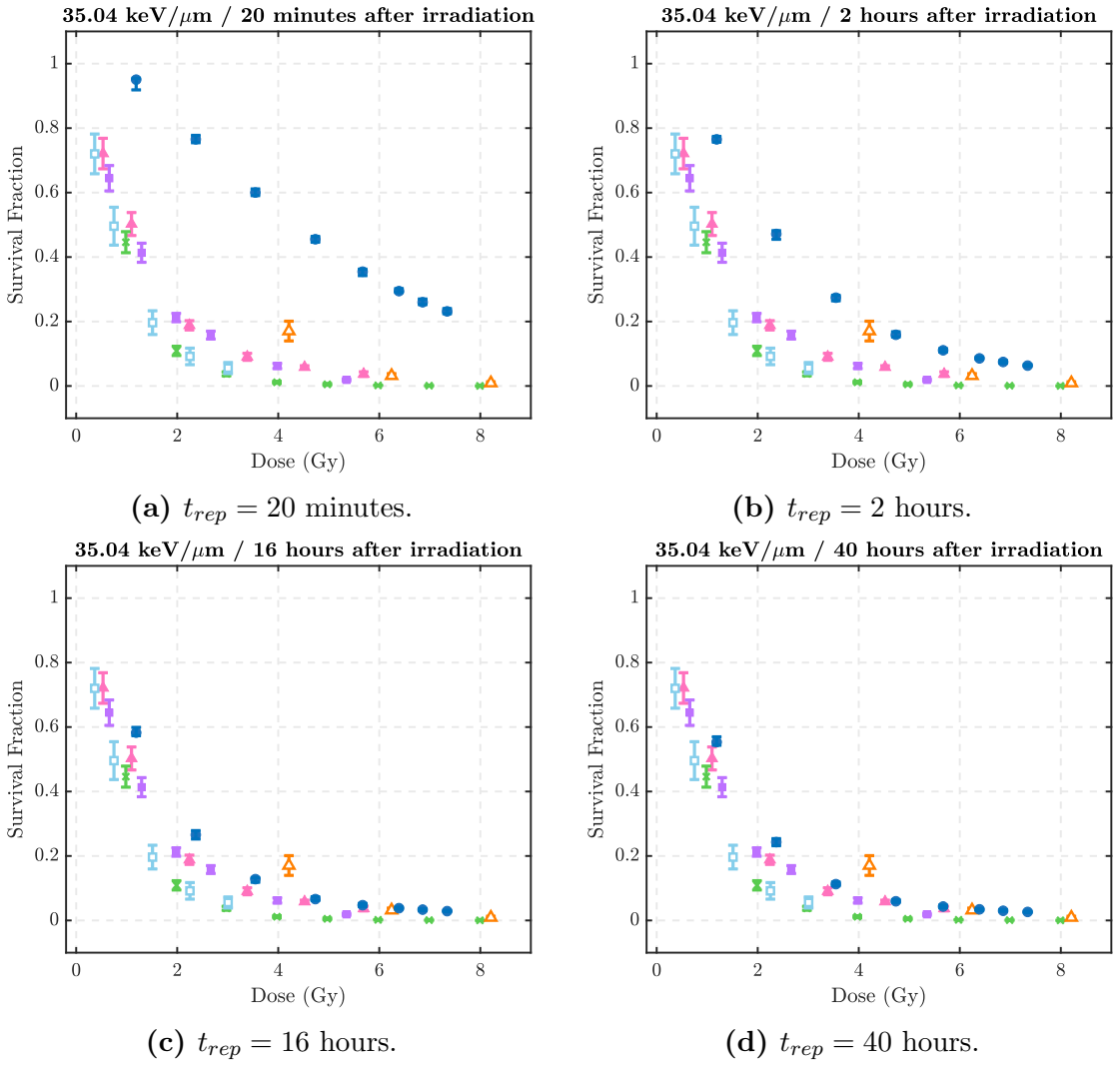
A total of 17 dose-response curves were generated using the TLK model framework, 10 with energy-varied LET and 7 with depth-varied LET. In this section a range of these results are presented with experimental data outlined in Appendix D. Below we present selected simulated data with increasing repair time for high, mid range and low LET proton beams. Due to the large number of data sets generated, not all of the results are depicted below. However, some additional results are presented in Appendix J to aid discussion.

At **high-LET**, the experimental data is shown to demonstrate higher cell killing efficiency with increasing dose compared with our low-LET results. The most varied results collected are in the region 30 - 35keV/μm and this is evident in the size of the error bars produced. The exact repair time at which the cell survival measurements were taken are not often reported in the literature, so it is logical to assume that the measurements are made after the system has become stable and no more lethally damaged cells remain. Any deviation from this in the experiments constitutes an unknown systematic error.

At 35keV/μm, the highest LET considered in this thesis, the measurements show good agreement with results from the literature (two data sets from [26], one from [96] and two [61]). As expected, the best agreement is observed at repair times greater than the time taken for a V79 cell to navigate through the four cell cycle phases (taken to be 15 hours in this study). In Stewart et al. [33], it is predicted that $\sim$20 DSBs Gy^{-1}cell^{-1} are re-joined quickly in CHO cells ($\sim$15 minutes) which are derived from the same species as the V79 cell[17]. To compare with this experimental work the survival fraction after 20 minutes repair time is depicted in Fig.(7.22(a)). The dose response relationship at this time appears to be approximately linear with a higher survival fraction than experiment. At this stage the cells have not been allowed sufficient time to fully repair or mis-repair, meaning that less lesions have been classified as lethal by the TLK model[18].

We begin to observe more agreement with experiment as the repair time increases. This is because the cells are now progressing through the four cell cycle phases where they encounter numerous checkpoints for the presence of damages. We can recall from our discussion in Chapter 2 that if these checkpoints in the cell cycle phase uncover a large number of DNA lesions, the cell will undergo apoptosis[19]. In Fig.(7.22(b)) where survival is measured following 2 hours of repair time, there is more agreement with experiment compared to Fig.(7.22(a)). It is likely, that each cell has at least progressed through one cell cycle phase at this time and hence a number have undergone apoptosis due to the presence of lethal damages.

[17]The species of origin is the Chinese Hamster.
[18]Recall that the survival probability is given by $S = \exp(-L_f)$.
[19]Programmed cell death.



(a) $t_{rep} = 20$ minutes. **(b)** $t_{rep} = 2$ hours.

(c) $t_{rep} = 16$ hours. **(d)** $t_{rep} = 40$ hours.

Figure 7.22: Our simulated cell survival measurements for the V79 cell line at various repair times (t_{rep}) after exposure to a 1.6 MeV, 35.04keV/μm proton beam. The experimental results presented utilise a monoenergetic proton beam (denoted by M) incident on cells with asynchronous cell cycle phases [26, 61, 96].

Figures (7.22(c)) and (7.22(d)), which depict the survival fraction measured after 16 and 40 hours of repair time, show very little deviation in their survival fractions and the most agreement with experiment. This is likely because the system has come to equilibrium and any repairable damages have either successfully or incorrectly repaired. This supports our prediction that most cell survival measurements are made for repair times greater than the length of time taken for a cell to complete all four cycle phases. The agreement with the experimental results from Belli et al. [26] where a 35 keV/μm proton beam is exposed to asynchronous V79 cells, is particularly noteworthy having the closest LET value to our results.

In the **mid-LET** range considered in this thesis in Fig.(7.23), the simulated data shows further agreement with experiment. Again, a similar trend of a large difference in survival fraction for small repair times is observed in Figs.(7.23(a)) and (7.23(b)). It is likely that the experimental results considered in this thesis were not measured soon after irradiation is completed, when the cellular system has yet to repair or mis-repair the damages induced. Good agreement is observed with experiment for repair times greater than the time taken for a single cell to traverse all cell cycle phases and is most likely due to the same behavior described in the high-LET case, shown in Figs.(7.23(c)) and (7.23(d)).

Figure (7.24) shows our results for the **lowest LET** considered in our simulation study with a proton energy of 50MeV corresponding to an LET of 1.23keV/μm. These low LET results are where we observe the least agreement with the experimental results. At repair times greater than the 15 hours required for a V79 cell to complete all four phases of the cell cycle, our measurements appear to overestimate the survival fraction compared with experimental results of similar LET for V79 cells from a Chinese hamster. Recalling that dose response measurements are highly dependent on the cellular environment, comparison of results across different experiments should be done with care. Therefore, the disagreement between our results and experiment in Fig.(7.24) should not be attributed to inaccuracy in our results based solely on the presence of experimental errors or inconsistencies from irradiation to cell preparation methods.

Furthermore, in the low-LET region indirect effects have a higher contribution to the total damage yield compared to high-LET radiation where direct effects dominate. This was discussed in Chapter 2, however we will briefly revisit the reason for this here. In a biological medium such as the nucleus, there are large amount of oxygen molecules with which ionising protons and secondary particles can react. In the low-LET region a majority of the interactions are via secondary particles because the primary protons are travelling so fast that they do not interact with the medium as much as high-LET protons. The secondaries can then propagate and react with the medium through radiolysis reactions. As discussed in Chapters 3 and 4, indirect effects can be simulated using the Geant4-DNA chemistry processes.

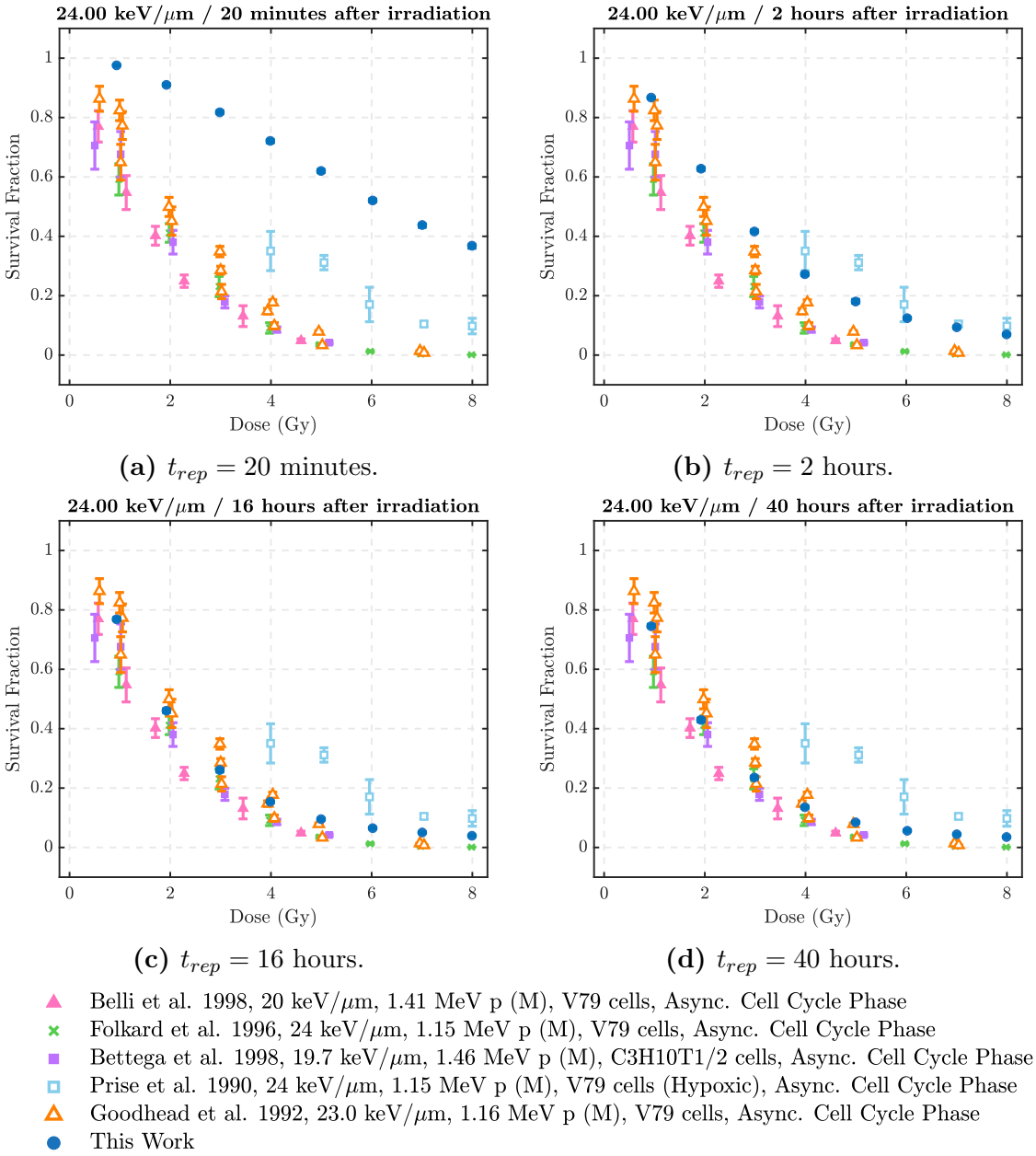
However for the purpose of our study we could not implement these chemical processes because the processing times and memory requirements were too large for us to perform repetitions of our measurements. By omitting the chemical processes in our simulations, it is likely that the survival fraction is overestimated, particularly in the low-LET region which explains the discrepancies between our results and experiment in Fig.(7.24) and Appendix J.

The cellular response with increasing LET at a repair time of 16 minutes (Figs.(7.22(a)), (7.23(a)) and 7.24(a)) appears to increase. The lowest LET demonstrates very little response at this time compared with the highest LET considered. A possible reason for the lack of response at short times for low LET could be that very few lesions have yet to be considered lethal due to the smaller presence of complex damages in particular. The fast-repairing lesions are more likely to be simple in nature compared with high-LET radiation, implying that many of these quick-repairing lesions may have already been repaired prior to this time.

7.4.2.1 Reliability and Validity of the Simulated Data

In addition to agreement with experiment, the reliability and validity of our measurements is important to the quality of results presented in Chapter 8. In addition to the size of the error bars in Figs.(7.22)–(7.24), Appendix J and Fig.(7.25), the variance in measurements appears to be the highest in our high-LET results. However at low-LET this uncertainty becomes so small that they are not visible in the figures. A possible reason for this is the additional spurious behavior of a high-LET particle traversing a medium, where the ionisation clusters occur with a higher probability. These additional interactions means that less primary protons are required to generate the doses simulated in this study, compared with that of low-LET particles. This can be explained in terms of the MC simulation uncertainties where the uncertainty is proportional to $1/\sqrt{N}$. Therefore the mean behavior per particle track is estimated more accurately for low-LET particles, thus explaining the smaller uncertainties present. Figure (7.25) demonstrates the extent of this observation, whereby the individual simulation repetitions of dose response are plotted for a high and low-LET beam. One can easily see the larger variance of our high-LET results in comparison with that of low-LET.

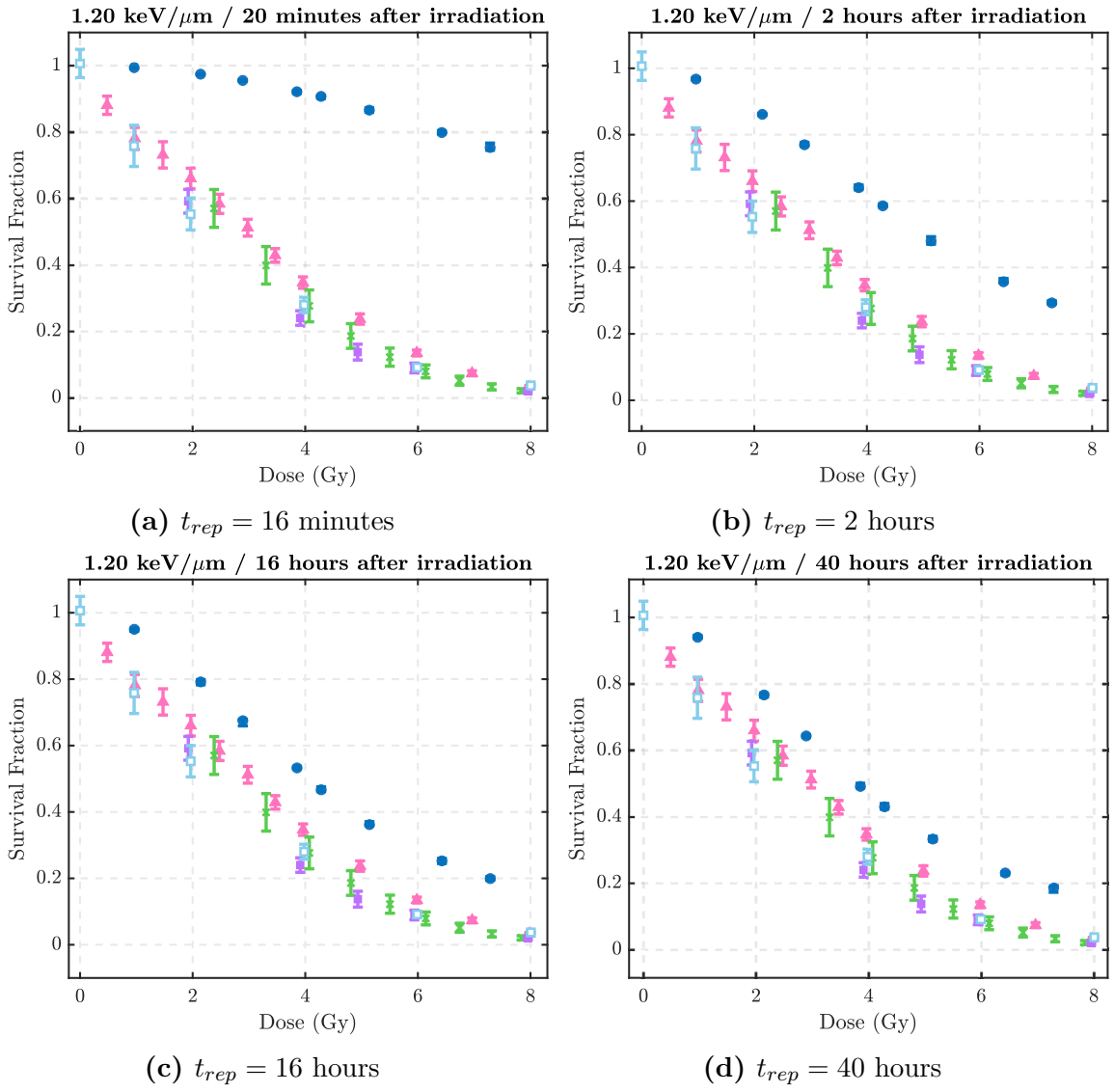
It is also important to compare our results with what we expect from theory. Figure (7.26) depicts four survival curves corresponding to different energy-varied LET, measured at 16 hours repair time. From theory, one expects more efficient cell killing from high-LET radiation. It is clear that the worst cell killing efficiency is present in a 1.20keV/μm proton beam, where the "shoulder" behavior discussed in Chapter 2 is most present in Fig.(7.26). As the LET reaches 11 and 16keV/μm, the "shoulder" behavior disappears and the cell killing becomes more linear in the low-dose region. The two highest LET considered not only demonstrate linear behavior



(a) $t_{rep} = 20$ minutes. **(b)** $t_{rep} = 2$ hours.

(c) $t_{rep} = 16$ hours. **(d)** $t_{rep} = 40$ hours.

Figure 7.23: Our simulated cell survival measurements for the V79 cell line at various repair times (t_{rep}) after exposure to a 2.0 MeV, 23.67 keV/μm proton beam. The experimental results presented utilise a monoenergetic or modulated proton beam (denoted by M and S respectively) incident on cells with asynchronous cell cycle phases [26, 29, 96, 168, 169]. Further details on the experimental data are presented in Appendix D.

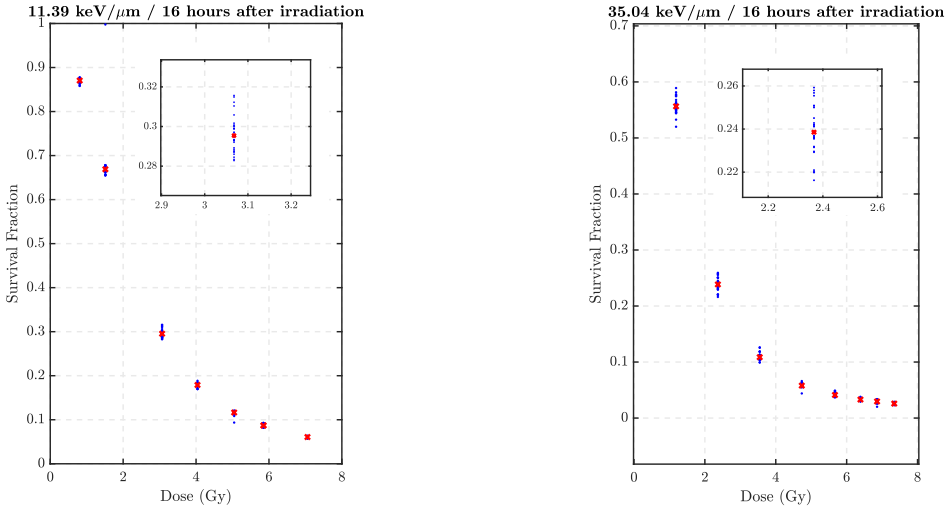


(a) $t_{rep} = 16$ minutes **(b)** $t_{rep} = 2$ hours

(c) $t_{rep} = 16$ hours **(d)** $t_{rep} = 40$ hours

Figure 7.24: Our simulated cell survival measurements for the V79 cell line at various repair times (t_{rep}) after exposure to a 50.0 MeV, 1.2keV/μm proton beam. The experimental results presented utilise a modulated proton beam (denoted by S) incident on cells with asynchronous cell cycle phases [38, 113]. Further details on the experimental data are presented in Appendix D.

in the low dose region, the cell killing becomes slightly less efficient in the high-dose region. This is in the region where the LQ model is shown to be the least effective (Chapter 5).



(a) Dose response of a 5 MeV (11keV/μm) proton beam incident on V79 cells.

(b) Dose response of a 1.6 MeV (35keV/μm) proton beam incident on V79 cells.

Figure 7.25: The "raw" data of each dose response simulation of 30 repetitions with a single dose point magnified.

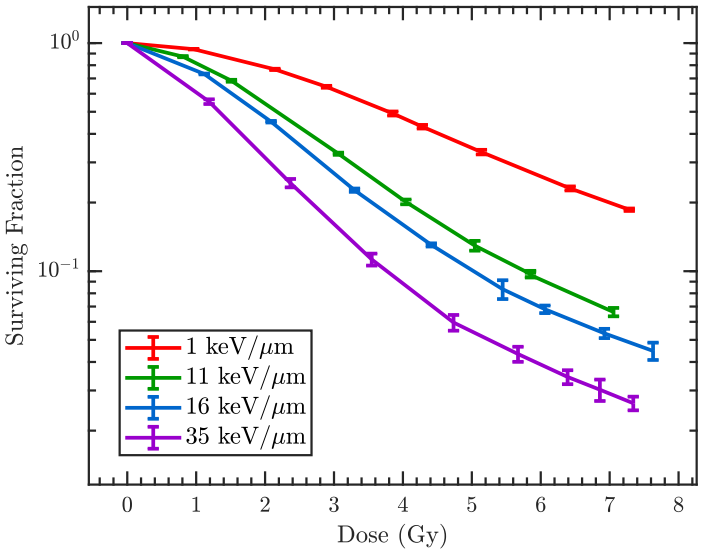


Figure 7.26: The trend in dose response of V79 cells with a range of LET values from our simulated data.

7.4.3 Relative Biological Effectiveness

The RBE calculations for our data will be performed in Chapter 8, where we present the results of testing the proposed model described in Chapter 6. However, it is still worth discussing in this section because we can infer the trend in RBE by observing the dose response curves presented in this chapter. Recall the definition of RBE in Chapter 2, where we can take the ratio of a given parameter required for the radiation to achieve the same biological effect as a reference radiation, commonly a photon beam. Photon radiation has not been the focus of this study, so it is worth noting that it has an extremely low LET in comparison with protons because photons are massless. Therefore to perform RBE calculations we require a photon dose response curve, i.e. a reference curve. A 6MV photon beam[20] is collected using an identical simulation setup to that described in Chapter 4. The small LET implies that a large number of primary photons are required to generate absorbed doses in the cells that are comparable to our results using proton radiation. This meant that numerous fractions of primary photons in the order of 10^7 are required for each of these doses. Aside from the nature of the radiation, all other parameters are kept consistent with Chapter 4. The resultant dose response curve that will be utilised for these RBE calculations is presented in Fig.(7.27).

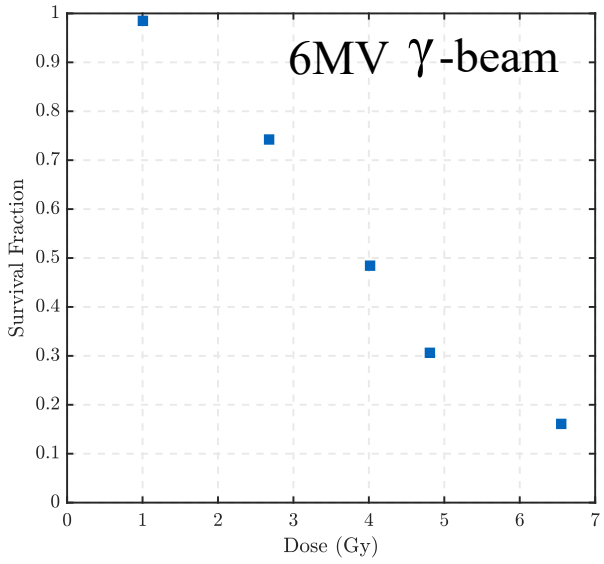


Figure 7.27: The reference dose response curve generated by exposing 530 V79 cells to a 6 MV photon beam in our simulations.

[20] 6MV does not mean a photon beam of 6MeV but a spectrum of energies where the maximum is 6MeV and the mean energy is ∼2MeV.

7.5 Simulation Limitations

Accompanying every scientific model are a series of assumptions and approximations. The radiobiological model from Chapter 4 used to generate the results presented in the current chapter is no exception to this. In the field of radiobiology, it is not uncommon for a model to require large amounts of computational power and memory. Often these models must include a series of approximations and assumptions to optimise processing times due to the complex nature of the cellular system and its response to ionising radiation. The methods used to achieve this balance are discussed in detail in Chapter 4. The current section discusses the impact that these limitations and assumptions have on the results.

The most noteworthy approximation is the exclusion of chemical processes that can induce indirect DNA damages via the production of charged radicals from the radiolysis of water. The result of this is the overestimation of the survival fraction after irradiation with low-LET particles. The literature has shown that indirect effects are the dominant process for inducing DNA damage for low-LET radiation (such as photon radiation [170]) - suggesting that such processes should be modelled for photon radiation-based studies. In the current work we mostly consider high-LET radiation (>10keV/μm), where the direct effects are dominant in DNA damage induction.

The exclusion of chemical processes from the simulations performed in this thesis is not ideal considering that the objective of the model is to predict DNA damage as accurately as possible. However, in the current workflow one would require at least 100GB of storage space per repetition and large amounts of computation power to perform the required post-processing necessary to achieve the desired results. This study is mainly statistical in foundation and therefore, the number of repetitions required to perform adequate analysis would not be possible to achieve in the available time. Re-designing this workflow to allow the inclusion of chemical processes should be explored in future work.

A realistic particle track contains ionisation yields with many orders of magnitude. As is the case with most particle transport simulations, this is not achievable even with the current computation power accessible to us. Therefore, the particle track is censored in regions where detailed track structures are not vital to the accuracy of the results. In our case, the track structure was simulated with more detail in the DNA-rich nucleus than other organelles. Overall, this does not impact the results since the nucleus region is the only organelle of interest in the current model. Therefore, the overall impact on the final endpoints of the model is negligible.

Various assumptions are made about the conditions inside of a V79 cell. Realistically, the morphology, material composition, organelles present and environmental

conditions are non-homogeneous due to the imperfections of cellular biology. The limitation this places on the current model is that we are assuming a similar dose response from every cell in the system. In our simulations we approximate each cell to be a perfect sphere with the same dimensions. In experiment this is not always the case - particularly when an asynchronous distribution of cell cycle phases are involved. The radiation sensitivity and physical dimensions of the cell can change according to biological parameters including the cell cycle distribution. Future efforts can focus on the dose response for cell lines in homogeneous cell cycle phases with more realistic geometries.

The targets of irradiation are ultimately the region of the cell that is sensitive to radiation, the nucleus. As discussed above, we approximate the shape of the V79 cell to be a sphere of radius 8.5μm and in the centre is the nucleus of radius 7μm. Our assumption does not reflect the imperfect biology of the cell where the shape and size can fluctuate according to the conditions in which they are kept and the parameters of the cell environment itself.

Finally, the large computation, memory and storage requirements of the results in the current model limits the number of repetitions performed. Statistically, the maximum sample size considered in the current model ($n = 30$) makes statistical analysis difficult and a larger sample size in future studies will improve our results. Our results demonstrate validity and reliability due to their agreement with theory and experiment and the measurement uncertainties - therefore not placing limitations on the results given by the model. We decided to take a "best of both worlds" approach by taking a sample size that is just adequate enough for statistical analyses whilst still being able maintain the level of detail in the model from a radiobiological perspective. The result of this is not so much a limitation on the model itself but on the statistical analysis performed on the generated results.

7.6 Conclusions

The objective of this chapter is to compare our simulated data against experiment and radiation response theory. Due to the complex nature of the system there is no data in the literature under identical physical, chemical and biological parameters as in our simulations. Our comparisons led to the following conclusions :

- The ionisation and DNA damage yields from our simulations are within the experimental range. There are discrepancies between our results and the literature however this is likely due differences in the classification of damage complexities, i.e. the definition of a simple and complex DSB differs between our simulations and the literature.
- The DNA damage repair stage of our simulations accurately reflects what occurs in experiment.

- The most important outcome of this chapter is the simulated survival fraction data and its agreement with the experimental data for high and mid-LET radiation after the system has reached equilibrium. For low-LET the simulated data does not agree with experiment as much as the high-LET region because we could not include the contributions to DNA damage from indirect effects (chemical processes).

8

Investigating The Theory of Fractionality in Radiation-Induced Cell Death

Since the development of target theory models for radiation-induced cell death in the 1960s, the Poisson process has remained predominant in describing cell response to ionising radiation. In this chapter we test our new stochastic model proposed in Chapter 6 and compare its performance against established radiobiological models based on the Poisson process (Pp), focusing on the famous Linear-Quadratic (LQ) model. Our model is tested against different biological endpoints and conditions under goodness-of-fit and error (residual) analysis.

In Chapter 5 we concluded that the LQ model has a limited range of applicability and that the most disagreement with experimental data is observed in the high-LET region. We concluded that in such regions the assumption that radiation-induced cell death obeys a Poisson process is no longer valid. Furthermore, the fits reported in the literature suffer from several deficiencies including the violation of regression analysis assumptions and a poorly determined goodness-of-fit. Additionally, we observed the extent to which the calculation of RBE is affected when the dose-response curve is fit to a model that poorly explains the data.

Chapter 6 aimed to explain the discrepancies between the Poisson process and experimental data that were uncovered in Chapter 5 by proposing a fractional Poisson process (fPp). Briefly, the fPp is a generalisation of the Pp, introducing more flexibility capable of explaining the presence of over/underdispersion[1] in the formation of lethal lesions in the cell as a result of ionising radiation - a property not accounted for in the Pp. The idea for the fPp model was developed in close collaboration with Dr. Markus Kreer of Johann Wolfgang Goethe-Universität, Frankfurt,

[1] Over/underdispersion is the result of the mean and variance not equating to each other which is a requirement of the Poisson process.

Germany.

In Chapter 7 we tested our Monte Carlo simulations described in Chapter 4 against experimental data and theory. We concluded that our simulated results are within the accepted experimental range at different biological endpoints including DNA damage yields and cell survival. Moreover, the fact that our simulated data is consistent with experiment gives us reasonable grounds to test our proposed model at different stages of dose-response.

In this chapter, our proposed model, in addition to established radiobiological models, will be tested against experimental data from the literature (given in Appendix D) as well as our simulated data (from Chapters 4 and 7) to verify the model's limit. This is achieved by testing the response of cells irradiated with proton radiation for the presence of fractionality at different stages of biological response including lethal DNA lesion yield and survival. We test our proposed model against the Pp (the LQ model) at the survival endpoint using goodness-of-fit and error (residual) analysis on both experimental and simulated data. Then we observe the effect on the calculation of RBE. As a benchmark study we compare the performance of our stochastic model against other established radiobiological models (all of which are discussed in Chapter 2), including the Multiple Hit Multiple Target (MHMT) [34], Repair-Misrepair (RMR) [31], Lethal-Potentially Lethal (LPL) [32] and LQ [23, 35] models.

8.1 The fPp Model: A Summary

It is useful to recapitulate some of the main concepts discussed in previous chapters (including Chapters 4, 5, 6 and 7) which are crucial to discussions in the current chapter.

Consider the simple sketch presented in Fig.(8.1) where the cells are considered to be in one of three states (i) $|i\rangle$ - the initial state, (ii) $|m\rangle$ - the intermediate state and (iii) $|f\rangle$ - the final state where the response is measured.

I - In the Real Picture (Experiment)

- The system begins in a relaxed initial state $|i\rangle$ where all the cells are assumed to be alive and contain no lethal lesions. From here, the system will try to equilibrate (repair the DNA damage) until the final state $|f\rangle$, where lethal lesions are present in the targets (cells) after sufficient repair time $[0, t_{rep}]$ has lapsed. Overall, our system's dynamics can be written abstractly as

$$|i\rangle \rightarrow |m\rangle \rightarrow |f\rangle, \quad t \in [0, t_{del} + t_{rep}]$$

and we omit the intermediate state by treating it as a "black box" leaving us with

$$|i\rangle \rightarrow |f\rangle \; t \in [0, t_{del} + t_{rep}].$$

- Cell exposure to radiation begins at $t = t_{del} = 0$ where it moves into the $|m\rangle$ state. Energy depositions due to ionising radiation produce DNA damage (potentially lethal lesions), some of which are repairable and some are not.
- The cells respond to the presence of damages via a series of biological and chemical response (repair) pathways which continue throughout irradiation, i.e. the radiation and repair timescales overlap.
- After delivery of the desired dose, the irradiation phase stops.
- After irradiation ceases, the repair process timescale will continue until the saturation state, $|f\rangle$ (i.e. there are no more repairable damages), is reached (when $t_{rep} = t_{max}$).

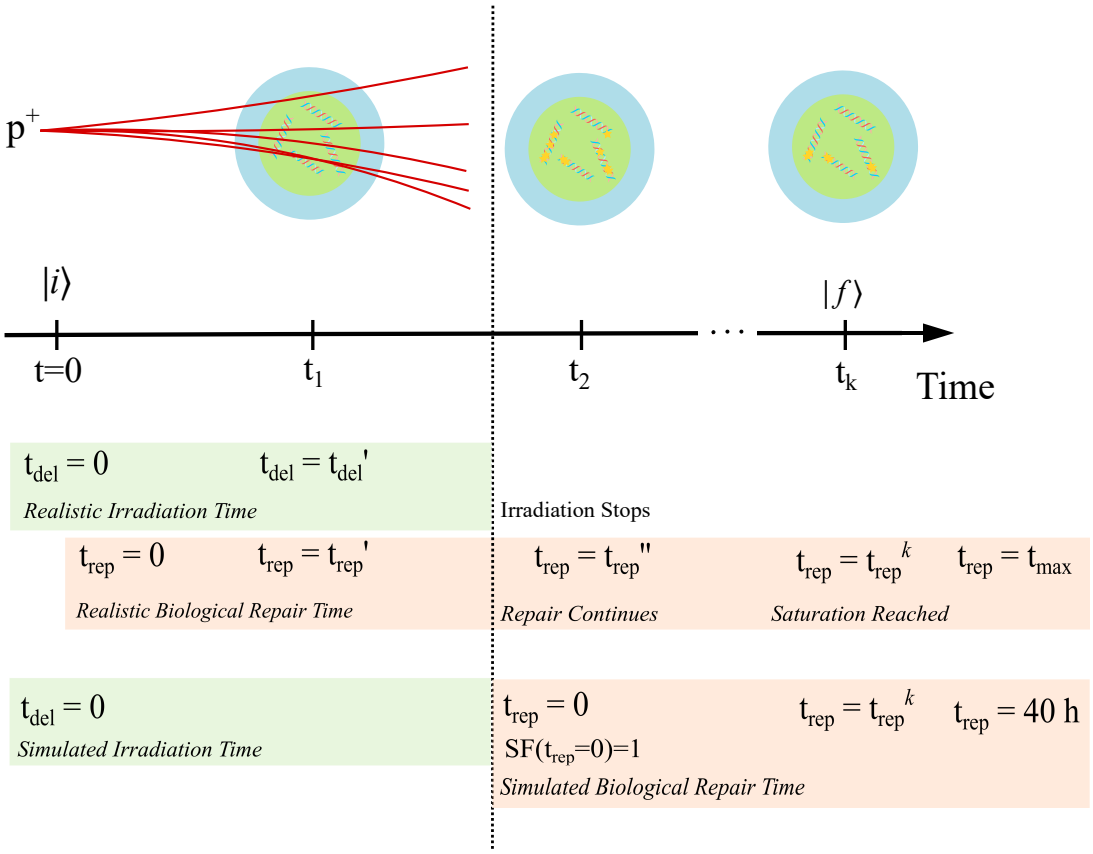


Figure 8.1: Timescales of the physical picture in experiment and our simulations.

II - Our Simulations

In our simulations the V79 Chinese Hamster cell is exposed to proton radiation of varying mean beam energies between 50MeV and 1.6MeV with a corresponding LET range of 1.2 to 35keV/μm. Irradiation is performed for eight different absorbed doses. All the simulation parameters are provided in Chapters 4 and 7, however the chemical processes (indirect DNA damage effects) could not be implemented due to time restrictions. The basic setup of our simulations is provided in Fig.(8.2).

We perform our simulations with the following in mind;

- The repair processes do not commence until the irradiation stage is complete ($SF = 1$ at $t_{rep} = 0$), which differs from the real picture.
- The Two-Lesion Kinetic Model (TLK) [33, 48] is used to define the repair processes and determine the survival fraction. This model contains three differential equations to describe the time dependent dynamics of simple and complex DSB repair and the formation of lethal lesions.
- When our simulated results are compared with experiment, it should be after the saturation (equilibrium) time is reached ($t_{rep} > 16$).
- The survival fraction is measured until 40 hours after irradiation has ceased.

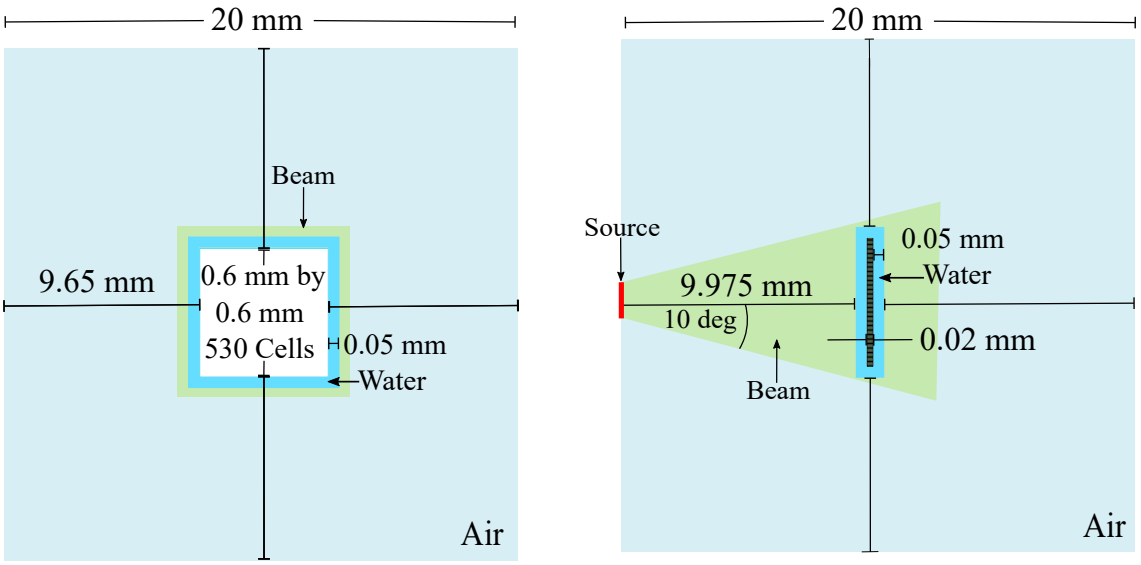


Figure 8.2: Experimental setup in the TOPAS simulation with dimensions indicated.

The lethal lesions that have formed within a cell when the system reaches its final state $|f\rangle$ governs its fate (death or survival). To our knowledge, nearly all established radiobiological models approximate the distribution of lethal lesions within a given cell using a Poisson process.

III - Poisson Process (Pp) as a Dose-Response Model

The Poisson process is the most common count model utilised in radiobiological research with the validity constraint that the mean and variance of the PDF must be equal - we call this the *equidispersion condition*. Detailed discussions are provided on this in Chapter 6 where :

- We showed that for high (35keV/μm) and low (1.2keV/μm) LET radiation, the equidispersion condition is violated (as shown in Fig.(8.3)). For high-LET radiation there is a strong presence of overdispersion of lethal damages (Figs.(8.3(a)) and (8.3(b))), whilst for low-LET radiation we observe *under-dispersion* (Figs.(8.3(c)) and (8.3(d))).

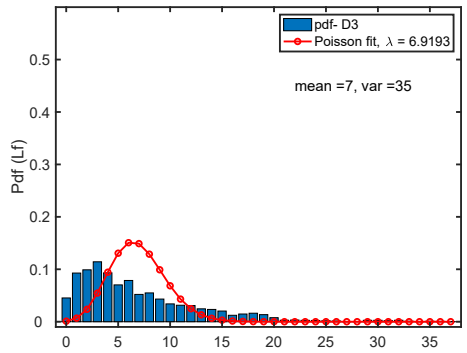
- We concluded that the presence of over/underdispersion in the distribution of lethal damages can influence the resulting cell survival probability. This provides an explanation for the limited range of applicability of the LQ model. Certainly for low-LET, the LQ model can describe the data reasonably (concluded in Chapter 5). Additionally, it yields a reasonable fit to low dose regions in many sets of experimental data as presented in Chapter 5. However, this is not the case under high-LET exposure.

- Similar observations have been made in the literature [20, 39, 69, 171].

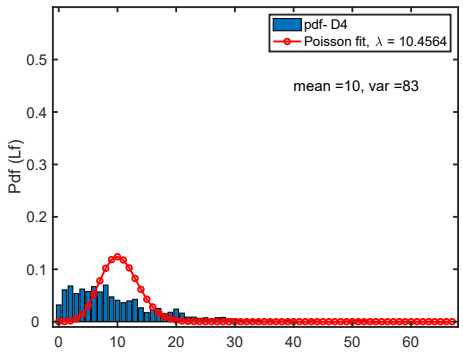
Now let us recall the specifics of the Pp in Table (8.1) :

CDF of Inter-arrival Time	$F(t) = 1 - e^{-\lambda t}$
PDF of Inter-arrival Time	$f(t) = \lambda\, e^{-\lambda t}$
Survival Function	$p_0(t) = 1 - F(t) = e^{-\lambda t}$
Probabilities	$P_n(t) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}$
Mean	λt
Variance	λt
Hazard Function	$h(t) = \frac{f(t)}{1-F(t)}$

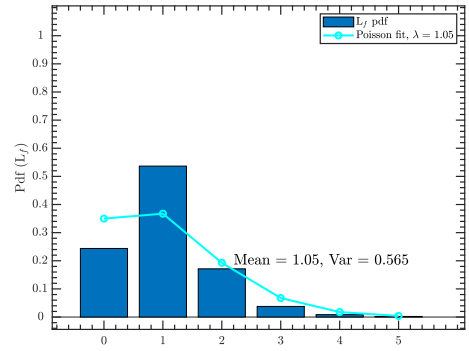
Table 8.1: A summary of the Poisson process properties.



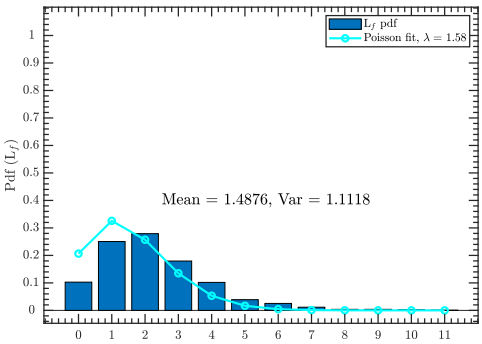
(a) LET = 35keV/μm, D = 3.54Gy



(b) LET = 35keV/μm, D = 4.73Gy



(c) LET = 1.20keV/μm, D = 4.28Gy



(d) LET = 1.20keV/μm, D = 5.13Gy

Figure 8.3: Normalised histograms (PDF) of the lethal lesions induced in our system fitted to the Poisson PDF. The upper plots show the results in the high-LET region and the lower plots are in the low-LET region.

IV - The LQ Model as a Dose-Response Relation

The LQ model is a mathematical model that describes the survival fraction as a function of absorbed dose due to the presence of lethal lesions inside the targets (cells). It is a time-dose relation and corresponds to the zero count state ($probN(t) = k$ with $k = 0$) of the Poisson process :

$$S = \exp(-\alpha D - \beta D^2) \tag{8.1}$$

where α and β parameterise the effect of the lethal lesions produced via a single hit (a single ionisation) or multiple hits (multiple ionisations), respectively.

Statistical validation in Chapter 5 using LQ model fits to the experimental data uncovered several deficiencies including a limited range of applicability and the violation of regression analysis assumptions outlined in Appendix E. Similar findings are present in the literature [17, 18, 20, 23, 41].

V - Fractional Poisson Process (FPp)

To account for the discrepancies of the Poisson process in radiation-induced cell death, we proposed the fPp as a counting process to describe the final lethal lesion distribution and cell survival fraction. It is more "flexible" than the Pp in that it accounts for the presence of over/underdispersion in addition to equidispersion.

- The probabilities of each target in the system sustaining $k = 0, 1, 2, ..., n$ lethal lesions in state $|f\rangle$ can be calculated under the fPp using

$$p_k^{(\gamma)}(t) = \frac{(\lambda t^\gamma)^k}{k!} \sum_{j=0}^{\infty} \frac{(k+j)!}{j!} \frac{(-\lambda t^\gamma)^j}{\Gamma(\gamma(k+j)+1)} \quad . \tag{8.2}$$

  at a given time t.

- The survival probability ($k = 0$) is written as

$$p_0^{(\gamma)}(t) = E_\gamma(-\lambda t^\gamma) \quad , \tag{8.3}$$

  where $E_\gamma(-\lambda t^\gamma)$ is the Mittag-Leffler function, which corresponds to the zero count state ($k = 0$) in the fPp as discussed in detail in Chapter 6 and Appendix I.

- Now let us recall the specifics of the fPp in Table (8.2) :

CDF of Inter-arrival Time	$F(t) = 1 - E_\gamma(-\lambda t^\gamma)$
PDF of Inter-arrival Time	$f(t) = \lambda t^{\gamma-1} E_{\gamma,\gamma}(-\lambda t^\gamma)$
Survival Function	$p_0(t) = 1 - F(t) = E_\gamma(-\lambda t^\gamma)$
Fractional Probabilities	$p_k^{(\gamma)}(t) = \frac{(\lambda t^\gamma)^k}{k!} \sum_{j=0}^{\infty} \frac{(k+j)!}{j!} \frac{(-\lambda t^\gamma)^j}{\Gamma(\gamma(k+j)+1)}$
Mean	$\mu = \frac{\lambda t^\gamma}{\Gamma(\gamma+1)}$
Variance	$\sigma(\mathrm{Var}) = \frac{\lambda t^\gamma}{\Gamma(\gamma+1)} \left\{ 1 + \frac{\lambda t^\gamma}{\Gamma(\gamma+1)} \left[\frac{\gamma B(\gamma, 1/2)}{2^{2\gamma-1}} - 1 \right] \right\}$
Hazard Function	$h(t) = \frac{f(t)}{1-F(t)}$

Table 8.2: A summary of the fPp properties.

- The fPp is a generalisation of the Pp, when $\gamma = 1$ the fPp goes to the standard Poisson process. In the literature, the composite Poisson process such as negative binomial processes and the modifications based on the Neyman type A distribution have been proposed to account for over/underdispersion of lethal lesions in cells [69]. However, the Neyman distribution can only account for non-equidispersed data, whereas the fPp is able to account for both equidispersed and over/underdispersed data.

- In Figs.(8.4) and (8.5) we present the resultant shapes of the dose-response curve under the Pp and fPp ($\gamma = 0.60$, i.e. a high degree of fractionality). In Fig.(8.4), we observe a one-to-one relationship between the dose and formation of lethal lesions. The result is a strictly exponential relationship between the number of lethal lesions present and the survival fraction meaning that the probability of a given cell surviving when a single lethal lesion ($\lambda = 1$) is present is 0.37.

  Under the fPp, the situation becomes a little more complex (Fig.(8.5)). The one-to-one relationship between absorbed dose and lethal lesion formation is no longer valid. The number of lethal lesions present in a given cell is governed by Λ and we must know the absorbed dose required for Λ to take a given value. Under a high degree of fractionality, the distribution of lethal lesions inside a cell has an increasingly large variance compared to the mean. Therefore, the predicted survival probability differs greatly from that of the Pp. The resultant survival fraction is "curved" compared to the Pp due to the deviation from the one-to-one dose-lesion relationship present under the Pp.

- Our proposed cell survival fraction model for dose-response is the "Mittag-Leffler" count model with a special argument which is discussed in detail in Chapter 6 and Appendix I :

$$SF = E_{\gamma}\left[-(aD + bD^2)^{\gamma}\right] \quad , \tag{8.4}$$

  where a and b are constrained to the positive domain and $\gamma \in [0, 1]$.

We will now test out proposed model against the experimental and simulated data.

8.2 Pp versus FPp Models to Predict DNA Damage

An important aspect to modern radiobiological modelling is the ability to explain cell response to ionising radiation at multiple biological endpoints. When we expose

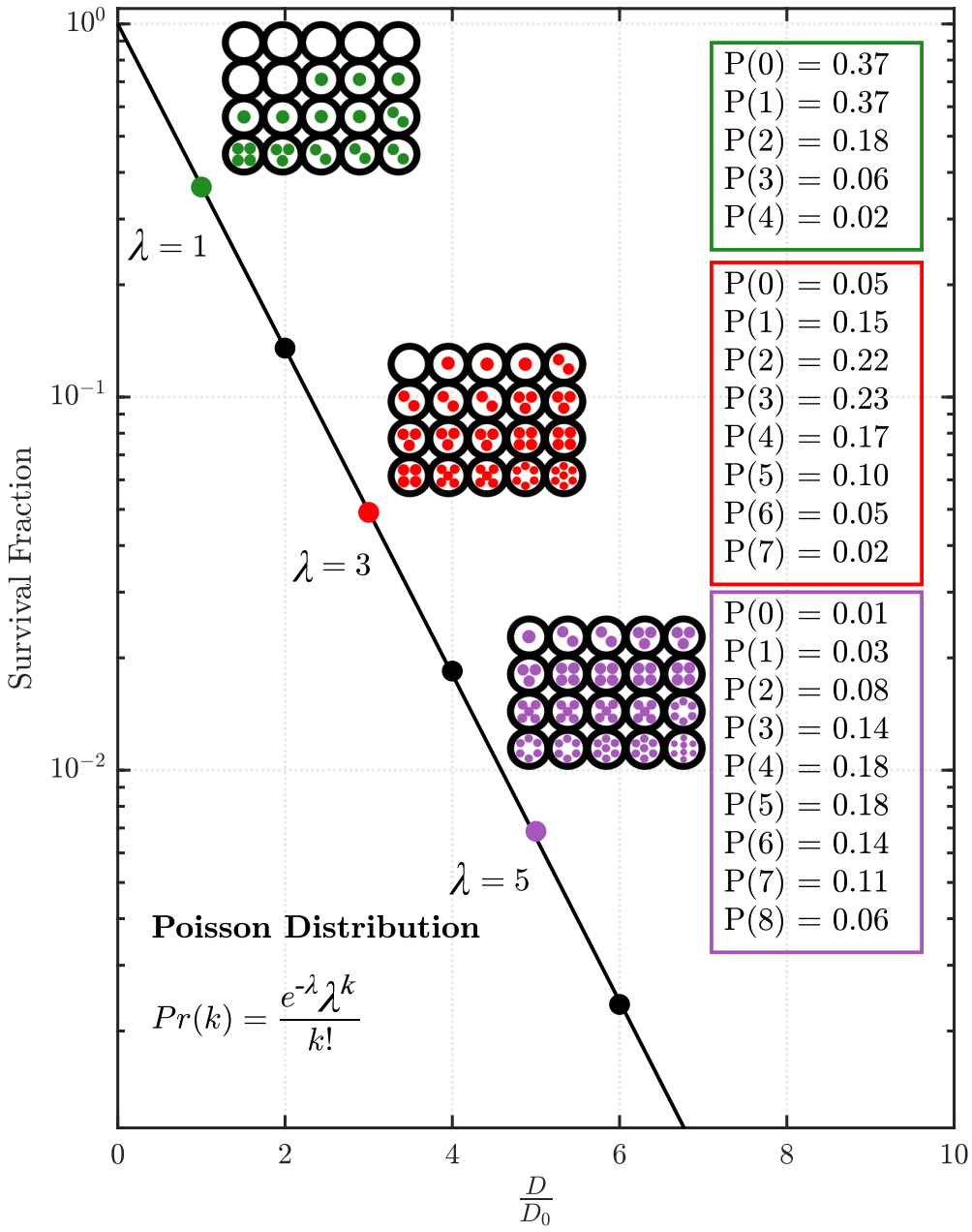


Figure 8.4: A survival fraction constructed from Poisson distribution probabilities, $1 - \exp(-\lambda)$, for the mean number of lethal lesions produced in the system, λ. Graphical depictions of the cell damage probabilities are shown for select SF. Credit to [111] for the original figure idea. Note: the total probabilities should add to 1, however probabilities < 0.01 are not shown.

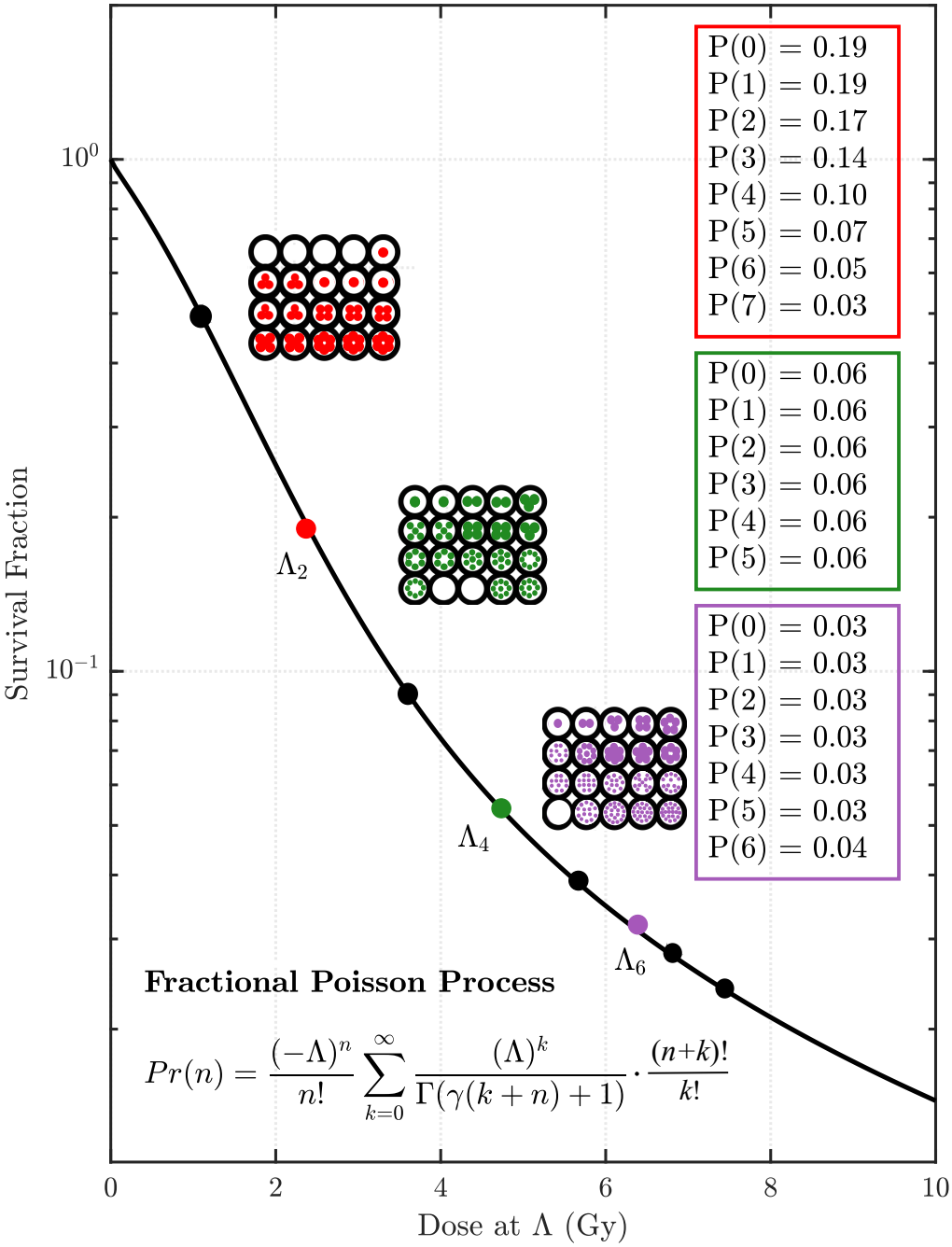


Figure 8.5: Survival probabilities as a function of the dose required to produce a given number of lethal lesions Λ assuming a fractional Poisson process with $\gamma = 0.60$ with a graphical depiction. Note: due to the large distribution of lethal lesion counts, not all probabilities are shown and the graphical depictions are approximations of the numerical probabilities. Note: the total probabilities should add to 1, however probabilities < 0.01 are not shown.

a cell to ionising radiation, the first endpoint to occur in our physical picture is the DNA damage yield.

In Chapter 4 we have outlined how the TLK Model is implemented into the simulations. To recall, we determine the probability of survival for every cell in the simulation using $\exp(-L_f)$, where L_f is the number of lethal lesions present in the cell at repair time t_{rep}. We then deduce whether the cell is alive or dead by selecting a uniform random number and if this is below the probability of survival, it is classified as "dead". This final step essentially eliminates the assumption of Poisson statistics, deeming our model appropriate for testing against survival. All the results presented in this section are measured at a repair time of 40 hours (after irradiation has ceased in the simulated picture) when the system reaches saturation.

Before we commence our investigation into lethal DNA lesions, it is useful to observe the distribution of damage clusters formed inside the cells immediately after irradiation has ceased ($t_{rep} = 0$, i.e. no damage has been repaired in the simulation). Figure (8.6) depicts the number of damage clusters that contain $k = 0, 1, 2, \cdots, n$ ionisations following exposure to increasing doses of high-LET radiation from our simulated data. The experimental fit $a\exp(-bx)$ is also performed in this data and is shown in Fig.(8.6), where x represents the ionisation clusters containing $k = 0, 1, 2, \ldots, n$ damages. Our analyses showed that the lethal damages in cells by increasing absorbed dose described by exponential decay with very good accuracy. We observe that the relative number of ionisation events increases with the absorbed dose as expected (directly observable with the increasing fit parameter a). The counts decrease exponentially at the same rate $\sim$0.63 (i.e. fit parameter b remains constant for all doses). We expect a linear increase in damage complexity from Chapter 7 where we found the yield of each damage to increase linearly with absorbed dose. Higher order values (corresponding to damages of higher complexity) also appear to grow with dose, implying a higher yield of complex damages with more than 3 individual strand breaks present - for example, in the histogram of the first dose only 6 bins are visible, whilst for dose eight, 10 bins are visible. Since lethal lesions are more likely to be complex in nature and the complexities increase linearly with dose (see Fig.(7.14)), we expect the variance of the damage complexity distribution in Fig.(8.6) to increase with dose.

Figure (8.7) depicts the distributions of lethal lesions (from our simulations) present in 530 cells for 500 repetitions of the survival fraction measurement from Chapter 4, at the 40^{th} hour of repair time, when the system has reached the equilibrium state $|f\rangle$. The PDF of the fractional and non-fractional Poisson processes are fit to each distribution. The computational method described in Chapter 6 is used to fit the fractional Poisson count model. To briefly recall, the fractional Poisson process PDF is computed using its integral representation defined as

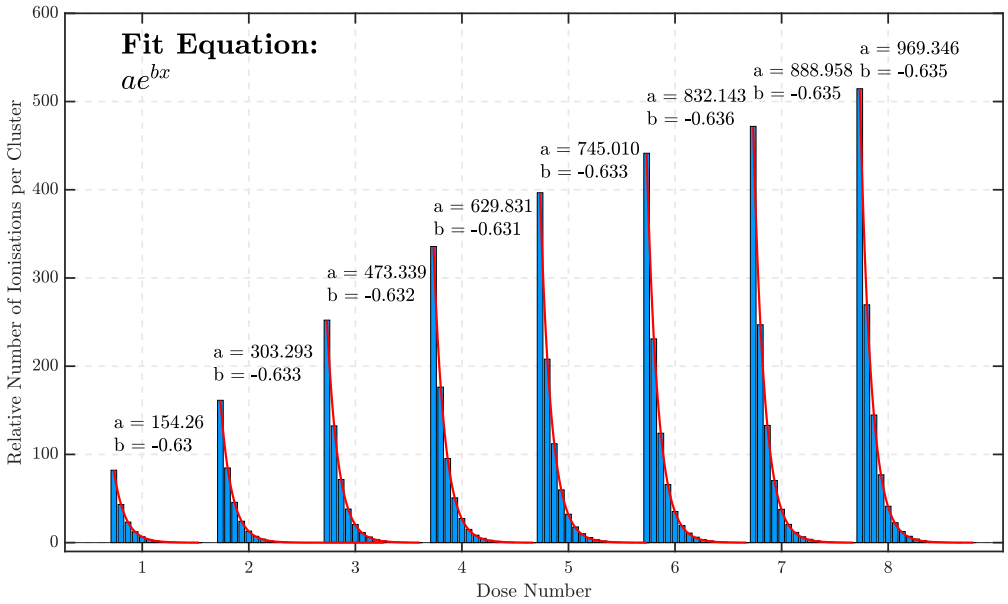


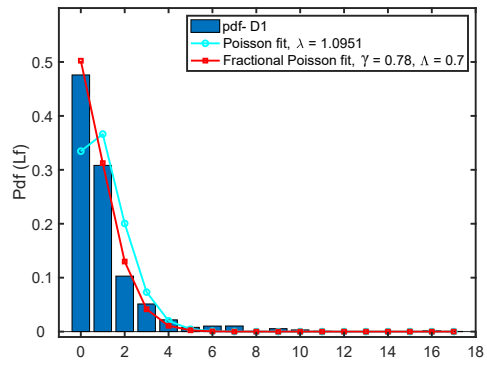
Figure 8.6: Histograms of DNA damage cluster complexities induced in 530 cells after exposure to a 1.6MeV (35keV/μm) proton beam at increasing doses.

$$P_k^{(\gamma)}(t) = \frac{\lambda^k t^{\gamma k}}{k!} \int_0^\infty dz \quad z^k e^{-z\lambda t^\gamma} M_\gamma(z) \quad , \tag{8.5}$$

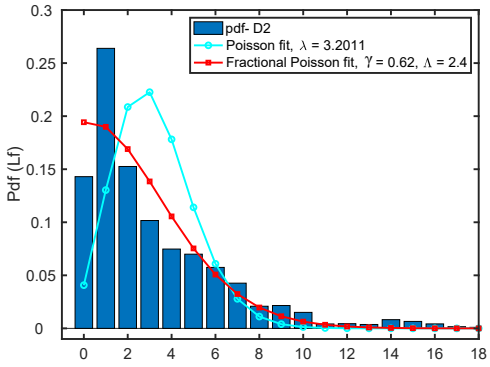
where $M_\gamma(z)$ is the Mainardi function at a given time t. Details of this fit procedure is given in Appendix I.

From observation of Fig.(8.7), the fPp (Eq.(8.5)) yields a far better agreement with data in comparison to the Pp PDF for all doses. The Poisson process consistently misrepresents the mean number of lethal lesions inside a single cell. The lowest dose (Fig.(8.7(a))) shows the best agreement with the Poisson process, however even in this case the fractional Poisson process provides the best fit to the data. This is expected because the Pp, i.e. LQ model describes low-LET data better as we have previously discussed (see our results for LET = 1.2keV/μm in Figs.(8.3(c)) and (8.3(d))). The Pp fit worsens with increasing dose such as in Fig.(8.7(h)) where the mean is very different between the two processes (22.41 for the Pp and 17.2 for the fPp). Furthermore, the fractionality parameter, $\gamma \approx 0.6$, which is also very far from 1 which further indicates a high degree of fractionality in the system. As shown in Chapter 6 and Figs.(8.3), the reason for the poor fit of the Pp to the data is the presence of overdispersion in the system (the variance exceeds the mean) in the high-LET radiation region.

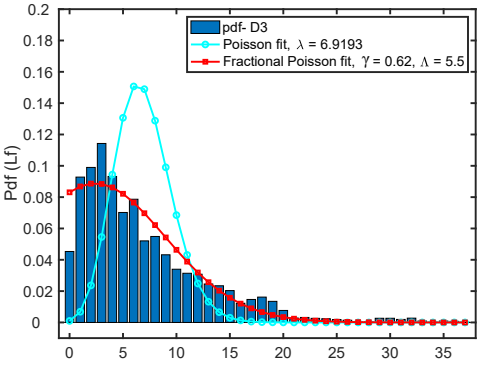
For consistency we have also performed fits of the Pp and fPp PDFs to normalised histograms for the number of lethal lesions per cell for high doses of low-LET radia-



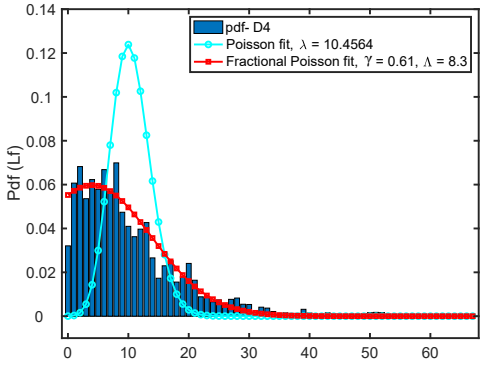
(a) 1.18Gy



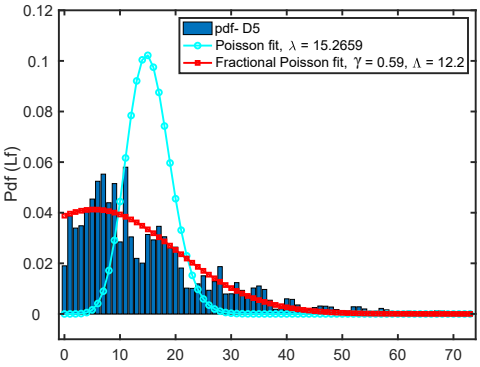
(b) 2.36Gy



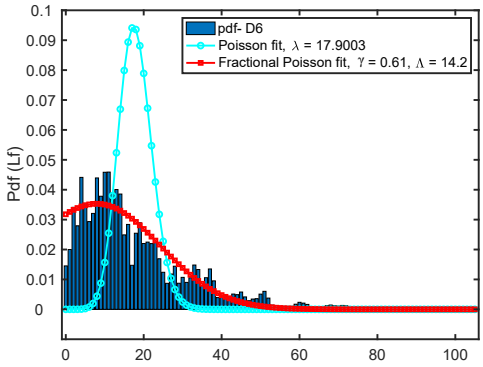
(c) 3.54Gy



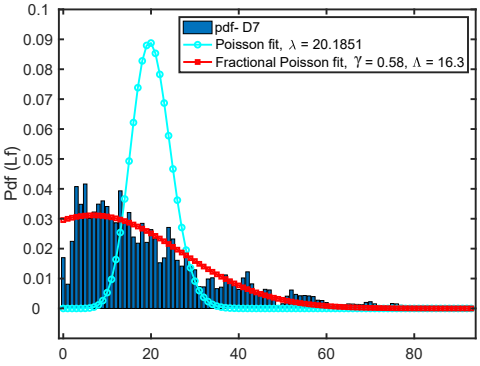
(d) 4.73Gy



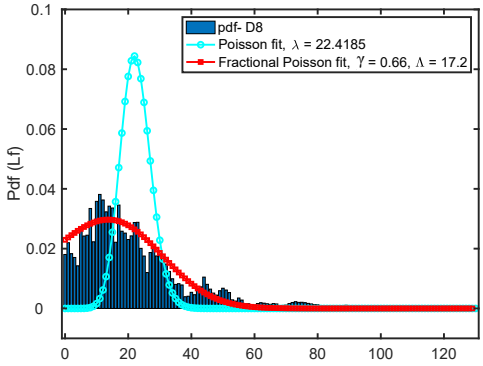
(e) 5.67Gy



(f) 6.39Gy



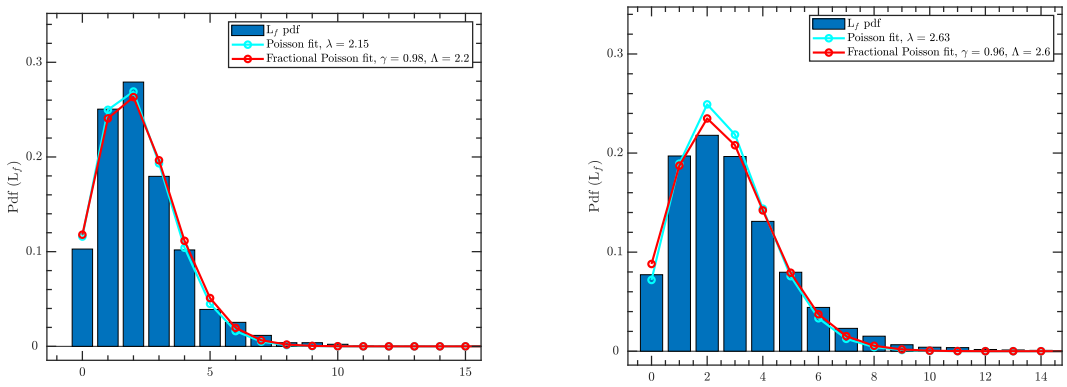
(g) 6.85Gy



(h) 7.34Gy

Figure 8.7: The distribution of lethal lesions in the cells after 40 hours of repair time induced by high doses of high-LET (35keV/μm) radiation from our simulations.

tion from our simulated results, which are depicted in Fig.(8.8). Here we note that due to the long computation time and instability in the numerical computation of the Mittag-Leffler function as $\gamma \to 1$, the comparison fits could not be performed for low doses of low-LET radiation in a timely manner, however this will be a focus of future work. For now, we observe a much closer agreement between the Pp and fPp models in Fig.(8.8) compared to all doses in Fig.(8.7). The reason for the closer agreement between the two models is reflected in the fractionality parameter, γ, which is approaching 1 ($\gamma = 0.96$ and 0.98), some of which are already in the Poisson process dominant region (i.e. the low-LET region).



(a) 6.40Gy **(b)** 7.30Gy

Figure 8.8: The distribution of lethal lesions in the cells 40 hours post-irradiation induced by high doses of low-LET (1.2keV/μm) radiation from our simulations.

Overall, we have observed strong evidence of fractionality in the high-LET region where the fPp model yields a substantially improved fit to the data compared to the Pp. Our observations support those made in the literature in that overdispersion is present under certain radiation exposure conditions [68, 69, 171] for proton radiation. Moreover, we observe close agreement between the Pp and fPp models for lethal lesion distributions under low-LET exposure conditions due to the fractionality parameter, γ, approaching 1 ($\gamma \to 1$) where the two processes become equivalent.

8.3 Our Proposed Model for Cell Survival

Our proposed model redefines the statistical assumption that the induction of lethal DNA damage as a result of ionising radiation, hence survival, obeys a Poisson process. It achieves this by describing the dose-response relation using a fractional Poisson process where the inter-arrival times of lethal damage produced on the DNA segment in each cell are no longer exponentially distributed, but instead with a "stretched" Mittag-Leffler (ML) distribution. A dose-time argument equivalent

to the expression in the LQ model is used to present the ML function in terms of absorbed dose D,

$$S(D) = E_\gamma[-(aD + bD^2)^\gamma] \quad . \tag{8.6}$$

We have shown in the previous section that the fPp model yields either a similar or highly improved fit to the distribution of lethal lesions inside the cells after reaching saturation (state $|f\rangle$). As discussed in Chapter 6, a poor fit to the lethal lesion distribution can result in a poor estimation of the survival probability, i.e. the zeroth order. In this section we focus of the zeroth order probability where we can test our model on available experimental survival data (Appendix D) and our simulated survival data (see Chapters 4 and 7). From our findings in the previous section we infer that the fPp model should yield a similar or improved fit to the survival data. The process of testing our proposed model on cell survival data is outlined in the flowchart of Fig.(8.9) below.

8.3.1 Testing Our Proposed FPp Model on Experimental Data

In this section we fit our new proposed fPp model to experimental data. Similar goodness-of-fit and hypothesis testing are performed on the fits as in Chapter 5. For clarity, the results corresponding to LET variance by energy and depth in the medium are considered separately and then compared. As a benchmark study, our results are also compared with some existing radiobiological models discussed in Chapter 2[2].

8.3.1.1 The Methods

In Chapter 6 we discussed in detail the methods used to compute the ML function, i.e. survival probability. We use an integral representation of [146] Corollary 1 to compute the probabilities for $k = 1, 2, \cdots, n$,

$$p_k^{(\gamma)}(t) = \frac{1}{k!}\,\lambda^k t^{k\gamma} \int_0^\infty dz\; z^k e^{-z\lambda t^\gamma} M_\gamma(z) \tag{8.7}$$

where the M-Wright function or Mainardi function is defined by

$$M_\gamma(z) = \frac{1}{\pi} \sum_{j=1}^{\infty} \frac{(-z)^{j-1}}{(j-1)!} \Gamma(\gamma j) \sin(\pi\gamma j) \quad . \tag{8.8}$$

[2]Namely the Repair-Misrepair (RMR) [31], Lethal-Potentially Lethal (LPL) [32] and Multiple Hit Multiple Target [34] models in addition to the LQ model.

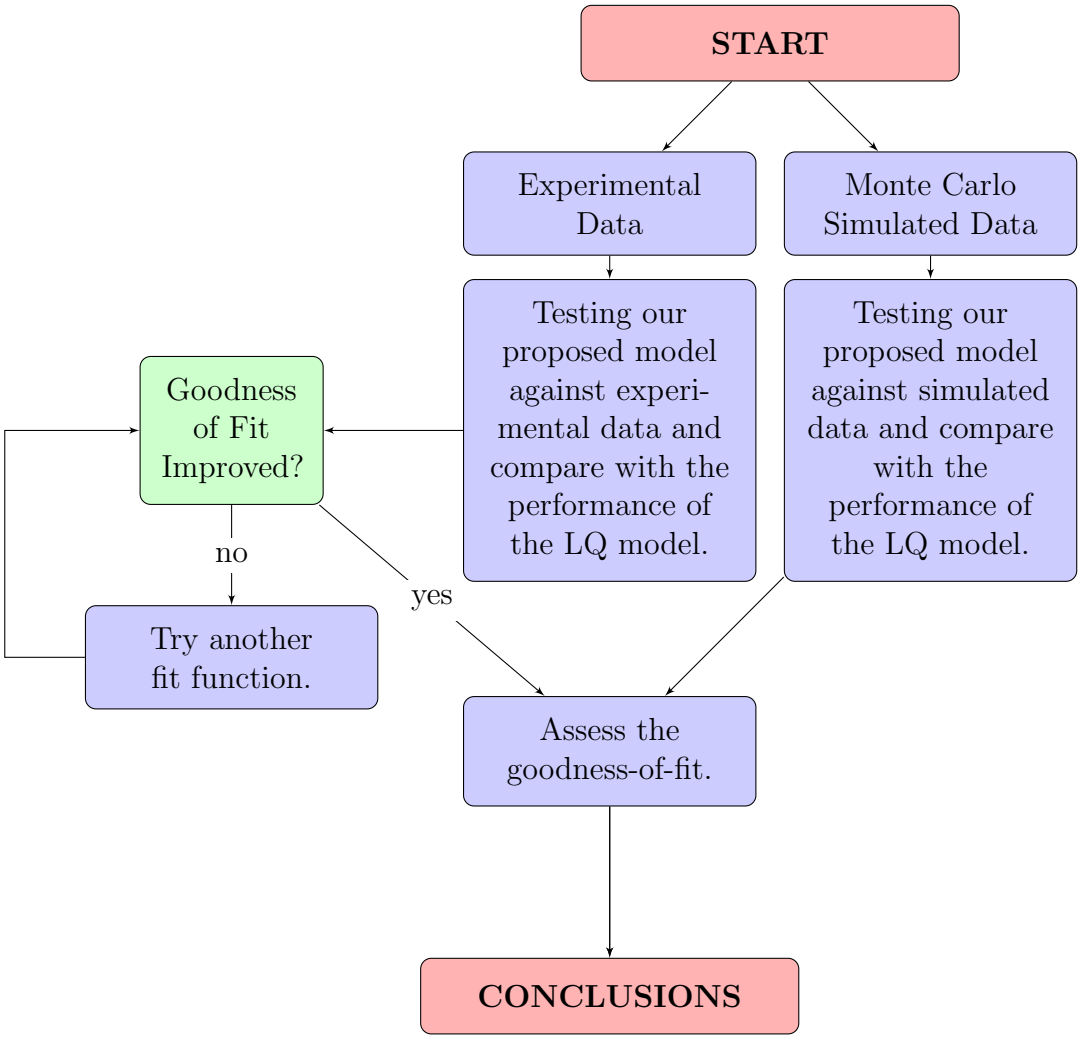


Figure 8.9: Flowchart describing how the proposed model will be tested.

As discussed in Chapter 6, the Mainardi function $M_\gamma(z)$ approaches a delta-distribution when $\gamma = 1$, making its numerical computation unstable. To observe the behavior of the function against the experimental data from Appendix D, we perform the fit using two methods :

1. Non-linear least squares method (NLLS)

2. A global optimisation brute force method using a search algorithm.

We apply a single fit, weighted by the standard error of the mean (SEM) of the measurements reported in the literature and unweighted using both the NLLS and search algorithm methods. In the NLLS method, the 95% confidence interval on each parameter is calculated. To calculate the parameter errors on our search algorithm method, we use Monte Carlo methods whereby random, Normally-distributed, small perturbations on the survival fractions are generated about the mean reported

in the literature for each dose point to which we fit the ML function. We repeat the fit 1,000 times to determine the mean and median parameters with errors. The latter method allows us to observe the behavior of the ML function parameters, particularly γ, for each data set which is discussed in more detail below.

We compare the performance of the fPp model in relation to the Pp (the LQ) model using goodness-of-fit and error (residual) analysis as we do in Chapter 5 for non-linear models on the survival data.

8.3.1.2 The Difficulties

The instability of the ML function computation as $\gamma \to 1$ meant that we could not fit all of the experimental data in Appendix D in a timely manner. Therefore, we have performed the fits using the methods described above for select experimental studies including Belli et al. [26] (our benchmark study), Bettega et al. [61] and Belli et al. [60].

The fitting methods described above allow us to observe the behavior of the fractional parameter, γ, under different LET. However, the computation time is large when performing an adequate number of repetitions using Monte Carlo methods. To observe the behavior of γ in the number of repetitions, n, we performed 1,000 fits for the data considered in the current study and 10,000 fits on the data of Belli et al. [26], our benchmark experimental study. The results are discussed later in this chapter.

Furthermore, the errors on the experimental data are very small and lacking information about how they are generated in the literature, which can impact on the χ^2 value. At times the model fits the data well, however the small measurement errors result in a large χ^2 value. In such instances, the χ^2 test does not adequately reflect the fit quality because it will penalise the model heavily if it does not pass through each point within the error bars. However, relative comparisons of χ^2 different models are still meaningful in choosing which model works better to explain the data. Moreover, one must compare other gof metrics including the RMSE and R^2. To note here, the experimental standard errors reported in the literature were used to generate weights for the least squares fit.

8.3.1.3 FPp on Energy-Varied LET

In Chapter 5 we performed a rigorous statistical evaluation of the LQ model against experimental data. Here we will extend this to test our new model against the same experimental data for a comparison of its performance against the LQ model.

To continue our benchmark study of Belli et al. [26], we will discuss the fPp and Pp fits to the data in detail. However, we will also discuss fits to the data of Bettega et al. [61] and Belli et al. [60] to compare our results across different cell lines. All fit results and gof metrics for the fPp and Pp models against our simulated, energy-varied LET data are presented in Table (K.1) of Appendix K.

Figures (8.10)–(8.12) show the survival fraction versus absorbed dose and performed fits to the data of Belli et al. [26] using the Pp (LQ) and fPp (Mittag-Leffler) models. Here we compare Figs.(5.3)–(5.5) where the LQ model is fit to the data of Belli et al. [26] with our ML model fitted to the same data. The fits are performed using the weighted NLLS method with the errors on the fit parameters corresponding to the 95% confidence intervals.

Let us analyse the fits systematically. Fig.(8.10(a)) corresponds to the response of the V79 Chinese Hamster cell line to a proton beam of energy 5.01MeV (7.7keV/μm). We observe that the Pp and fPp fits are in excellent agreement with perhaps a slightly improved fit from the Pp model (Pp: $\chi^2 = 0.420$, $R^2 = 0.995$ and fPp: $\chi^2 = 0.484$, $R^2 = 0.994$). The reason is clear upon observation of the fit parameters. The fractional fit parameter, γ, is close to one (0.98) meaning that the fPp and Pp are nearly equivalent. The α and β parameters of the Pp model (LQ) are similar to a and b of the fPp model, showing further agreement. Recalling the lethal lesion distribution in the low-LET limit of our simulated data (Fig.(8.7)), the result in consistent with our expectation that the fPp will approach the Pp for low-LET radiation.

In Fig.(8.11) we observe that at an LET of 11 and 20keV/μm the fractional parameter, γ, is $\sim$0.9. The Pp and fPp processes are therefore not exactly equivalent in this LET region compared to the low-LET region of Fig.(8.10(a)) but are in the fractional Poisson region instead. The gof metrics indicate that the fit is improved for the fPp model compared to the Pp (fPp: $\chi^2 = 0.449$, 0.315 and Pp: $\chi^2 = 0.863$, 0.366 for 11 and 20keV/μm, respectively). This is consistent with our expectation that for LET >10keV/μm we observe deviations from the Pp as observed in Section (8.2).

In Fig.(8.11(b)) we present fPp and Pp fits to the 30.5keV/μm dose response curve of Belli et al. [26]. Upon observation of the gof metrics, the fPp yields a consistently better fit than the Pp ($\chi^2 = 1.719$ for the Pp and 0.078 for fPp). With a fractionality parameter of $\gamma = 0.99$, this data set is inside the Poisson region. In fact this is the LET where γ is closest to 1 out of all the data in Belli et al. [26] and is therefore the closest to the Poisson region. One can see why by observing the fPp and Pp fits in Fig.(8.11(b)), where β from the Pp model is only just negative and therefore the dose-time relationship is convex. In Chapter 5 we concluded that when the dose-time relation is convex, the LQ (Pp) model does not fit the data as well. Using the fPp model, the dose-time relationship is still concave despite being

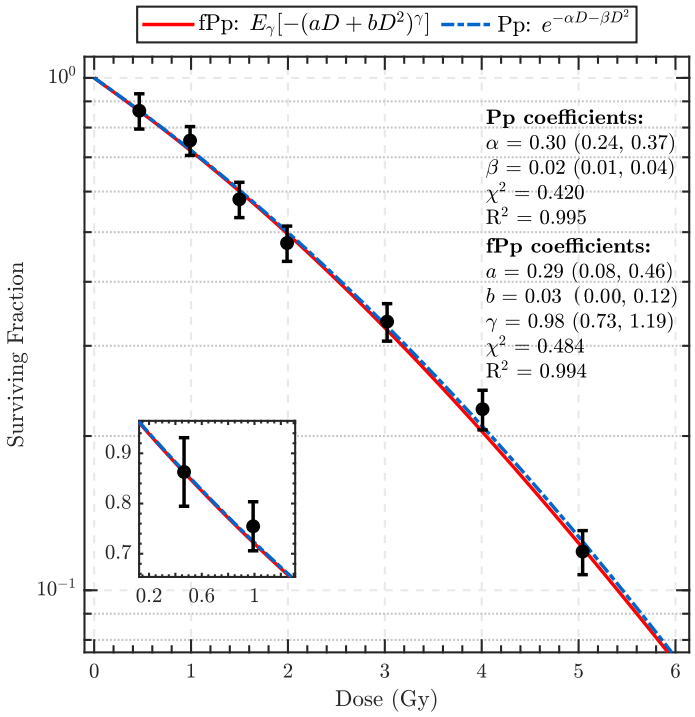
in the Poisson region.

Finally, Figs.(8.12(a)) and (8.12(b)) depict the fPp and Pp fits to the two highest LET data sets we possess from the experimental data, 34.6keV/μm and 37.8keV/μm. In both cases, a high degree of fractionality is observed ($\gamma = 0.75$) with an improved fit compared with the Pp model ($\chi^2 = 1.766$ and 1.746 for the 34.6keV/μm and 37.8keV/μm Pp fits, respectively). We recall that it was this region where the LQ model, hence the Pp, was found to least adequately explain the data. Linearised LQ fits showed that the logarithmic survival fraction as a function of dose began to decrease. Therefore, we expect the fPp model to yield substantially better fits than the Pp model, which is exactly what we observe here. This is further evidenced by the difference in fit quality to the lethal lesion distributions from our simulated data in the high-LET region, where the Pp model did not properly estimate the zeroth order (survival) probability.

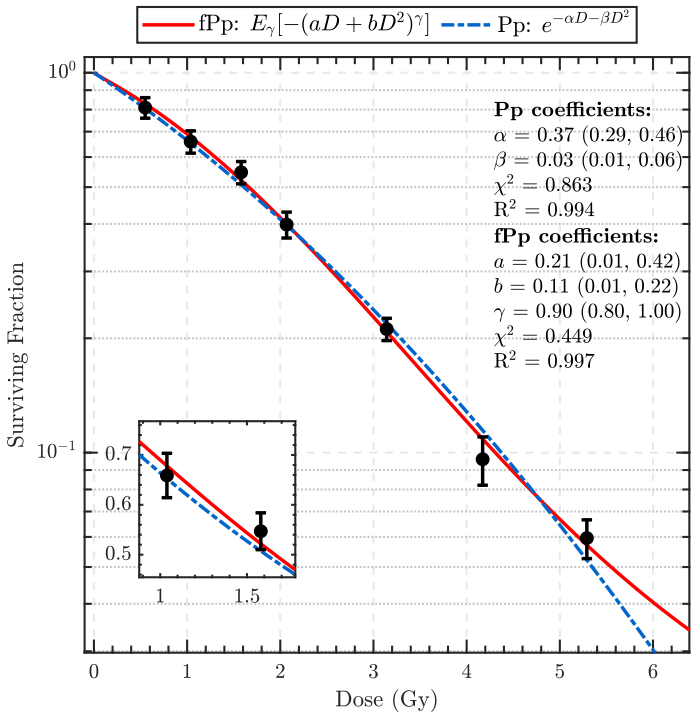
Using Table (K.1), we can compare the fits to Bettega et al. [61] and Belli et al. [60]. Both studies utilise cell lines of human and rodent origin (different from the V79 cell). A similar trend to Belli et al. [26] is observed in Bettega et al. [61] where γ decreases as LET increases. The highest LET considered in Bettega et al. [61] is 33.2keV/μm, which results in a fractionality parameter of $\gamma = 0.805$ before gradually increasing to 0.969 at LET 11keV/μm.

Belli et al. [26] and Bettega et al. [61] contain a data set with an LET of 11kev/μm which both result in high, but slightly different γ values (0.909 was reported for Belli et al. [26] compared to 0.969). Nevertheless, all fPp fits from both data sets yield a significantly better gof compared to the Pp model. It is plausible to suggest that there are additional factors that influence the degree of fractionality in a system aside from LET, such as the radiation-sensitivity and availability of certain repair processes for different cell lines.

The degree of fractionality in the system is related to the inter-arrival times between the induction of DNA strand breaks and hence the final yield of lethal lesions in per cell. In Chapter 2 we introduced a number of parameters that influence the production of lethal lesions inside the cell. Biological environment parameters within the cell such as its cycle phase, oxygen content and DNA genome length can influence a cell's response to radiation and hence its radio-sensitivity. We assume that the survival fraction is measured after a sufficient repair time has elapsed in all of the experimental studies considered here (i.e. they have reached equilibrium). Therefore, it is plausible to suggest that the dose-response curves considered here are influenced by many biological parameters. Furthermore, it is not stated in the literature what the distribution of cells in each cycle phase actually is when they are irradiated, therefore we assume they are asynchronous. By taking cellular biology conditions into account, we provide a possible explanation for the difference in γ

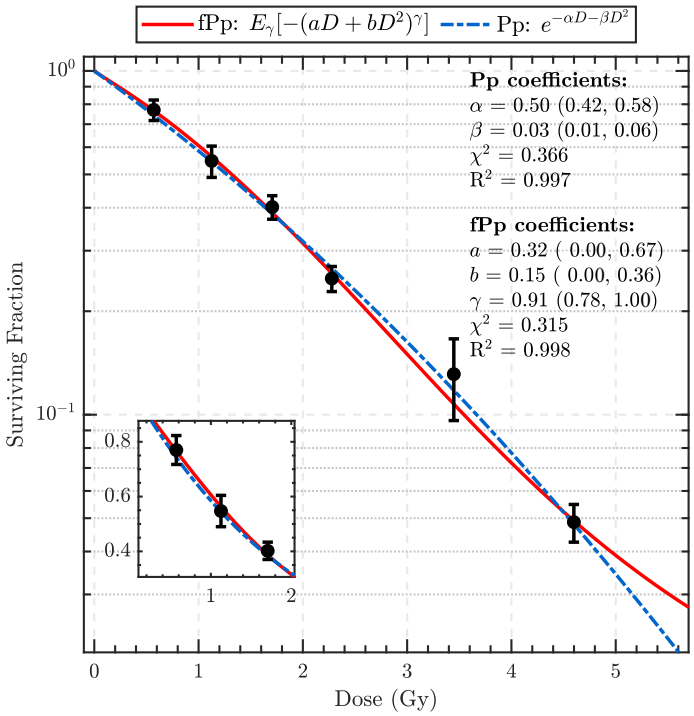


(a) Belli et al. (1998) [26]: 5.01MeV protons 7.70keV/μm incident on V79 cells.

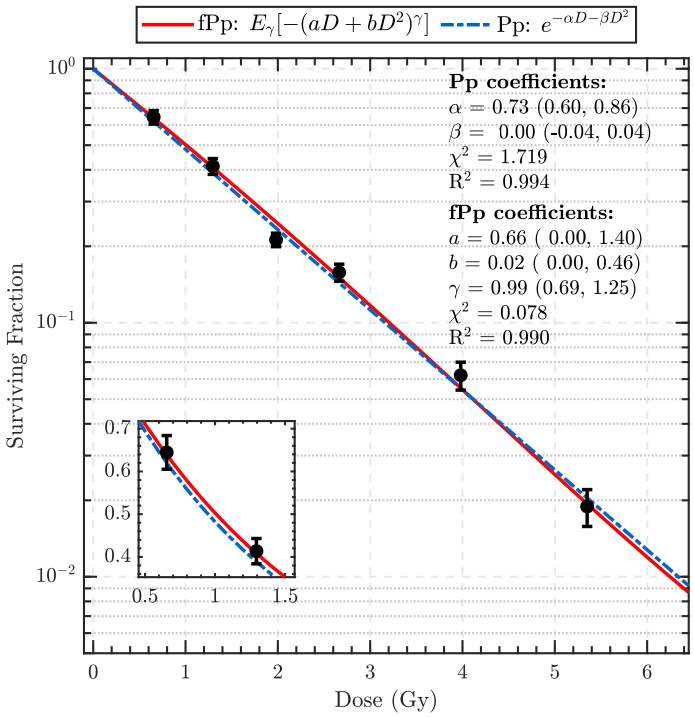


(b) Belli et al. (1998) [26]: 3.20MeV protons 11.0keV/μm incident on V79 cells.

Figure 8.10: Belli et al. [26] fPp and Pp model tests using a single least squares fit and global search algorithm weighted by the SEM (error bars).

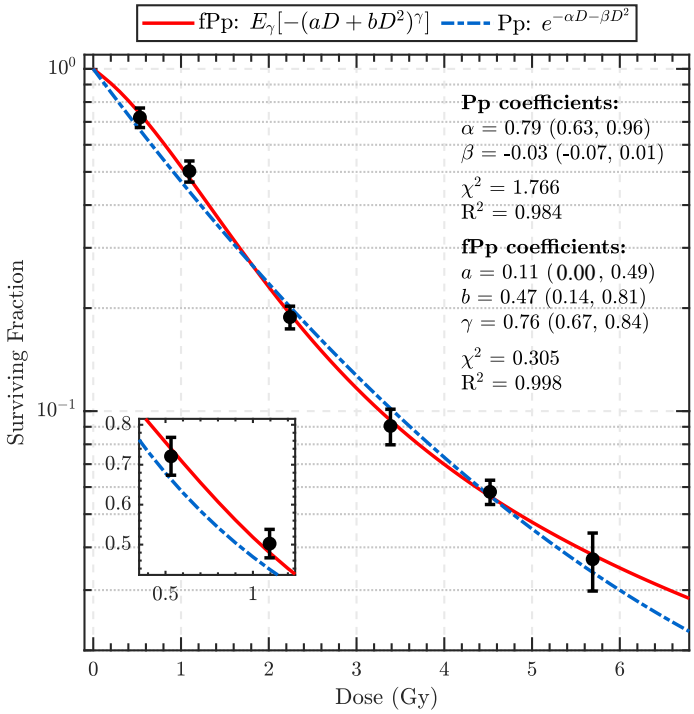


(a) Belli et al. (1998) [26]: 1.41MeV protons 20.0 keV/μm incident on V79 cells.

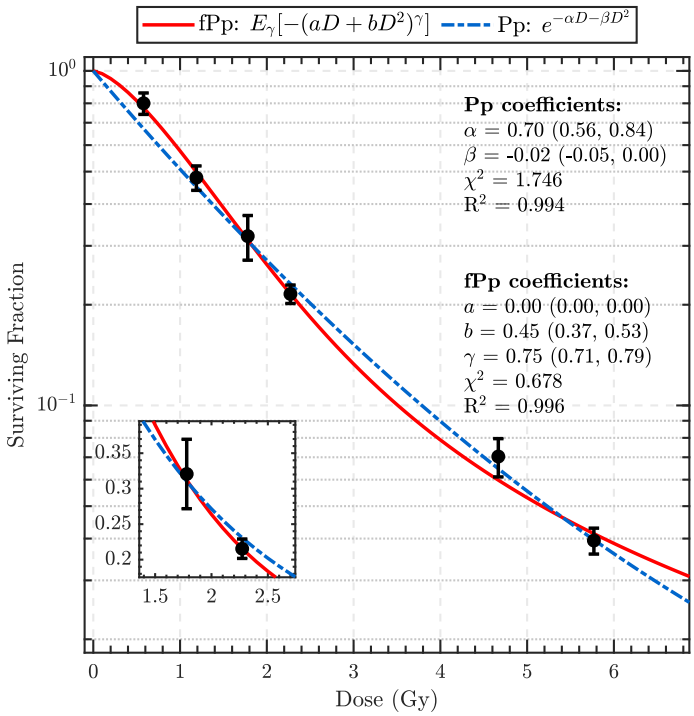


(b) Belli et al. (1998) [26]: 0.72MeV protons 30.5keV/μm incident on V79 cells.

Figure 8.11: Belli et al. [26] fPp and Pp model tests using a single least squares fit and global search algorithm weighted by the SEM (error bars).



(a) Belli et al. (1998) [26]: 0.64 MeV protons 34.6 keV/μm incident on V79 cells.



(b) Belli et al. (1998) [26]: 0.57 MeV protons 37.8 keV/μm incident on V79 cells.

Figure 8.12: Belli et al. [26] fPp and Pp model tests using a single least squares fit and global search algorithm weighted by the SEM (error bars).

for the data of Belli et al. [26] and Bettega et al. [61] under similar exposure conditions and different cell lines. The V79 and C3H10T12 cell lines likely have different radiation-sensitivities which will influence the final lethal lesion distribution per cell and hence the survival fraction.

The fits to the data of Belli et al. [60] yield a similar trend in γ compared to Belli et al. [26] and Bettega et al. [61]. In the mid-LET range γ is further from 1, suggesting that DNA damage is induced via a fractional Poisson process. As the LET increases to $\sim$30keV/μm, γ approaches 1 again thus implying that the system obeys a Poisson process where the dose-time relation moves from a concave to a convex function (see Fig.(8.11(b))). If we recall our discussions on RBE from Chapter 5, we know that the dose-response can differ across cell lines due to differing biological conditions within the cells. Moreover, some biological conditions influence the lethal lesion yield in the cell when it reaches its equilibrium state ($\langle f \rangle$) thus affecting the survival probability and RBE. We expect that the differences in cell response across cell lines will be reflected in the fit of our fPp model, namely the value of γ. Therefore, we can infer that the turning point of the M10, SCC25, SQ20B and HF19 human cell lines experience a similar change in behavior when exposed to proton radiation of LET 30keV/μm.

Table (K.3) of Appendix K shows the results of hypothesis testing for non-linear regression described in Appendix E. Recall from Chapter 5 that the Pp LQ fits had a high failure rate when fit to the same experimental data using the least squares method (the results from the LQ fits are presented in Table(G.2)). We observe that when using the Mittag-Leffler model, the pass rate is improved across the data of Belli et al. [26,60] and Bettega et al [61]. A notable improvement is observed in the fits to Belli et al. [26], our benchmark study, in that they all pass the tests for normality and homoscedasticity. The LQ model fits from Chapter 5 violate the assumption of normally distributed residual errors or homoscedasticity for LET >30.5keV/μm. Therefore, hypothesis testing supports our theory that the fPp model is superior to the Pp, particularly in the high-LET region. Some of the fPp fits that fail the normality test, however this can be due to the presence of outliers in the data. Nevertheless, many data sets perform better under hypothesis testing for the fPp compared to the Pp.

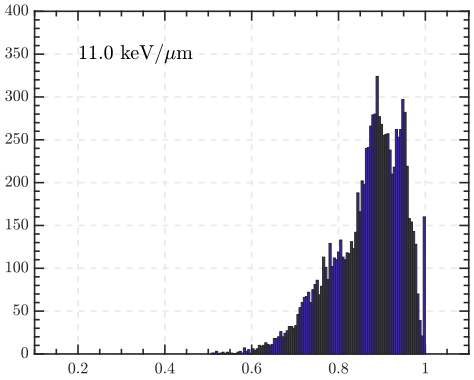
In Table (K.1) we observe that the least squares and optimisation fitting methods generally result in a similar γ, a and b parameters of our Mittag-Leffler model. Recalling our discussion from Chapter 6, the numerical computation of the Mittag-Leffler model can be unstable which results in slightly different fit parameters. To investigate the extent to which γ deviates in the data of Belli et al. [26], we used the Monte Carlo method described at the beginning of this section. Briefly, because we do not have any knowledge of how the errors in the data from the literature are calculated or handled, we use Monte Carlo methods to generate uncertainties on

each parameter in our model. This is achieved by introducing Normally-distributed, small perturbations on the measured survival fractions from the literature according to the measurement error (SEM). We used this method to perform 10,000 fits of our Mittag-Leffler model to the data allowing us to observe the distribution of the fractional parameter, γ.

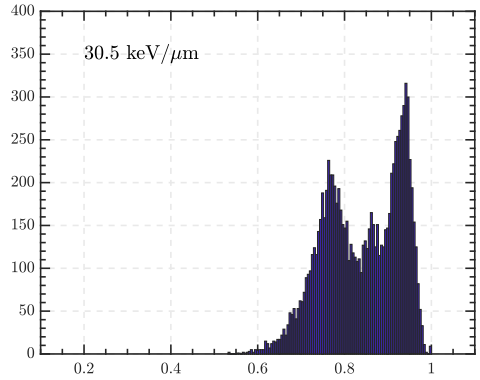
The distribution of fractional parameter γ from 10,000 unweighted fPp fits to the data of Belli et al. [26] is presented in Fig.(8.13) for LET 11, 30.5, 36.4 and 37.8keV/μm. Fig.(8.13(a)) depicts the distribution from the fPp fits for an LET of 11keV/μm. As discussed above, the least squares and optimisation methods reported a γ of 0.9, which is in agreement with the distribution in Fig.(8.13(a)) where the peak occurs at $\sim$0.9. If we compare Fig.(8.13(a)) with Figs.(8.13(b)), (8.13(c)) and (8.13(d)), we observe two distinct peaks at $\gamma \approx 0.75$ and 0.95, suggesting two different convergence points of the solutions. This is despite the NLLS and search algorithms resulting in similar parameters for all the experimental data (see Appendix K).

It is not entirely clear yet why the clustering effect occurs in the high-LET region in particular, however we are currently exploring this. We recall that in experiment we are considering the full picture with all physical, chemical and biological processes present. It is likely that 10,000 Monte Carlo repetitions is not adequate at explaining the γ parameter distribution, therefore we require more repetitions to observe the region between the two peaks increase.

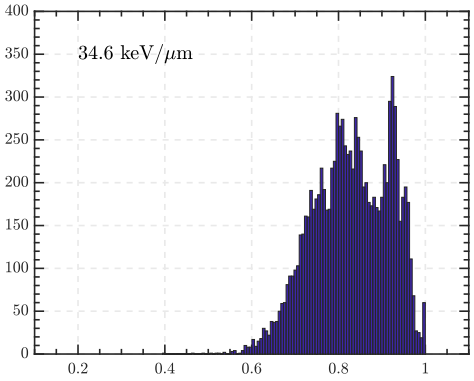
To conclude, the fPp (Mittag-Leffler) model consistently yields a similar (in the low-LET region) and a more superior fit (in the mid to high-LET region) to the experimental data compared to the Pp model. The reason is clear by observing Fig.(8.14) depicts the fPp fits to the data of Belli et al. [26] and shows how the behavior changes in the high dose region for high-LET. The decrease in cell killing efficiency in the high-LET region is the result of DNA damage saturation effects and the fPp model is a much better fit to data with this behavior. The improvement in the fit was expected from our results in Section (8.2), where the distribution of lethal lesions per cell was better explained by the ML distribution compared to the Poisson due to overdispersion effects. Our observations are consistent across different studies and conditions where the Mittag-Leffler parameter, γ, is in the high 0.9 region for low-LET conditions where the fPp is in better agreement with the Pp. Then as the LET increases, the γ value decreases where the fPp moves away from the Pp.



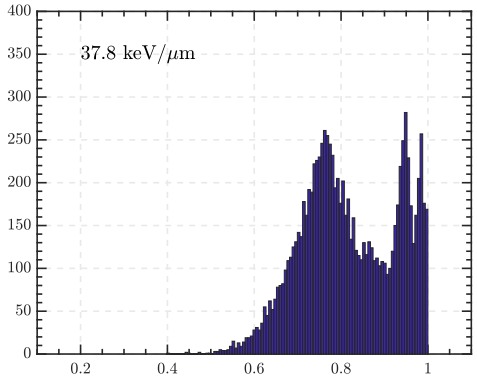
(a) Belli et al. (1998) [26]: 3.2MeV protons 11.0keV/μm incident on V79 cells.



(b) Belli et al. (1998) [26]: 0.72MeV protons 30.5keV/μm incident on V79 cells.



(c) Belli et al. (1998) [26]: 0.64MeV protons 34.6keV/μm incident on V79 cells.



(d) Belli et al. (1998) [26]: 0.57MeV protons 37.8keV/μm incident on V79 cells.

Figure 8.13: Fractional parameter, γ, distribution for 10,000 Monte Carlo fits to the data of Belli et al. [26].

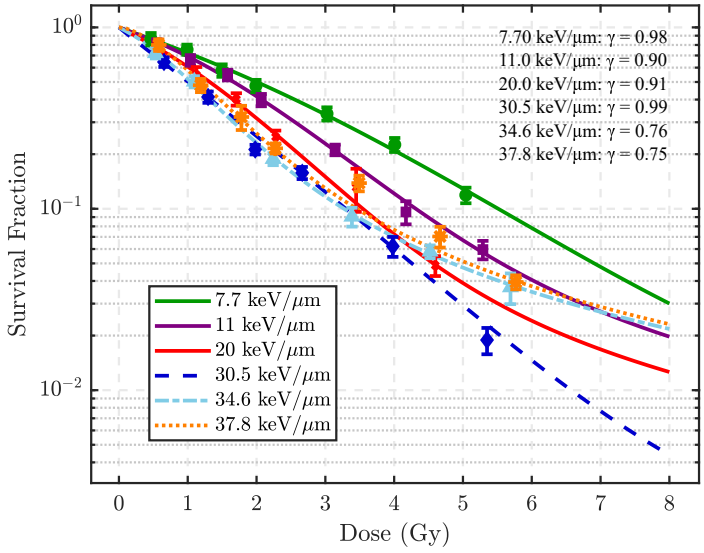


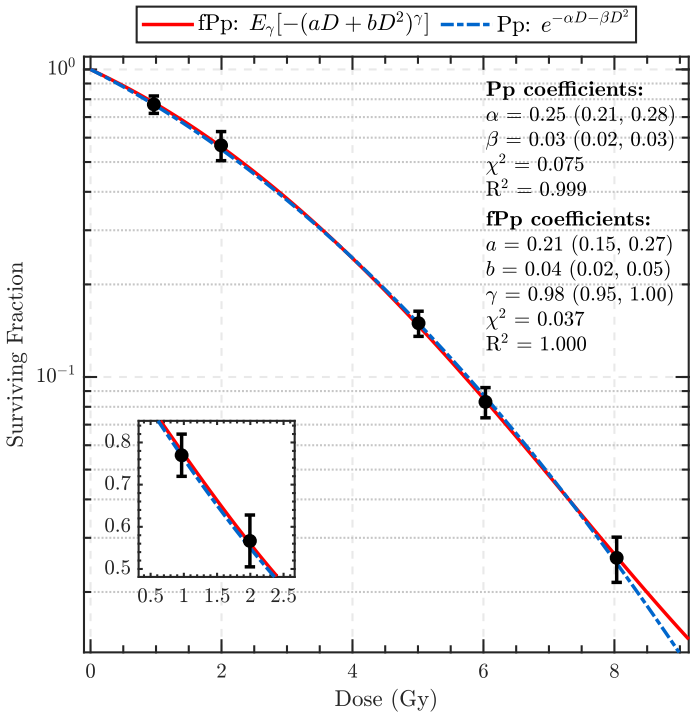
Figure 8.14: FPp fits against the data of Belli et al. [26].

8.3.1.4 FPp on Depth-Varied LET

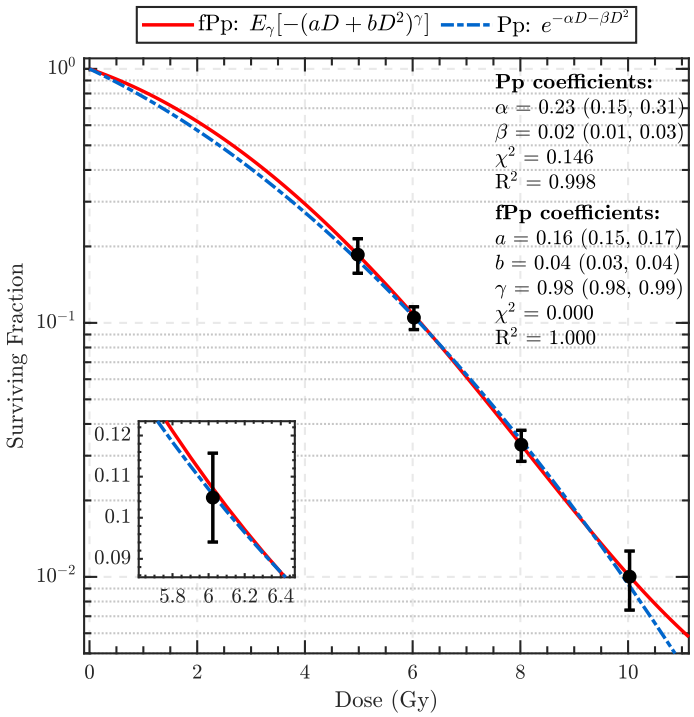
The experimental data considered in the current study is sorted into one of two categories, (i) LET varied by beam energy and (ii) LET varied by the depth in the Bragg Peak. In this section we discuss the performance of the fPp model for the experimental depth studies of Guan et al. [27], Howard et al. [123] and Wouters et al. [38]. We note that most depth studies involve a LET range much smaller than those considered in the energy-varied LET case. The three studies we will analyse correspond to LET ranges of 10.5 to 19keV/μm, 0.99 to 7.29keV/μm and 1.03 to 4.74keV/μm for Guan et al. [27], Howard et al. [123] and Wouters et al. [38], respectively. Some of the fPp and Pp fits are presented in Figs.(8.15) and (8.16) whilst the fit parameters and gof metrics for fits to the data of each study are presented in Table (K.1) of Appendix K. All the studies considered here measure survival in the proximal (shallow depths), inter-peak and distal (after the Bragg Peak) regions for pristine and SOBPs for energies 71 to 160MeV and 160 to 230MeV, respectively.

Firstly, we observe that the fractionality parameter γ is consistently in the high 0.9 region (i.e. close to 1) for all fits to the data of Howard et al. [123] and Wouters et al. [38] ($\gamma > 0.95$) in Table (K.1) since all this data belongs to the low-LET region. The Pp and fPp fits from Fig.(8.15(a)) are in almost perfect agreement with each other, with a fractional parameter of $\gamma = 0.98$, i.e. the fPp has approached the Pp region. Additionally, the gof metrics are also similar between the Pp and fPp models, thus further evidencing the presence of the Pp in the small LET region. Figures (8.15(b)), (8.16(a)) and (8.16(b)) visualise the close agreement between the Pp and fPp for the data of Howard et al. [123]. The gof metrics indicate an extremely good fit from both models with $\chi^2 < 1$ and $R^2 \approx 0.99$ in all depth varied cases from Howard et al. [123] and Wouters et al. [38].

FPp fits to the data of Guan et al. [27] in Table (K.1) also give γ values in the high 0.9 region, however they do not quite follow the same trend as that of Howard et al. [123] and Wouters et al. [38]. The fractionality parameter, γ, increases between LET values 10 and 17.7keV/μm ($0.95 \rightarrow 0.97$) before decreasing to 0.94 at 19keV/μm. We expect a decrease in γ as LET increases, however for LET 15.2 and 17.7keV/μm, this trend is violated. Guan et al. [27] reports that the highest three LET in this investigation occurs after the Bragg Peak, whilst the 10.5keV/μm LET occurs inside the peak. The mechanics of proton energy deposition after the Bragg Peak is likely different to the proximal and inter-peak regions in that proton propagate to distal depth less frequently. Many of the energy depositions likely occur from secondary particles produced inside the Bragg Peak, which propagate to the distal region. It is difficult to interpret this without analysing similar studies, however we consider a series 20MeV pristine Bragg Peak depth-varied dose-response curves in our simulation study, which yields a higher LET range than Guan et al. [27] and will be analysed in the following section.

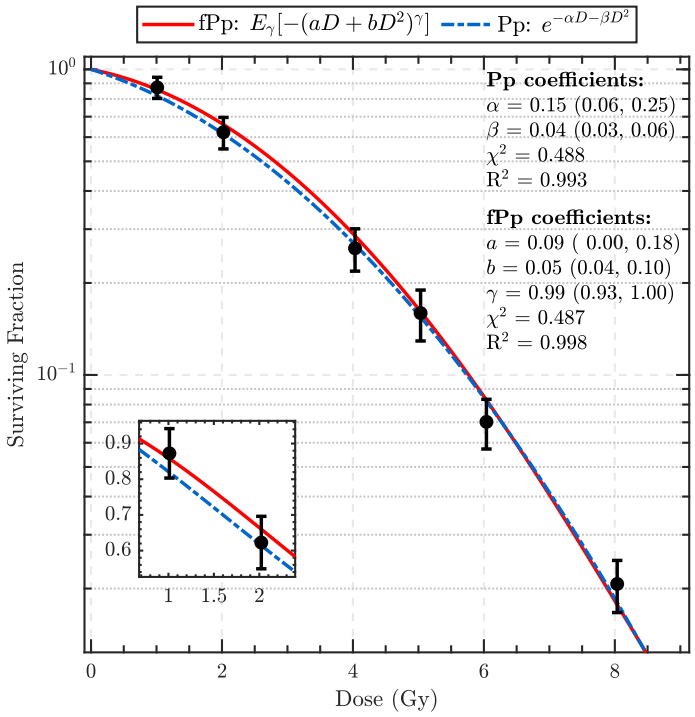


(a) Howard et al. (2018) [123]: 160MeV protons 0.99keV/μm incident on A549 cells.

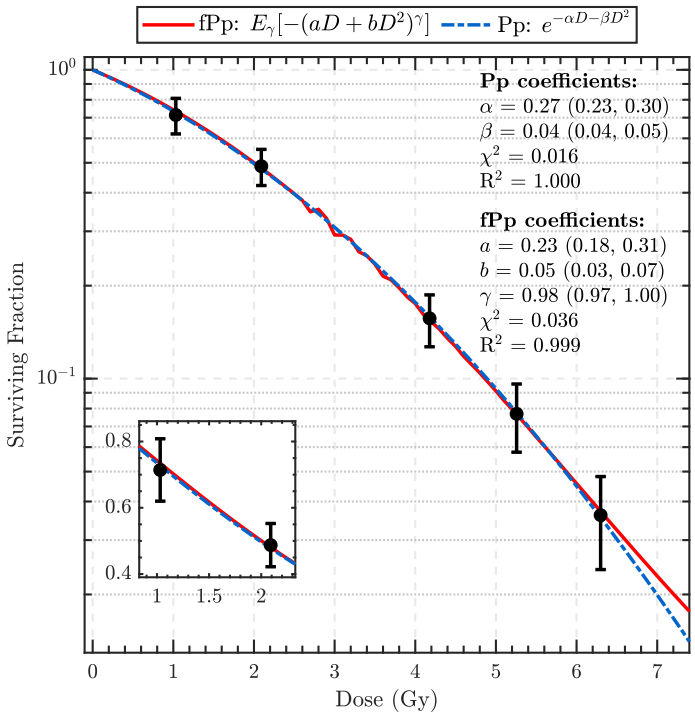


(b) Howard et al. (2018) [123]: 160MeV protons 2.26keV/μm incident on A549 cells.

Figure 8.15: Howard et al. [123] fPp and Pp model tests.



(a) Howard et al. (2018) [123]: 160MeV protons 4.19keV/μm incident on A549 cells.

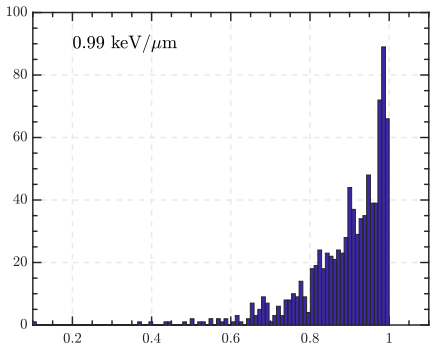


(b) Howard et al. (2018) [26]: 160MeV protons 7.29keV/μm incident on A549 cells.

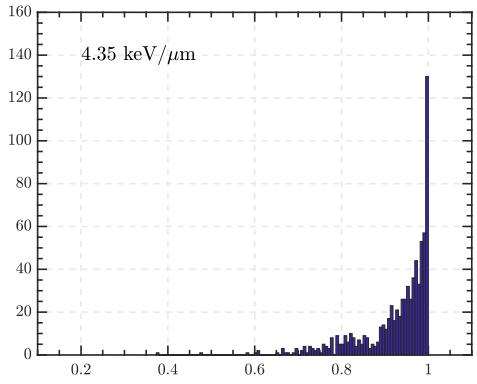
Figure 8.16: Howard et al. [123] fPp and Pp model tests.

The superior goodness-of-fit of the fPp model in comparison to the Pp further evidences our observations of fPp and Pp fits to the low-LET lethal lesions distributions from our simulated data in the previous section (see Fig.(8.8(b))). From the lethal lesion endpoint we observe the two processes to agree and the fractionality parameter $\gamma \approx 0.96$. We have shown that for the very low-LET region the Poisson process still yields an adequate approximation to the experimental data.

Similar to the previous section we observe close agreement in the parameters of the fPp using the least squares and optimisation fitting techniques, indicating that our fits are stable, reliable and reproducible. To compare the behavior of γ in the low-LET region, we present Fig.(8.17), which depicts the distribution of fPp parameter, γ for fits to the data of Howard et al. [123] in the proximal and Bragg Peak regions after 1,000 Monte Carlo fits. We acknowledge that more repetitions were performed in the previous section on the data of Belli et al. [26], nevertheless we have presented the low-LET γ parameter distributions for comparison. First, we do not observe the clustering effect observed at high-LET in Fig.(8.16(b)). Many of the γ estimates are > 0.9 indicating the Pp is a good approximation for low-LET, which is consistent with our observations thus far in this chapter.



(a) Howard et al. (2018) [123]: 160MeV protons 0.99keV/μm incident on CHO cells.



(b) Howard et al. (2018) [26]: 71MeV protons 4.35keV/μm incident on CHO cells.

Figure 8.17: Fractional parameter, γ, distribution for 1,000 fits from the Monte Carlo method to the data of Howard et al. [123].

Overall, most depth studies involve a small LET range and therefore making the results consistent with our theory in that the Pp is still applicable for small LET values. Compared to the energy-varied LET case, the LETs in the current section are still small enough such that the equidispersion condition of the Poisson process holds. Moreover, the fits in this section demonstrate the robustness of the fPp model in that it is capable of representing the data that can be explained using a Poisson process as the fractional parameter, γ, approaches one.

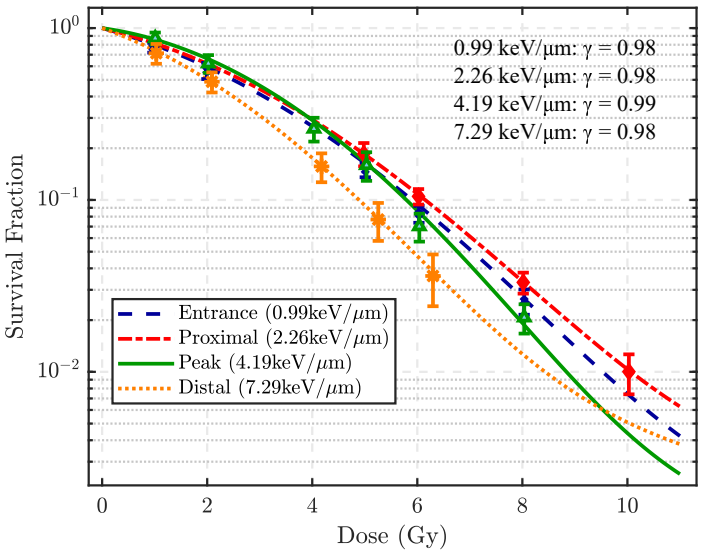
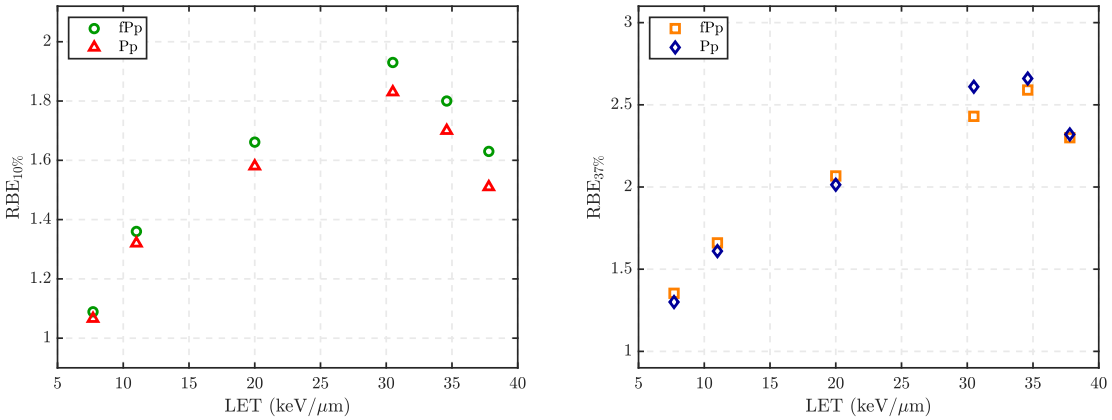


Figure 8.18: FPp model fits to dose-response curves at different positions in a 160MeV pristine Bragg Peak using the data of Howard et al. [123].

8.3.1.5 Predicting RBE with the fPp Model

We have discussed the impact of a superior dose-response curve fit to the prediction of RBE in Chapter 5. In this section we compare our previous RBE predictions with those made by the fPp model. Due to time constraints we could not derive an analytic expression for the RBE as a function of proton dose from the Mittag-Leffler function, however this is possible with the fractional survival probability function and we will leave this for future studies. Regardless, we can still observe the RBE as a function of LET, depth in the Bragg Peak/SOBP and across different cell lines.

Figure (8.19) depicts the RBE predictions at 10% and 37% survival for the data of Belli et al. [26] using the fPp and Pp model fits presented in Tables (K.1) and (F.4), both of which correspond to a single NLLS fit, weighted by the SEM. In the low-LET limit each model yields a close RBE prediction. As the LET increases, the discrepancies between the Pp and fPp begin to increase.



(a) 10% survival **(b)** 37% survival

Figure 8.19: RBE comparison at 10% and 37% survival the fPp and Pp models using fits to the data from Belli et al. [26].

At 10% survival the disagreement in RBE between the fPp and Pp models is the most pronounced in the mid to high-LET region. The turning point occurs at the same LET (30.5keV/μm), however, the fPp predicted RBE is much higher. Recalling that a 10% survival fraction is likely achieved with doses >3Gy and that the LQ model does not perform well in the high dose region [39], we expect to see a more accurate prediction from the fPp model which yields a better fit. At 37% survival (low dose region), the RBE predictions using the Pp and fPp models are in good agreement and the discrepancies are smaller than for the 10% survival RBE.The reason for this is shown in Fig.(8.20), which depicts the fits to the 34.6keV/μm data set of Belli et al. [26] on a logarithmic scale. At 37% survival the curves corresponding to the Pp and fPp are nearly overlapping, whereas at 10% survival the Pp predicts a slightly lower RBE, which explains why the Pp model predicts a smaller RBE compared to the fPp. This shows the extent to which small deviations in the dose-response curve can lead to substantially different RBE predictions.

In Chapter 5 we observed the RBE trend for the data of Belli et al. [26] to increase until 30.5keV/μm at 10% survival. The peak represents an "effective" LET where values after the peak result in a damage saturation or "overkill" effect [172]. We observe the same trend in Fig.(8.19), which demonstrates that the fPp model can correctly predict trends in RBE as a function of LET.

In Chapter 5 the RBE was observed to increase as a function of depth - and inherently LET. The fPp model predicts similar behavior as depicted in Fig.(8.19). The resultant RBE predictions made between the Pp and fPp models in Fig.(8.21) are closer in the low-LET region (at shallow depths in the Bragg Peak), showing more scatter at larger depths (inside the Bragg Peak) - another observation made in Chapter 5. The fPp also predicts the same trend of an increased RBE at low doses

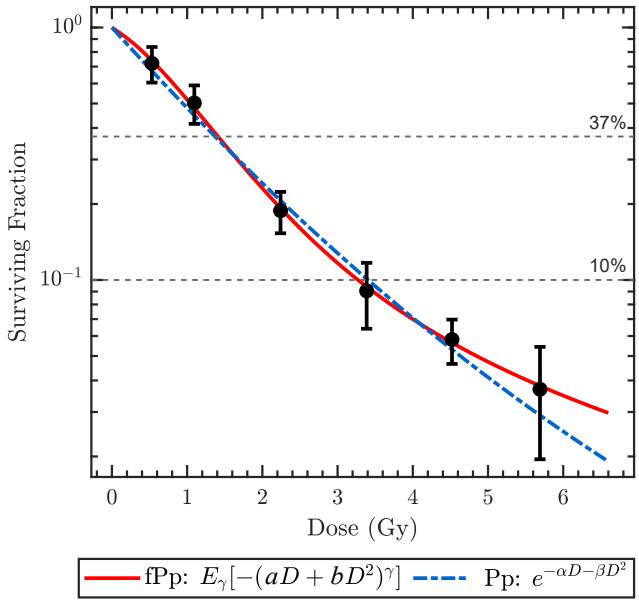


Figure 8.20: The fPp and Pp fits presented on a logarithmic scale to the data of Belli et al. [26] at an LET of 34.6keV/μm.

compared to the high dose range (37% and 10% survival respectively).

The 10% survival RBEs are in very close agreement for low-LET, whereas at 37% they are more scattered. This can be explained using Figs.(8.15) and (8.16), where the fPp and Pp models agree in the high dose region. However, in the low dose region they tend to deviate from each other. This can be attributed to the relative size of the error bars. Therefore, in a weighted fit the curve will most likely go through the points with small error bars and not those with larger error bars. Therefore, even small deviations between the models can impact the predicted RBE.

If we recall that the fPp approaches the Pp in the low-LET region, we expect the RBE predictions of each model to be similar. Comparing this with Fig.(8.19), the differences in the RBE are much smaller.

It has been previously reported [25, 172] that the RBE can vary between cell lines. We observed that the cells have different radiation-sensitivities and hence respond differently to ionising radiation. The data of Belli et al. [26] suggests that the RBE of the V79 Chinese Hamster cell peaks at ~30keV/μm due to damage saturation effects in the high-LET region. We should therefore expect the same behavior in other cell lines. Figure (8.22) depicts the RBE trend as a function of LET for the V79 and C3H10T12 cell lines, both derived from a rodent. Firstly, we observe in Fig.(8.22(b)) that the C3H10T12 cell line does not contain a peak within

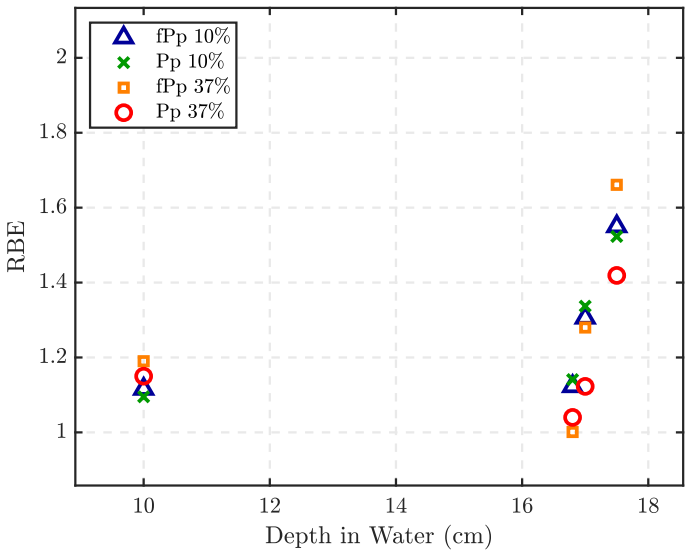
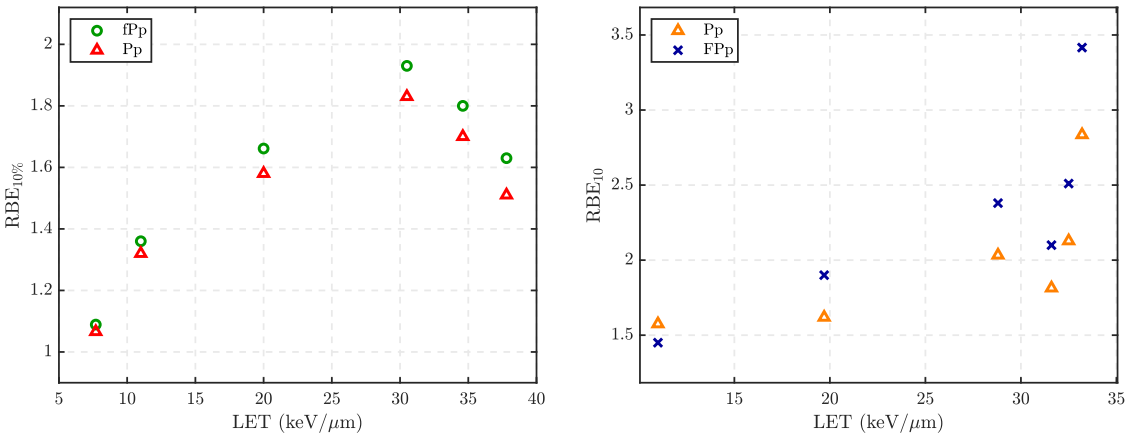


Figure 8.21: RBE calculated at 10% and 37% survival using the fPp and Pp model fits to the data from Howard et al. [123] (160MeV proton beam incident on CHO cells).

the same LET region as the V79 cell line. It is possible that the peak RBE occurs after the range considered in the current study for the C3H10T12 cell line, however some fluctuations in the RBE are observed with increasing LET. The RBE decreases just after 30keV/μm before sharply increasing, which is not consistent with theory. Bettega et al. [61] reports that the discrepancy is the result of a larger cell thickness to that considered in Belli et al. [26], however there is no further justification to this claim.

Overall, we have observed the expected trends of RBE increasing as a function of LET, depth and survival rate (inherently dose). Differences are observed between the RBE predictions made with the Pp and fPp models in the high-LET region where the LQ model does not adequately explain the data (see Chapter 5). The RBE predictions using the Pp and fPp models show more agreement in the low-LET region due to the fractionality parameter γ approaching one where the fPp approaches the Pp. The MHMT, RMR and LPL models have smaller (better) RMSE values than the LQ by a small amount, however the fPp model still yields the smallest RMSE overall. This shows that the fPp model is capable of accurately predicting endpoints such as RBE in the high-LET region due to the improved gof whilst yielding similar predictions in the low-LET region where the fractional and standard Poisson process become equivalent and the fits are close to each other.



(a) Belli et al. (1998) [26]: energy-varied proton radiation incident on V79 Chinese Hamster cells.

(b) Bettega et al. (1998) [61]: energy-varied proton radiation incident on C3H10T12 rodent cells.

Figure 8.22: A comparison of RBE across the V79 and C3H10T12 rodent cell lines for the computed using the fPp and Pp (Chapter 5).

8.3.1.6 Comparison with Established Survival Models

In this section we fit the several established radiobiological survival models to an experimental dose-response curve under low and high-LET exposure conditions. We compare the RMSE, R^2 and χ^2 of these alternative established models with the Pp (LQ model) and fPp models. The aim of this investigation is to provide a benchmark of the fPp model performance against other models.

The models chosen for this investigation are:

1. **Multiple Hit Multiple Target (MHMT):** $S = 1-(1-\exp(-\alpha D))^m$ where m is the number of targets [34],
2. **Repair-Misrepair (RMR):** $S = \exp(-\alpha D)\left(1+\frac{\alpha}{\beta}D\right)^{\beta}$ [31],
3. **Lethal-Potentially Lethal (LPL):** $S = \exp(-(\alpha+\beta)D)\left(1+\frac{\alpha}{\beta}D\right)^{c}$ [32],
4. **Linear-Quadratic (LQ):** $S = \exp(-\alpha D - \beta D^2)$ [23, 35].

All of which differ slightly from the LQ in that repair processes and the induction of lethal DNA lesions are characterised differently (see Chapter 2 for a detailed description of these models). A single fit, weighted by the SEM was performed on two dose-response curves,

1. **High LET:** A 0.64MeV/34.6keV/μm proton beam incident on V79 cells of asynchronous cell cycle from Belli et al. [26],

2. **Low LET:** A 3.2MeV/11keV/μm proton beam incident on V79 cells of asynchronous cell cycle from Belli et al. [26].

The relative R^2, RMSE and χ^2 statistics for fits of each model to the data are presented in Figures (8.23)–(8.25). Moving systematically, the χ^2 values in Fig.(8.23) indicate that the LQ model performs the worst for high-LET, although the low-LET fit is below 1 and hence can still be considered to be a good fit. The MHMT, RMR and LPL models yield a similar χ^2 value. The RMR and LPL models account for biological repair processes in their terms, which explains why they result in a better χ^2 than the LQ model, which only considers ionisations to the target. The MHMT model interestingly performs better than the LQ, which is likely the results of the extra parameter in the power. The fPp model yields the smallest (best) χ^2 in the low-LET region compared to the other models, even if all of the χ^2 statistics imply a good fit for each model. The R^2 values are the largest (best) for the LQ and fPp models, however the fPp model is the highest. This indicates that the fPp model most adequately accounts for variances in the data. The RMSE values are the highest (worst) for the MHMT and LPL models with values of $\sim$2 and 2.7, respectively, implying the fits are further from some of the points. The RMSE is for the fPp model whilst the LQ model RMSE is quite large in comparison. We expect this because we are considering an LET of 11keV/μm, it is possible that there is some deviation from Poissonian behavior. Overall, all the models considered here fall within the error bars of the low-LET data of Belli et al. [26], however there is slightly more variance not explained by the Pp-based models which is also reflected in the higher (worse) RMSE values.

Figure (8.23) the high-LET region the fPp model yields the smallest (best) χ^2, followed by the LQ, RMR, MHMT and LPL models. The χ^2 values for all the models in Fig.(8.23) are less than one indicating that all fits are within the error bars of the data. Although the LPL model is the worst performer, followed by the RMR, MHMT and LQ models. Figure (8.24) depicts the R^2 values for each model. In the high-LET region the fPp clearly explains the variance in the data better than the other models, which yield values of < 0.98. Recall from Fig.(8.20), the fit of the fPp and Pp model to the data considered in the current study, that at high-LET the cell killing efficiency decreases at high doses due to saturation effects. This behavior is better explained by the fPp model, whilst the other Pp-based models would fail to account for the "tail" region of the dose-response curve at high-LET. The RMSE is the largest for the LQ model (a smaller RMSE results in a better fit), which is expected due to the model's simplicity and inability to describe the complex process of high-LET cell irradiation. The LPL and RMR models perform reasonably to this extent due to the inclusion of biological repair processes in their survival equations. The RMSE of the LQ model is slightly larger than the fPp model, which is shown by the quality of fit in Fig.(8.20) (the fPp and Pp fits to the data considered here)

due to the disagreement in the high dose region as discussed above.

Overall, we observe a highly improved performance of the fPp model for both low and high-LET against other survival models. The versatility and flexibility of the fPp model (due to its ability to converge to the Poisson process) allows it to perform well where Poissonian and non-Poissonian statistics is obeyed. The remaining models, whilst yielding a reasonable fit, also do not perform well in the high-LET region because they are all based on the Pp which has been shown in Chapters 5 and 6 to lose validity.

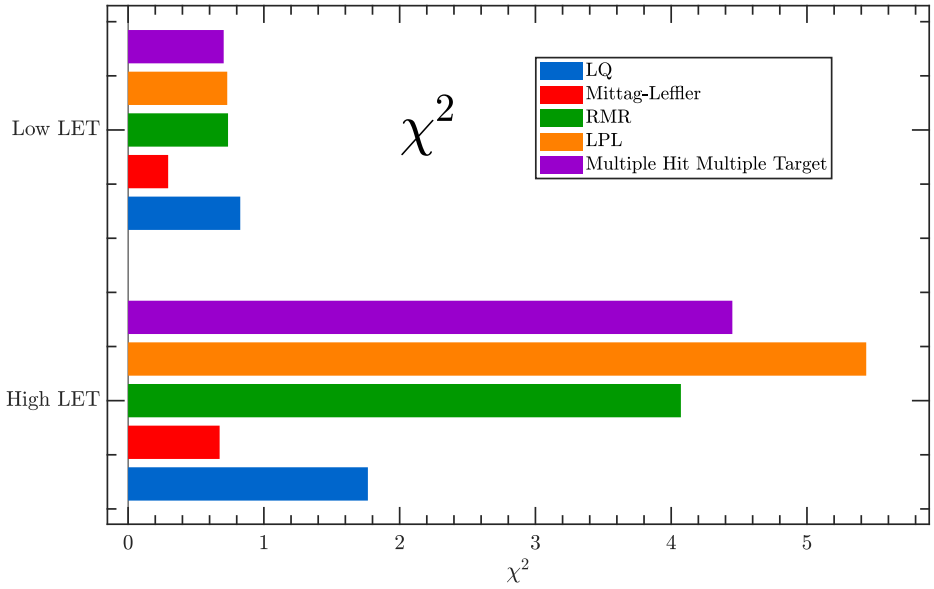


Figure 8.23: Comparison of χ^2 goodness-of-fit statistics for five survival models fitted to high and low-LET data [38, 96].

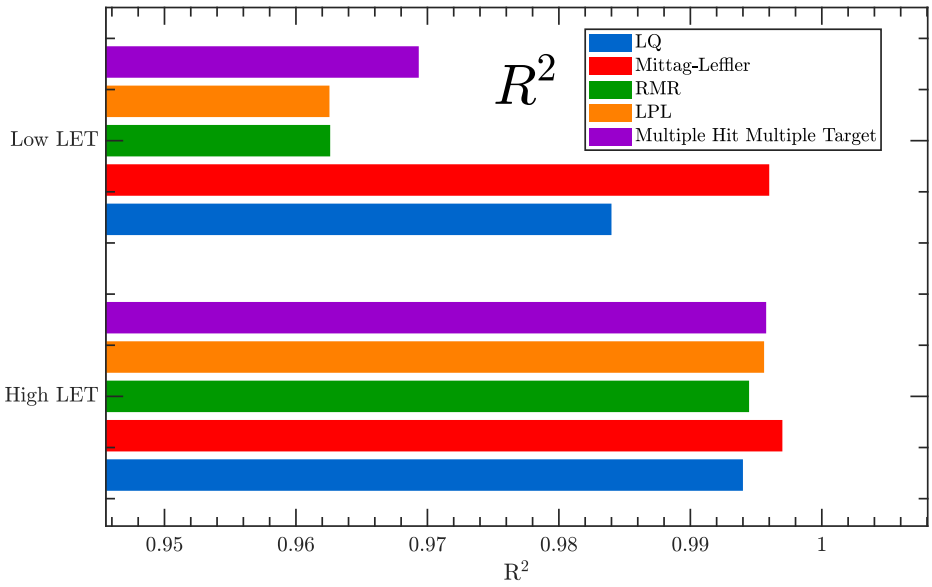


Figure 8.24: Comparison of R^2 for five survival models on high and low-LET data [38, 96].

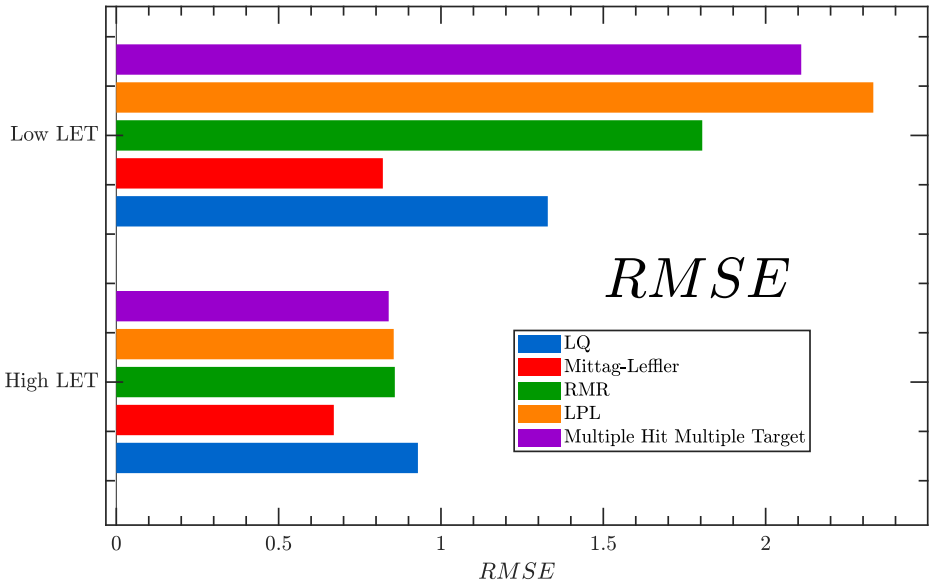


Figure 8.25: Comparison of RMSE for five survival models on high and low-LET data [38, 96].

8.3.2 Testing the FPp Model on Simulated Data

We have already introduced some of our simulated data in Chapter 6 and Section (8.2) in the form of a lethal lesion distribution per cell. The data from the literature to

which we have access does not contain data for multiple endpoints of cell response, such as DNA damage yield and repair time. Therefore, we will use our simulated data to observe the cell survival (zero count state $k = 0$ of fPp) probability as a function of dose and repair time in this section. We will use three methods to assess the stability of our fits to the simulated data :

- **Method 1 -** Fit the Mittag-Leffler distribution to the mean of our independent cell survival fraction measurement repetitions using the standard deviation for weighting.
- **Method 2 -** Fit the Mittag-Leffler distribution to the data for each repetition individually for a given absorbed dose range, resulting in a parameter range for a, b and γ of Eq.(8.6). We take the mean of each parameter to be the final fit.
- **Method 3 -** Fit the Mittag-Leffler distribution to all our cell survival fraction measurements versus all doses.

As we did in Section (8.3.1), we will assess the quality of fit of our fPp (Mittag-Leffler) model for comparison with the Pp (LQ) model. Making use of the three fitting methods described above we will compare the gof metrics (R^2, RMSE, etc.) and χ^2 values against the LQ model to determine which yields the best explanation of the data.

8.3.2.1 Energy-varied LET

The LET is varied by energy whilst positioning the cells in the centre of a cube filled with air and surrounded by a water medium (see Chapter 4 for all the simulation parameters). The fit results are presented in Tables (L.1) (fPp - method 3), (L.2) (Pp - method 3) and (L.3) (fPp and Pp - method 1) of Appendix L. We exhibit our simulated survival fraction data as a function of dose for one high, intermediate and low-LET simulated data set in Figs.(8.26) to (8.31) after 2 and 40 hours of repair time have lapsed using fitting methods 1 and 3. The remaining fits fPp and Pp fit plots are presented in Appendix L.4. The LQ and Mittag-Leffler distribution fits on the data are shown in the figures. Additionally, the experimental data in a similar LET range are shown. We also provide the relevant gof metrics are presented in the tables below each figure.

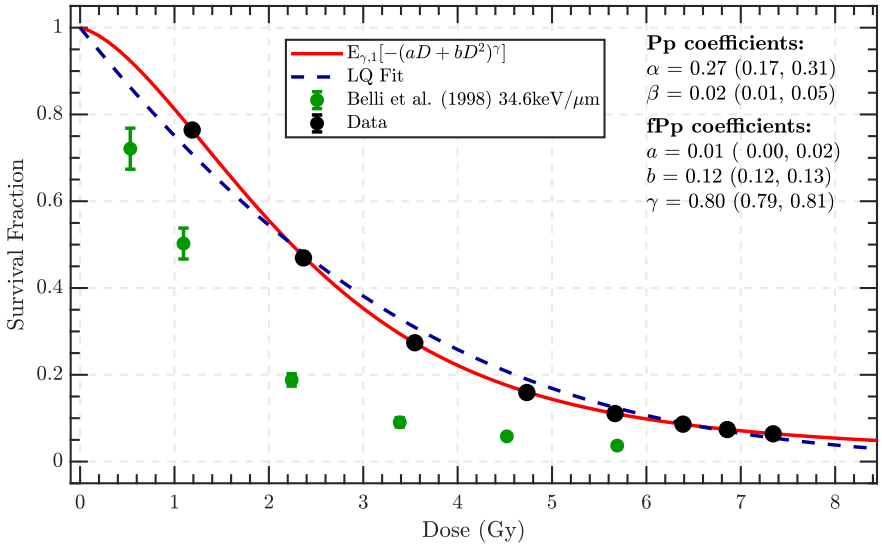
Upon observation of Tables (L.1), (L.2) and (L.3), one sees that the fPp model consistently yields a superior fit compared to the Pp model with a substantial improvement in the gof metrics despite the fact the errors are extremely small. The most noticeable improvement is in the high-LET region, where we concluded that the LQ model (hence the Pp) failed to adequately explain the data in Chapter 5. We expect this when we recall Chapter 5 where LET beyond 30keV/μm resulted in

dose-response moving from a concave to a convex relation for the LQ model and the previous section where a considerable improvement in fit was observed for the fPp model. The same trend is reflected in our simulated results, where the difference between the χ^2 statistics for the fPp and Pp models is the largest. As previously mentioned, method 1 yields χ^2 statistics greater than 1 for the fPp model in addition to the Pp model. However, this is not a reflection on the fits themselves but on the small error bars. To resolve this issue we require more data which could not be achieved in the current study due to time constraints.

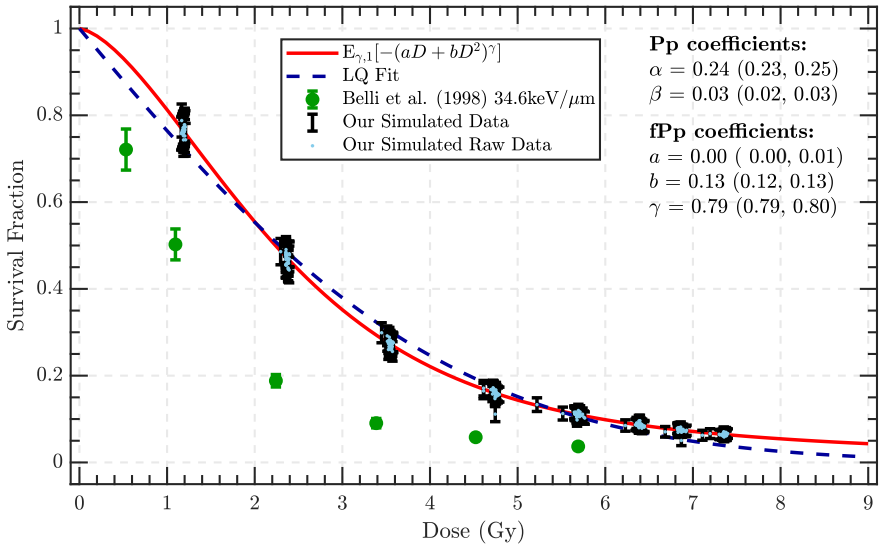
Figure (8.26) depicts the fPp and Pp fits to the highest LET data set considered in our simulations (35keV/μm) after 2 hours of repair time, with experimental data from Belli et al. [26] for comparison. As discussed in Chapter 7, the TLK model used to simulate the repair processes in our simulations determines that survival probability by accumulating lethal lesions with repair time. When the number of lethal lesions in a given cell increases, its probability of cell survival will decrease. Therefore, at 2 hours of repair time not all the lesions have undergone repair yet meaning that the number of lesions classified as lethal by the TLK model is smaller than for longer repair times. In the realistic picture we expect that the survival fraction will increase with repair time because the lesions present in the cell do not need to be classified as lethal like in the TLK model. Regardless, we decided to fit the fPp and Pp models at small repair times when the system is considered to be "out of equilibrium".

Both fitting methods yield similar fits with a fractionality parameter, γ of $\sim$0.80. The fPp model fit is extremely good compared to the Pp model ($\chi^2 = 0.158$ using method 1 and 0.040 using method 3). Since the χ^2 alone is not an adequate measure for gof in the current context, we must compare other metrics. The RMSE of the fPp fit in the table below Fig.(8.26) is substantially smaller than that of the Pp model, indicating that the Euclidean distances between the model predicted and observed survival fractions is minimal. The SSE paints a similar picture with consistently small values. The R^2 value is very close to 1 for both the Pp and fPp models, however the fPp fit is higher where $\sim$99% of the variance in the data is accounted for.

Figure (8.27) depicts the fPp and Pp model fits using methods 1 and 3 to the highest LET data set (35keV/μm) considered in our simulations after 40 hours of repair time has elapsed, i.e. when the system has some to equilibrium. From Chapter 7, the survival fraction is much more consistent with experiment under similar exposure conditions with the same cell line (Belli et al. [26]). The fractionality parameter, γ, is 0.81 using both methods, indicating that the fit parameters are stable. We expect γ to be far from 1 for this LET region where Poissonian behavior breaks down (Section (8.2), Chapter 5 and 6). The gof metrics underneath Fig.(8.27) further reflect the superior fit of the fPp model compared to the Pp model. Comparing Figs.(8.26) and (8.27), the fractionality parameter is very similar despite the slower



(a) Method 1



(b) Method 3

Figure 8.26: Fits of the fPp and Pp models to our simulated cell survival data of V79 cells exposed to 1.6MeV (35keV/μm) proton radiation after 2 hours of repair time.

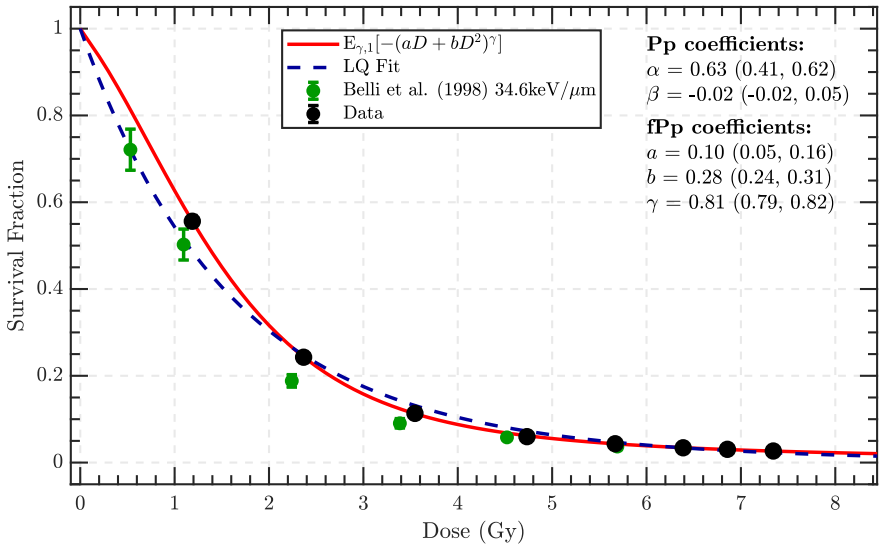
Method	Model	χ^2	R^2	RMSE	SSE
1	**ML** (LQ)	**0.158** (2.473)	**0.999** (0.989)	**0.008** (0.0245)	**0.014** (0.133)
3	**ML** (LQ)	**0.040** (11.564)	**0.999** (0.986)	**0.0017** (0.238)	**0.004** (0.011)

Table 8.3: Fit results from Fig.(8.26) above. The fPp fits are presented in **bold** and the corresponding Pp results are presented below.

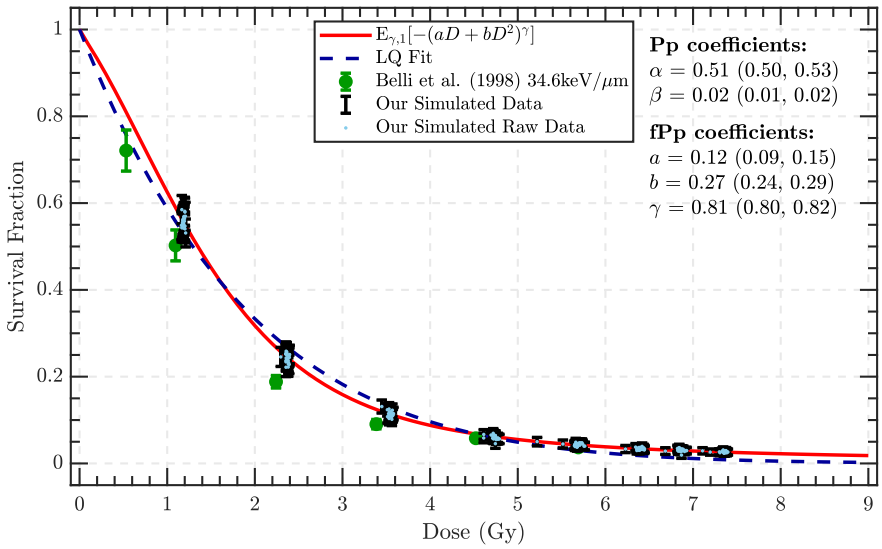
decrease in survival fraction in Fig.(8.26). However, if we compare the parameters a and b across repair times, they are higher for longer repair times. We expect this behavior because the lethal lesions accumulate with time within the TLK model framework, meaning that contributions to lethal lesions will be higher for both the linear and quadratic components of the survival function. The mid-LET range exhibits similar behavior as shown in Figs.(8.28) and (8.29) where the fPp and Pp models are in more agreement after 2 hours of repair time, when the system has not yet reached equilibrium. As the system reaches equilibrium after 40 hours of repair time, the fPp performs much better than the Pp.

Figures (8.30) and (8.31) depict the fPp and Pp fits to the data of the lowest LET considered in our simulations following repair times of 2 and 40 hours, respectively. The survival curve is much more gradual at low-LET due to the smaller DNA damage yields incurred by the cells compared to high-LET radiation as shown in our results from Chapter 7. The fractionality parameter is closer to 1 than for high-LET conditions which is expected behavior due to the equidispersion condition being satisfied in this region (see Section (8.2)). However, in the low-LET region of the experimental data discussed in Section (8.3.1), γ is greater than 0.9, implying that the Pp is satisfied. For our simulations, γ is 0.85 which implies that fractionality is still present, but this does not contradict with our observations of the experimental data. This is an expected outcome since our simulations did not include indirect effects (chemical processes) and the full dose-response of a cell as in experiment. Furthermore, the cells in our simulation do not go through the exact physical conditions as in experiment. Currently we are running our simulations with the chemical processes included to quantify these effects for which we expect the outcome to give similar results to the experimental data.

Recall from Chapter 4 that not all process of cell irradiation and response were included in our simulations due to longer processing times and larger memory requirements. The simulated and realistic physical pictures are therefore quite different as we do not include all repair processes undergone by a cell, conditions within the cell environment and the indirect contributions to DNA damage (chemical processes/radiolysis reactions). The discrepancy between simulation and experiment implies that there are additional parameters aside from LET that can contribute to fractionality. The exclusion of indirect DNA damage contributions, in particular, can change the arrival times between DNA strand break inductions. This effect is most visible in the low-LET region where indirect effects have a greater contribution to DNA damage than for higher LET. Furthermore, there are many biological parameters such as cell cycle effects and temperature can impact a cell's response to ionising radiation. Under some conditions the cell is less likely to repair damage via a given pathway, meaning that the amount of lethal damage for increasing repair times can differ across cells and ultimately changing the degree of fractionality



(a) Method 1



(b) Method 3

Figure 8.27: Fits of the fPp and Pp models to our simulated cell survival data of V79 cells exposed to 1.6MeV (35keV/μm) proton radiation after 40 hours of repair time.

Method	Model	χ^2	R^2	RMSE	SSE
1	**ML** (LQ)	**3.699** (32.451)	**0.998** (0.987)	**0.007** (0.019)	**0.011** (0.084)
3	**ML** (LQ)	**0.056** (0.775)	**0.999** (0.993)	**0.021** (0.021)	**0.003** (0.014)

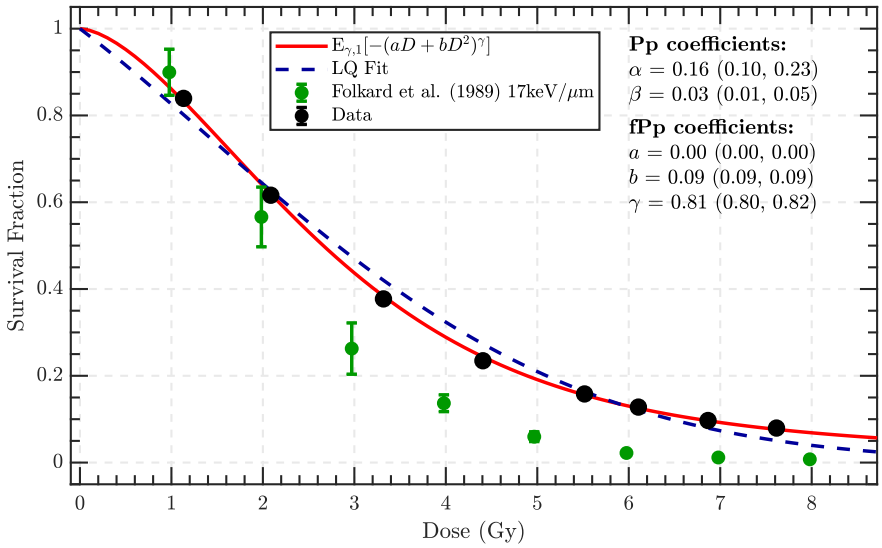
Table 8.4: Fit results from Fig.(8.27) above. The fPp fits are presented in **bold** and the corresponding Pp results are presented below.

within the system.

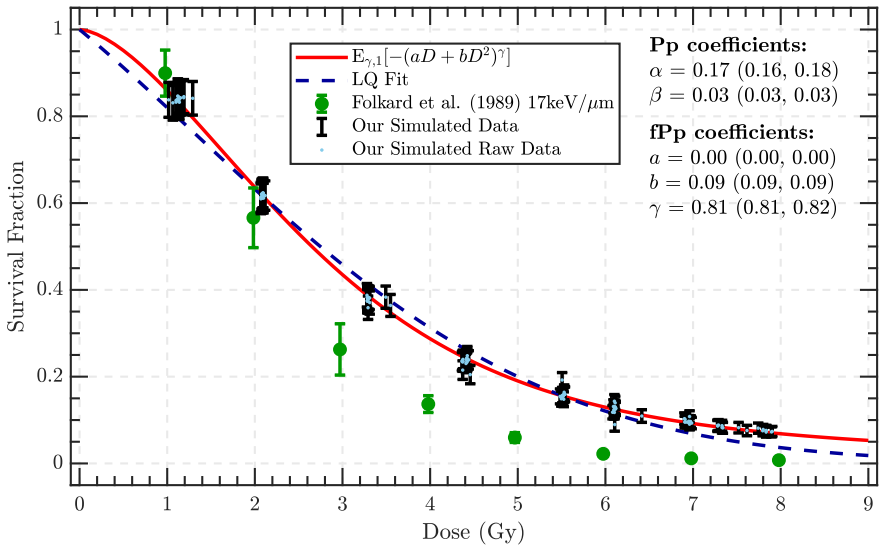
Regardless of the processes included in our simulations, the fPp fits to the low-LET simulated data gives quite a comparable or even much improved fit compared to the Pp. In Section (8.3.1) a similar observation is made in that the fPp and Pp model fits tend to agree more in the low-LET region than high-LET. Figure (8.30(a)) in particular is a good example of the agreement between the models due to the fPp approaching the Pp as γ increases.

In Chapter 5, the hypothesis testing methods for non-linear regression described in Appendix E on a non-linear least-squares LQ fit, weighted by the SEM. Table (L.7) summarises the hypothesis testing of the fPp fits performed on our simulated data for repair times of 2 and 40 hours. We make use of our fit results from method 1 to do this. In Chapter 5 we observed very high fail rates for LQ fits to many experimental data sets (see Appendix G for the results). The superior fit of the fPp model is further evidenced by an almost unanimous pass rate in the fPp for our simulated data. After two hours of repair time, most of the fPp fits pass the test of Gaussian residuals (Kolmogorov-Smirnov and Anderson-Darling test) and homoscedasticity (Breusch-Pagan and White tests for equal variance of residual errors) with the exception of our 32keV/μm data. This is the result of more variance about the fit at two hours because the system has yet to equilibrate. The same data set also fails the BP test for the Pp fit. After 40 hours of repair time, the cells have equilibrated and it is likely there are not more potentially lethal damages present by this time - only lethal lesions contributing to cell death. All data passes the four tests with the exception of 29keV/μm, which fails the AD test despite a good fit (see Table (L.2) for the gof metrics). The Pp model fits have a much higher failure rate than the the fPp model, particularly in the high-LET region. A similar observation is made for the hypothesis testing results on experimental fPp fits discussed in Section (8.3.1). Therefore, we have observed that the fPp model can adequately explain the data, particularly in regions where the LQ model (Pp) does not.

Considering all the fits presented in this section (see Tables (L.1)–(L.3)), the expected decrease in γ with increasing LET is observed and is consistent with what was observed in the experimental data. As expected, the under simulated conditions the range is much smaller (0.78-0.86) than in experiment (0.75-1.00). As discussed above, we can attribute the small range in γ to the fact the our simulated picture is an approximation of the realistic picture where not all biological and chemical processes within the cells could feasibly be included without requiring excessive computation power and processing time. Regardless, we find the fPp to be highly improved upon the Pp model when considered in terms of gof. The "clustering effect" in the γ parameter observed in Section (8.3.1.3) was not observed using the two fitting methods considered for our simulated data. This is likely the result



(a) Method 1

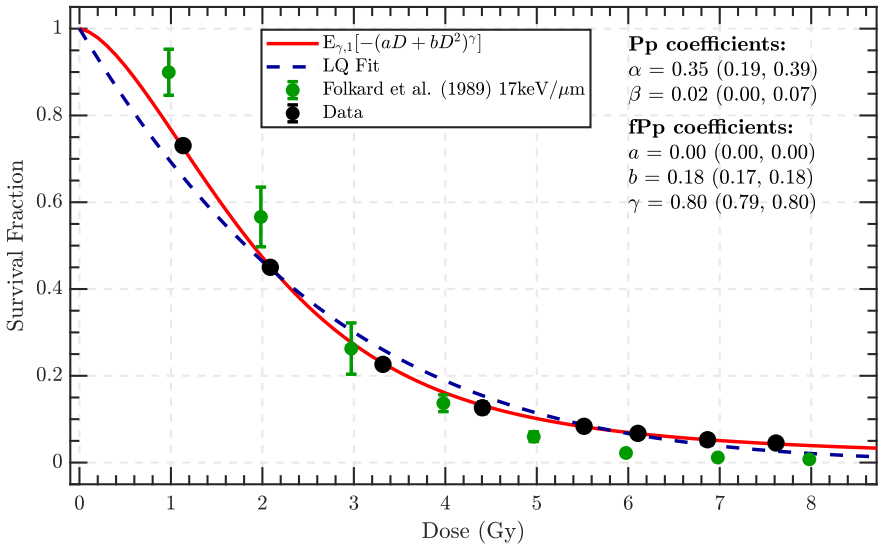


(b) Method 3

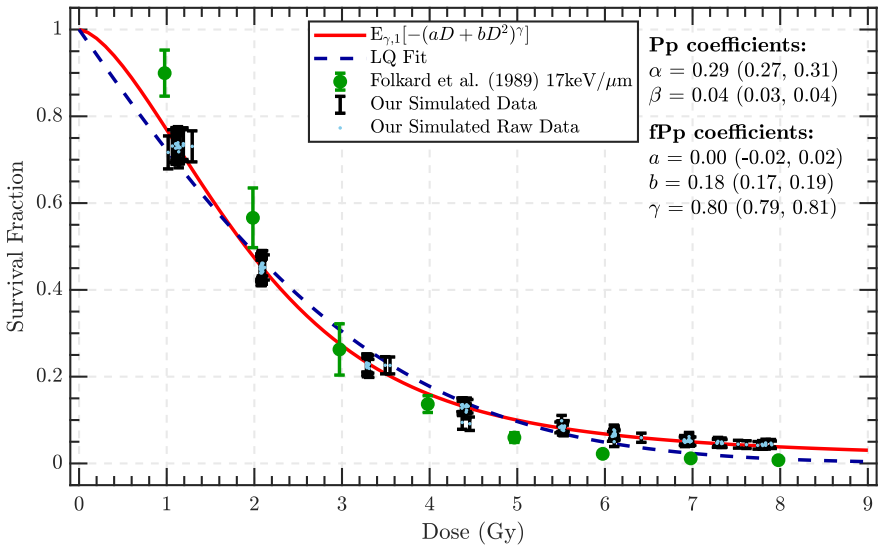
Figure 8.28: Fits of the fPp and Pp models to our simulated cell survival data of V79 cells exposed to 2.5MeV (16keV/μm) proton radiation after 2 hours of repair time.

Method	Model	χ^2	R^2	RMSE	SSE
1	**ML** (LQ)	**0.751** (26.229)	**0.999** (0.999)	**0.004** (0.004)	**0.011** (<0.001)
3	**ML** (LQ)	**0.247** (2.975)	**0.998** (0.988)	**0.011** (0.028)	**0.003** (<0.001)

Table 8.5: Fit results from Fig.(8.28) above. The fPp fits are presented in **bold** and the corresponding Pp results are presented below.



(a) Method 1



(b) Method 3

Figure 8.29: Fits of the fPp and Pp models to our simulated cell survival data of V79 cells exposed to 2.5MeV (16keV/μm) proton radiation after 40 hours of repair time.

Method	Model	χ^2	R^2	RMSE	SSE
1	**ML** (LQ)	**0.373** (47.502)	**0.999** (0.986)	**0.004** (0.004)	**<0.001** (<0.001)
3	**ML** (LQ)	**0.229** (8.476)	**0.998** (0.985)	**0.010** (0.028)	**0.031** (0.106)

Table 8.6: Fit results from Fig.(8.29) above. The fPp fits are presented in **bold** and the corresponding Pp results are presented below.

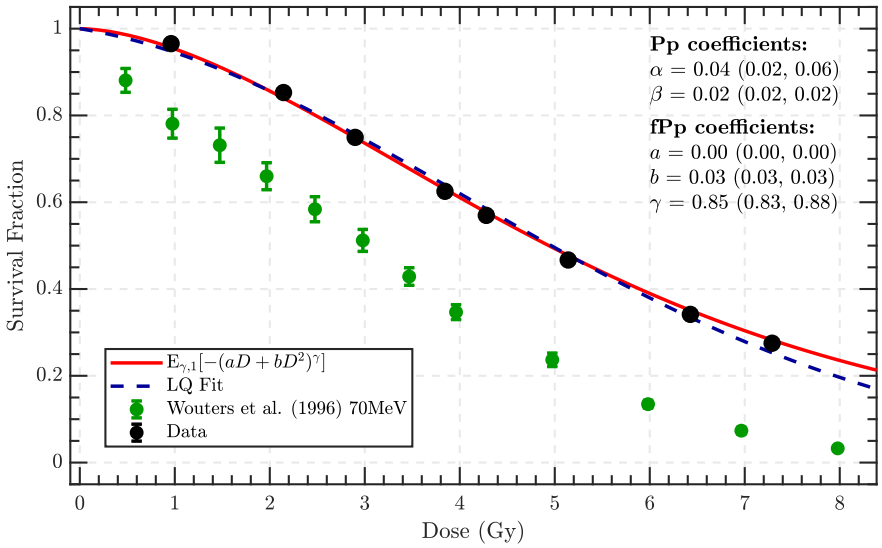
of fewer outliers presented in our simulated data compared to the experimental data.

To conclude, our newly proposed Mittag-Leffler distribution fPp model fits exceptionally well to the simulated data further evidencing the fractional Poisson process as a much better alternative to the less flexible Poisson process at describing radiation-induced cell death. The over/underdispersion effect observed in Chapter 6 and Section (8.2) does influence the cell survival probability estimated by each model, which can lead to a misleading dose-response curve. The improved fit of the fPp model compared to the Pp model is more pronounced in the high-LET region where we concluded in Chapter 5 that the LQ model is no longer valid.

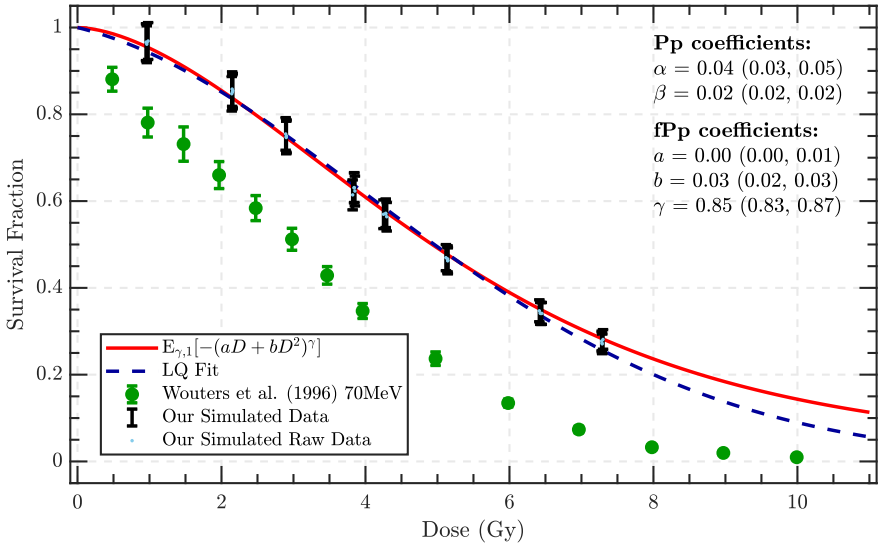
8.3.3 Depth-varied LET

Varying LET by depth in a medium is another comparison we will conduct here with our simulated data. Although we were subject to time constraints, we felt it was necessary to compare our model in the energy and depth-varied LET cases as we do in Section (8.3.1). The detailed simulation parameters are presented in Chapter 4, however it is helpful to briefly recall the depths considered here. We note that the number of simulation repetitions for all of the data sets considered here is 5. Therefore, the χ^2 statistic alone is not a reliable measure of goodness of fit because the errors are so small that most curves would struggle to fit within all error bars.

We generate a pristine Bragg Peak with a mean proton beam energy of 20MeV inside a water medium and increase the depth of the cells inside the medium (Fig.(8.32)). When we approach the Bragg Peak, we are inherently increasing the LET. In Section (8.3.1.4), the LET range was very small compared to what we have observed here due to the lower mean energy of the proton beam (20MeV in our simulations versus 70-160MeV in experiment). Recall our results from this section that the degree of fractionality in the system is very low in the low-LET region but increases slightly within the Bragg Peak. Although this effect was small due to the small LET range, it was consistent with our theory that fractionality increases with LET. In this section we expect to see a more pronounced effect because the LET ranges from 3 to 28keV/μm. However, the way a primary proton deposits energy in the entrance/proximal (0 to 3mm), inter-peak (3 to 4.2mm) and distal (>4.2mm) regions will likely be reflected in the fPp fits as they do in Section (8.3.1.4).



(a) Method 1

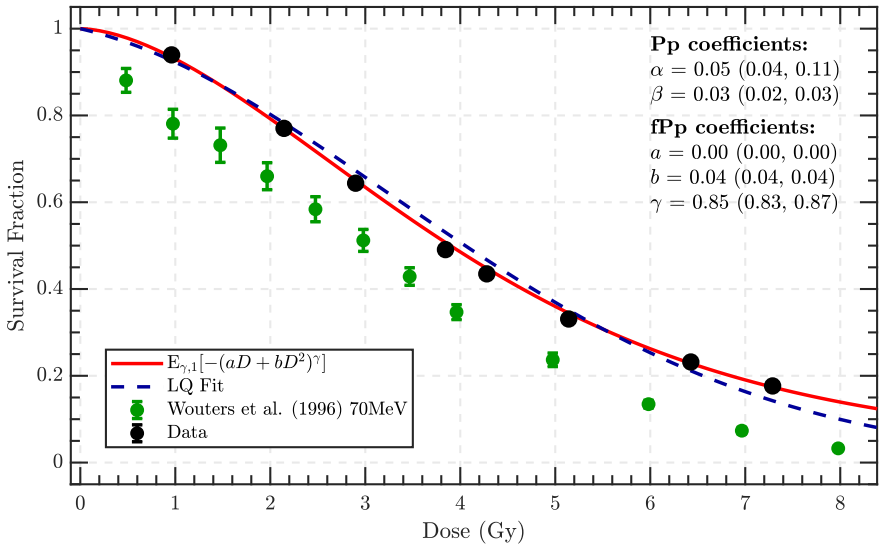


(b) Method 3

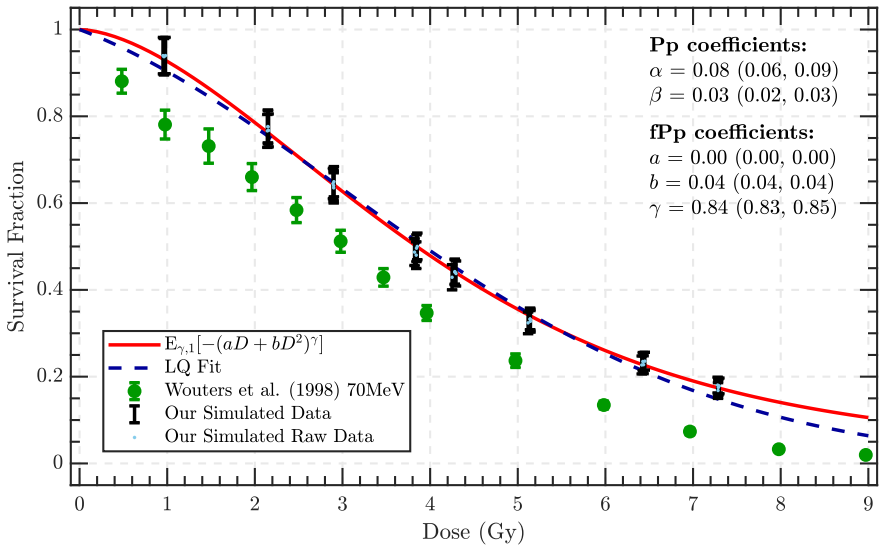
Figure 8.30: Fits of the fPp and Pp models to our simulated cell survival data of V79 cells exposed to 50MeV (1.2keV/μm) proton radiation after 2 hours of repair time.

Method	Model	χ^2	R^2	RMSE	SSE
1	**ML** (LQ)	**7.637** (20.169)	**0.996** (0.998)	**0.009** (0.015)	**0.003** (0.007)
3	**ML** (LQ)	**0.087** (0.251)	**0.998** (0.996)	**0.0090** (0.0150)	**0.0070** (0.0070)

Table 8.7: Fit results from Fig.(8.30) above. The fPp fits are presented in **bold** and the corresponding Pp results are presented below.



(a) Method 1



(b) Method 3

Figure 8.31: Fits of the fPp and Pp models to our simulated cell survival data of V79 cells exposed to 50MeV (1.2keV/μm) proton radiation after 40 hours of repair time.

Method	Model	χ^2	R^2	RMSE	SSE
1	**ML** (LQ)	**0.157** (8.245)	**0.998** (0.987)	**0.007** (0.019)	**0.011** (0.084)
3	**ML** (LQ)	**0.094** (6.003)	**0.999** (0.987)	**0.008** (0.0190)	**0.002** (0.0840)

Table 8.8: Fit results from Fig.(8.31) above. The fPp fits are presented in **bold** and the corresponding Pp results are presented below.

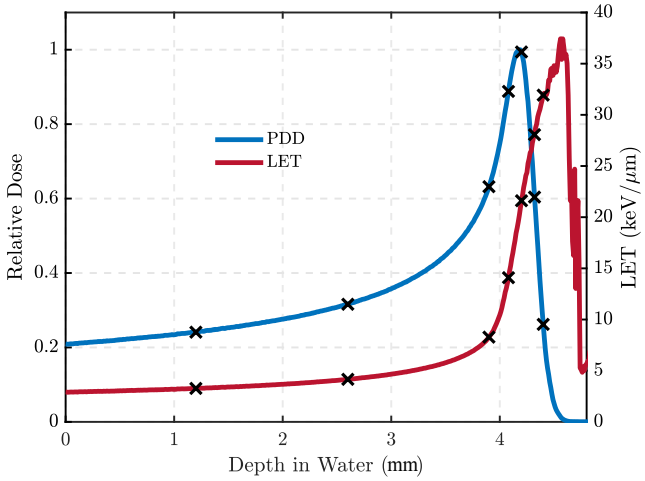


Figure 8.32: The 20 MeV proton beam PDD and dose-averaged LET as a function of depth in liquid water with the depths studied using the current model indicated. The "×" denotes the LET and relative dose at the depths chosen for the current study.

Tables (L.4)–(L.6) of Appendix L depict the parameters and gof results for the fPp and Pp models using methods 1 and 3. We observe a similar trend as the energy-varied case in that the fractionality parameter, γ, is the highest in the low-LET regions, i.e. in the proximal region of the Bragg Peak. When the cells are positioned within the Bragg Peak, γ begins to decrease. This is similar behavior to what was observed in Section (8.3.1.4), where the experimental data resulted in a γ parameter of $\sim$0.99 for the proximal regions of the Bragg Peak, only increased within the peak region. This is the result of a relatively consistent LET in the proximal region (see Fig.(8.32)). The LET only varies greatly within the Bragg Peak. After the Bragg Peak, γ increases slightly which is consistent with our observation in Section 8.3.1.4 where two of the LET data sets from Guan et al. [27] increased γ at depths after a 79.7MeV Bragg Peak.

A similar γ range is observed in Tables (L.4) and (L.6) (0.79 to 0.84) for fits using methods 1 and 3 as in the energy-varied case despite observing a range of 0.95 to 0.99 in the experimental data. As discussed in the previous section, there are parameters we could not include in our simulations due to the large processing times involved. Some of the parameters not included in the simulations can influence the damage yield and availability of repair processes, which ultimately affects the number of lethal lesions in each cell as a function of time. Therefore, it is plausible to expect different results to experiment where the full physical picture is considered.

The γ parameter is $\sim$0.83 in the proximal regions before the Bragg Peak. We can see that the LET does not change much until depths within the Bragg Peak (see Fig.(8.32)), where it increases dramatically. Then γ decreases to $\sim$0.8 in the peak

region until the largest depth in the distal region where it increases slightly to 0.81.

In the distal region, however, primary protons penetrate to this depth less frequently. Some protons and secondary particles can propagate further, especially if the secondary particle is formed within the peak region. Recalling our discussion of interaction cross sections from Chapter 7, we know that the highest ionisation yield occurs for protons of low energy. Therefore, in the distal region one should expect a higher degree of fractionality because the protons are at their maximum LET as they slow down. However, we do not observe a lower γ in the distal region from our simulated results where it increases to 0.81. The reason for this could be that a smaller fraction of protons will propagate to the distal region of the peak compared to before and inside the peak. Moreover, in the distal region a large number of ionisations could be caused by secondary interactions from particles created from a proton interaction, which can have a lower LET than a primary proton. Moreover, the protons that propagate to the distal region will be of high-LET but they will be less frequent, hence affecting the inter-arrival times between DNA damage. A decrease in DNA damage contributions to the targets (cells) in the distal region of the Bragg Peak by primary protons could therefore explain the slight increase in γ for the largest depth in Tables (L.4) and (L.6).

The goodness of fit metrics also indicate a superior fit for the fPp model compared to the Pp model. However, the χ^2 statistic does not accurately reflect the fit of each model due to the extremely small errors as discussed above. In Fig.(8.33) for example, the fPp fit passes through all the points upon observation, whilst the LQ fit is close, but does not pass within the error bars. The extremely large χ^2 of the LQ model is therefore the result of the fit not passing through the small error bars unlike the fPp model. Nevertheless, we can see that the fPp model results in consistently lower χ^2 values than the Pp. For a more explanatory measure of gof, we can observe the R^2, SSE and RMSE metrics. The R^2 values remain at 0.99 for all cases of the fPp model and it is clear why by observing Figs.(8.33) and (8.34) where the fPp curve (red) passes through every point. The LQ model comes close but does not pass through all the points, resulting in a smaller R^2 value. The SSE and RMSE in the tables underneath Figs.(8.33) and (8.34) reflect the closeness of the fPp fit to the data points compared to the Pp, with the differences in the order of one magnitude smaller for the fPp model at least.

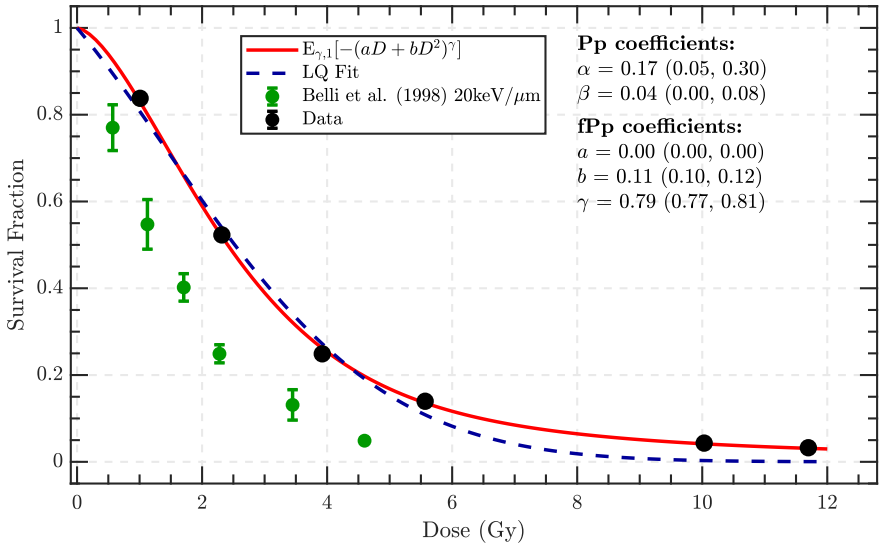
As for to the energy-varied case, we performed hypothesis testing for non-linear regression (see Appendix G for a detailed description) after 2 and 40 hours of repair time had elapsed. The results are presented in Table (L.7) of Appendix L. However, we can attribute this to the fact that the system has not equilbrated. After 40 hours of repair time, the system has reached equilibrium. All fPp fits pass tests for Normally-distributed and homoscedastic (constant variance) residual errors except the final fit which fails the Kolmogorov-Smirnov and Anderson-Darling tests. How-

ever, this is the result of an outlier which impacts both fits. As previously discussed, the behavior in the distal region of a Bragg Peak could lead to scattered survival measurements due to less frequent high-LET primary protons propagating to the depth it is situated at. Regardless, the pass rate is higher for the fPp model than the Pp model for positions within a pristine 20MeV proton Bragg Peak, thus further evidencing the superiority of the fPp model.

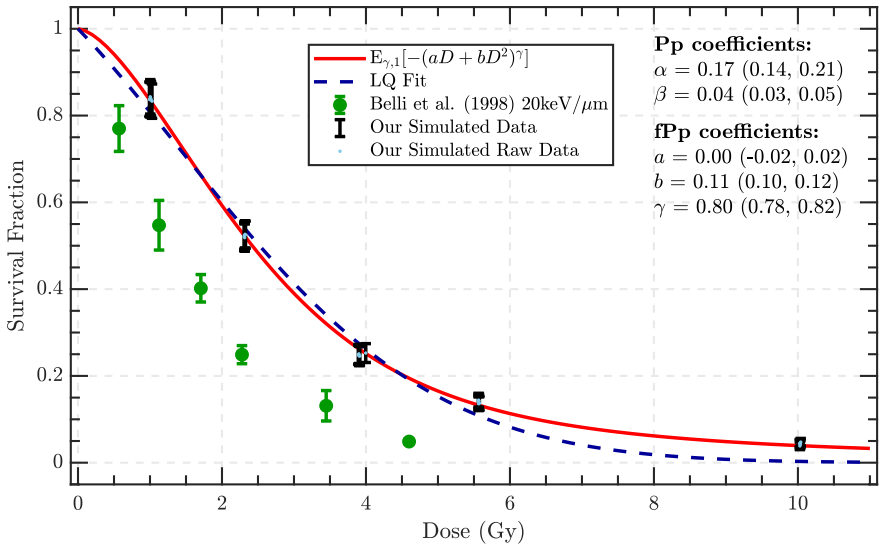
Overall, the fPp model yields comparable fit to the Pp model for the data of Howard et al. [123] and Wouters et al. [38]. As expected, the fractionality parameter, γ, is close to 1 due to the extremely low-LET radiation considered in the two studies. We observe γ to gradually decrease for positions inside the Bragg Peak compared to the entrance and proximal regions. Fig.(8.18) depicts fits to data of Howard et al., where the cell killing efficiency is at its highest inside the Bragg Peak which is expected.

8.3.4 Conclusion

To conclude, the Mittag-Leffler distribution yields a superior fit to the cell survival fraction as a zero count ($k = 0$) of the fractional Poisson process model for our simulated data compared to the LQ model, thus further evidencing the observations made in Section (8.2). The ML model behaves similarly in the energy and depth-varied LET cases of our simulated data, in that the fractionality parameter, γ, increases with LET and depth (inherently LET). The range of the fractionality parameter is narrower for our simulated data compared to experiment because we make many approximations of the full physical picture in our simulations. In experiment the full physical picture with all irradiation and repair processes, not to mention cell states, present results over a wider range in γ. The parameters of the fPp model have some correlation with LET, however, $a \to 0$ for many of the cases in the low-LET region (see Appendix L). In the context of the LQ model, we know that a and b are correlated with the lethal lesion count due to single and multiple hits to the target. However, in the high-LET region where γ is far from 1, the interpretations of these parameters will change. Moreover, we plan to test alternative dose-time relations to determine if the fits using Eq.(8.6) can be improved.



(a) Method 1 (Note: error bars are present but too small to be visible)

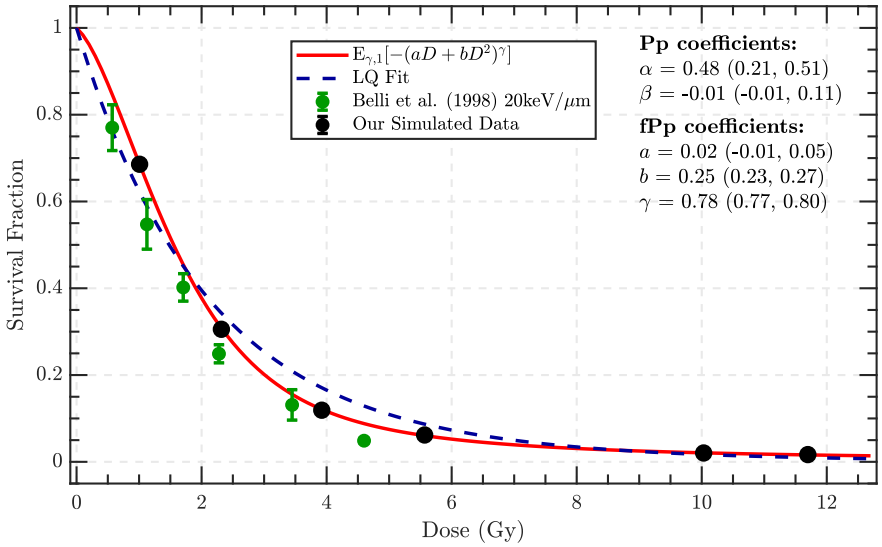


(b) Method 2

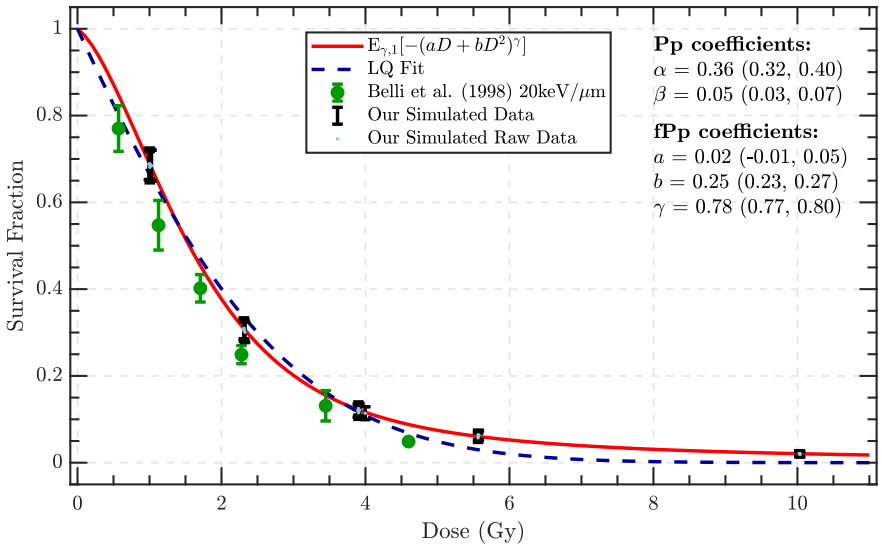
Figure 8.33: Fits of the fPp (Mittag-Leffler) and Pp (LQ) models to our simulated survival data of V79 cells exposed to 20MeV (20keV/μm) proton radiation at 4.20mm depth in liquid water after 2 hours of repair time.

Method	Model	χ^2	R^2	RMSE	SSE
1	**ML**	**8.93**	**0.999**	**0.006**	**<0.001**
	(LQ)	(345.85)	(0.969)	(0.0504)	(0.0153)
2	**ML**	**0.198**	**0.999**	**0.007**	**0.001**
	(LQ)	(256.76)	(0.989)	(0.0302	(0.0274)

Table 8.9: Fit results from Fig.(8.33) above. The fPp fits are presented in **bold** and the corresponding Pp results are presented below.



(a) Method 1 (Note: error bars are present but too small to be visible)



(b) Method 3

Figure 8.34: Fits of the fPp (Mittag-Leffler) and Pp (LQ) models to our simulated survival data of V79 cells exposed to 20MeV (20keV/μm) proton radiation at 4.20mm depth in liquid water after 40 hours of repair time.

Method	Model	χ^2	R^2	RMSE	SSE
1	**ML** (LQ)	**0.048** (188.71)	**0.999** (0.988)	**0.003** (0.0302)	**<0.001** (0.0274)
2	**ML** (LQ)	**0.094** (1697)	**0.999** (0.990)	**0.008** (0.0230)	**0.002** (0.0159)

Table 8.10: Fit results from Fig.(8.34) above. The fPp fits are presented in **bold** and the corresponding Pp results are presented below.

8.3.5 Cell Survival as a Function of Repair Time

Here we observe the change in cell survival fraction (SF) with repair time, t_{rep}. The SF is observed to decrease and plateau before reaching the equilibrium state, and is discussed in Chapter 7. The time at which the system reaches this state is after the cells have likely progressed through the four stages of the cell cycle. It is prior to this region that fast-repairing damages are still present inside of the cell. The repair processes begin to repair the damage over time, leaving behind only lethal lesions on the DNA within the cell. The cell's probability of survival decreases when more lethal lesions are present (discussed in Chapter 7) and is governed according to whether we are using a fractional or non-fractional Poisson process. Within the model framework used to generate our results (see Chapter 4), the survival fraction is observed to decrease as more damages are classified as lethal. We observe the evolution of survival fraction within our model timescale, $t \in [t_{del}, t_{rep}]$.

In this section we will observe this time scale behavior for selected low and high-LET conditions by fitting different expressions and assessing the gof. We tested a series of functions on the data, however due to the steep "fall-off" region at small repair times and fast approach to the "plateau" region for higher repair times we required a two-part function (see Fig.(8.35)). A pure exponential relationship is not sufficient to describe the relation because of the steep "fall-off" region and a pure power-law relationship was not sufficient for low doses. Whilst it fit well to repair time versus SF at high doses, the "fall off" region is not steep enough at low doses to be appropriately described by the power-law. Therefore, we fit an exponential-power-law relation to capture the behavior for all doses across all LET. We also trial the Mittag-Leffler distribution (Eq.(8.9)) to compare with the exponential-power law expression (Eq.(8.10)).

$$SF(t_{rep}) = E_{\gamma}(-\lambda t_{rep}^{\gamma}) \tag{8.9}$$

$$SF(t_{rep}) = \frac{e^{at_{rep}^{b}}}{1 + ct_{rep}^{d}} \tag{8.10}$$

Each equation is chosen such that conditions $SF(t_{rep} = 0) = 1$ and $SF(t_{rep} = \infty) = 0$ are satisfied. This is because within our simulations, the repair processes do not commence until irradiation has ceased and all cells are assumed to be alive until $t = t_{del}$. Then the cells are assumed to die for $t >> t_{rep}$ within the confines of our model, because we know that cells will die eventually whether they are damaged by radiation or not. Fits using both equations for SF as a function of repair time are depicted in Fig.(8.36). Figure (8.36(a)) depicts the SF-repair time relation for high and low doses of low-LET radiation and Fig.(8.36(b)) for that of the high-LET region. We observe that the time at which the system approaches equilibrium (i.e. state $|f\rangle$) is smaller at large doses due to the larger number of damages present after irradiation, resulting in a faster accumulation of lethal lesions. For smaller doses the

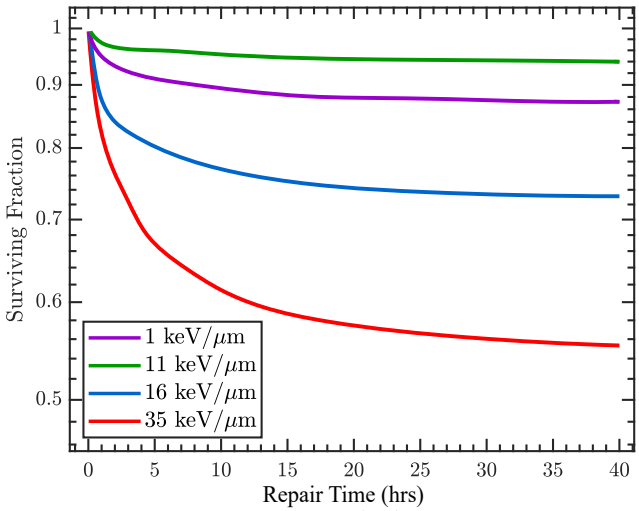
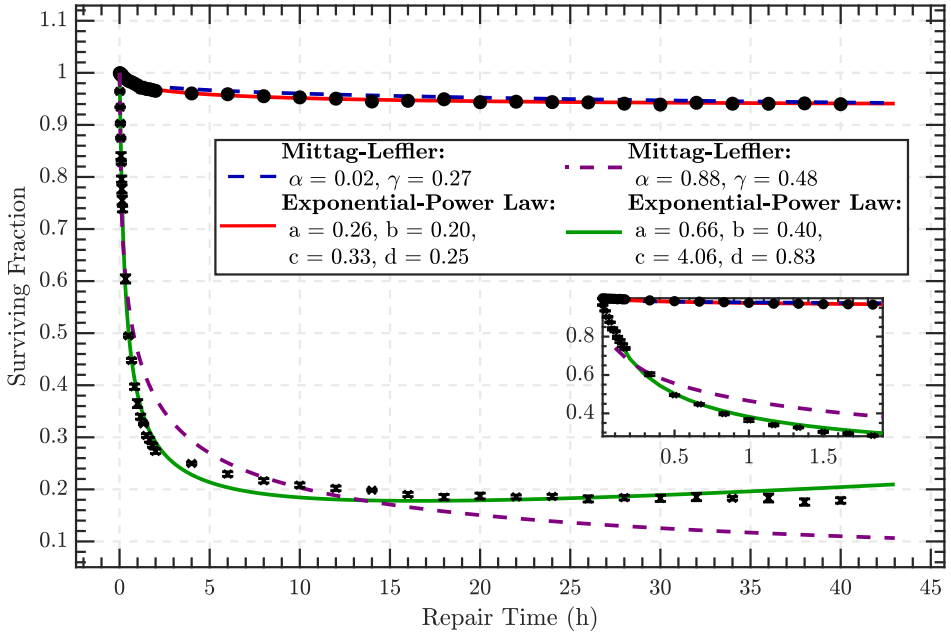


Figure 8.35: The evolution of survival with repair time after the delivery of ≈1Gy to cells for different LET radiation.

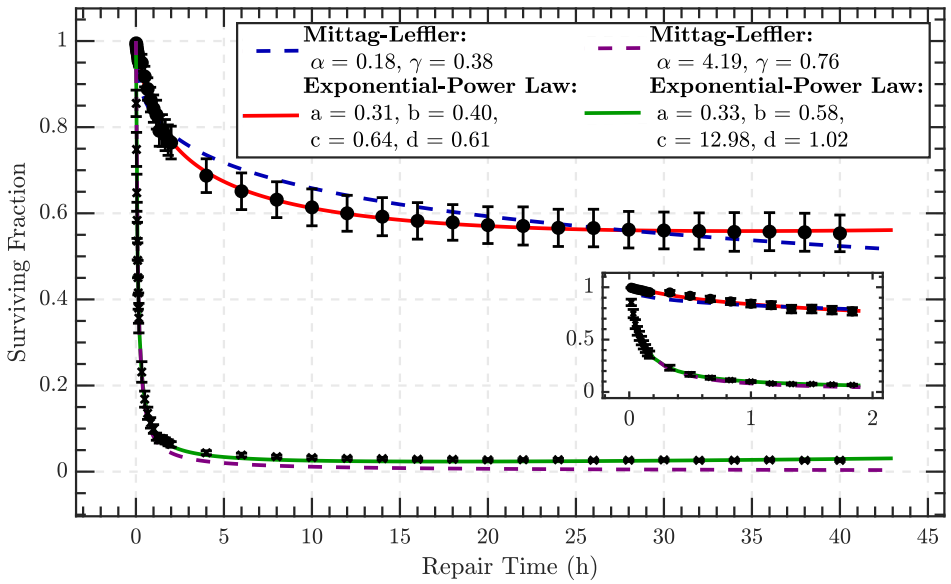
process is more gradual because there are less fast-repairing lesions present at time t_{del}. It is therefore possible that there are multiple stages involved. For short repair times the fast-repairing damages and characterisation of lethal damages is a sudden process resulting in a steep decrease in the survival fraction. At a given time - for example after the cell has progressed through the four cell cycle phases - a phase transition occurs where the accumulation of lethal damages decreases resulting in a "plateau" region of the survival fraction as a function of time. The plateau will occur at different survival fractions according to the absorbed dose and LET of the beam (see Fig.(8.35)) thus warranting a relation that is exponential for small t, with a slower change as $t \to \infty$.

Equation (8.9) demonstrates similar exponential behavior for small repair times followed by a phase transition to the plateau region. However it declines too gradually when the survival fraction is higher (Figs.(8.36(a)) and (8.36(b))). Alternatively, Fig.(8.36(b)) shows that both equations approach the equilibrium state adequately, resulting in a good fit for both instances ($R^2 = 0.996$ for Eq.(8.9) and 0.999 for Eq.(8.10)).

The exponential-power-law equation, Eq.(8.10) is the most effective at describing the evolution of survival fraction with repair time $t \in [t_{del}, t_{rep}]$ because it yields a better fit in the flat region where the SF reaches equilibrium after longer repair times, as shown from the gof metrics in Table (8.11). The relationship is exponential for small t where many fast-repairing lesions are forming lethal lesions and many cells undergo apoptosis as they progress through the four cell cycle phases. As t increases the system begins to sustain permanent damage before coming to equilibrium. Future efforts should focus on describing this system with a single timescale,



(a) Low LET (1.2keV/μm)



(b) High LET (35keV/μm)

Figure 8.36: Fits to our simulated cell survival fractions as a function of repair time for high and low LET proton radiation, at low (∼1Gy) (red and blue) and high (green and purple) doses (∼8Gy), incident on V79 cells.

where $t = 0$ before irradiation begins and the repair time commences when the beam is "switched on" and irradiation begins.

Data	Dose (Gy)	GOF Metric	Exponential Power-Law	Mittag-Leffler
Low-LET	1	SSE	< 0.001	< 0.001
		RMSE	< 0.001	< 0.001
		R^2	0.966	0.942
High-LET	1	SSE	< 0.001	0.043
		RMSE	0.002	0.033
		R^2	0.992	0.963
Low-LET	8	SSE	0.011	< 0.194
		RMSE	0.071	< 0.017
		R^2	0.996	0.934
High-LET	8	SSE	< 0.001	0.011
		RMSE	0.001	0.017
		R^2	0.996	0.994

Table 8.11: GOF metrics for the fits in Fig.(8.36).

8.4 Predicting RBE: FPp versus Pp

In Chapter 5, Sections (8.3.1.3) and (8.3.1.4) we discussed the RBE predicted using fits from the Pp and fPp models to experimental data. Here we will analyse the RBE predictions using our simulated data and test for the expected trends against LET and depth, as we have done previously. We also know the repair times for our simulated data and hence we can compare the evolution of RBE with repair time.

Our expectation from the simulated results of the V79 Chinese Hamster cell line is as per Belli et al. [26], where the optimal RBE is achieved at $\sim$30keV/μm before decreasing for higher LET due to damage saturation effects. Figure (8.37) depicts the expected behavior of the V79 cell line is response to proton and heavy ion radiation. We could not investigate heavy ion radiation due to time constraints so instead we focused on proton radiation - the region where the first peak occurs in Fig.(8.37). However, judging by our promising results thus far, we expect to see a similar trend in heavy ion RBE as we do in Fig.(8.37).

In Section (8.3.3), we examined our fPp fits to simulated data at varying positions within a 20MeV pristine Bragg Peak. We do not possess experimental data for Bragg Peak that is similar in energy and hence we cannot compare these RBE results with experiment. Recall our definition of RBE in Chapter 2, where we measure the RBE of a beam by taking the ratio of endpoints due to two different radiation types (the radiation in question and a reference beam). Photon beams are the most commonly used reference radiation and are therefore used in this investigation (namely a 6 MV photon beam). The RBE was calculated in this investigation by taking the ratio of doses to for our reference and proton beams to achieve a survival fraction of 10% and 37% respectively (Fig.(8.38)).

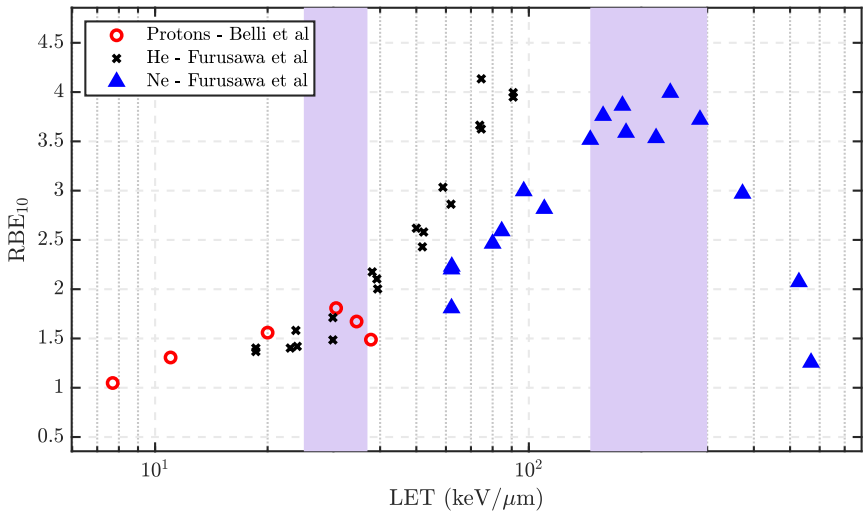


Figure 8.37: The correlation between RBE and particle LET for different radiation types. The RBE measurements in this figure are derived from Belli et al. [26] and Furusawa et al. [131].

8.4.1 RBE as a Function of LET

Figure (8.39) depicts the trend in fPp and Pp predicted RBE from the energy-varied case of our simulated results. Continuing with our discussion on the study from Belli et al. [26], we know that the RBE peaks at an LET $> 30\text{keV}/\mu\text{m}$. This peak is not visible in our simulated data, which is likely the result of not simulating the full physical picture. The RBE peak is a characteristic of different cell lines and is governed by their internal biological response mechanisms to ionising radiation. Our simulations did not include a number of biological conditions and is likely the result of the RBE peak not being visible. In Chapter 5 we did not observe RBE peaks for other cell lines such as Bettega et al. [61]. Furthermore, it is not reasonable to directly compare the RBE values of Belli et al. [26] with our simulated results because we do not know all the biological conditions of the V79 cells used in the experiment. However, the expected trend of increasing RBE with LET is observed in our simulated results and is consistent with experiment - with the exception of the RBE peak at $30\text{keV}/\mu\text{m}$.

Moreover, we have shown in Section (8.3.1) that subtle differences in the fPp and Pp fits can yield very different RBE. Despite our survival measurements being close to that of Belli et al. [26], the fPp and Pp fits are very different. The difference is also the result of the reference beam used in Belli et al. [26], a 200kV photon beam, in comparison with our simulations which use a 6MV photon beam. Overall, we can conclude that the RBE differences are likely the result of different exposure and biological conditions used in our simulations and Belli et al. [26].

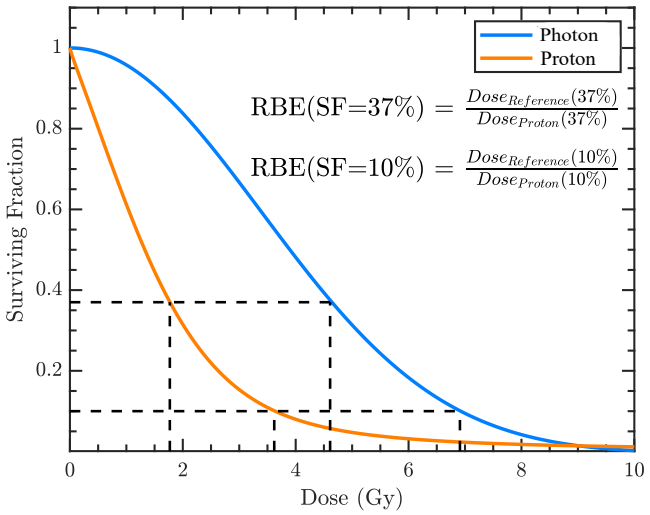


Figure 8.38: An example of the RBE calculations performed in this investigation using fPp fits to our reference photon beam curve and a high-LET proton curve, namely 35keV/μm.

Even though the predicted RBEs are different between our simulations and Belli et al. [26], the expected trend of an increased RBE in the low dose (high survival) region is observed. This is consistent with our observation of less efficient cell killing in the high dose region where the effect is more enhanced for high-LET due to saturation effects (see Section (8.3.1.3)). This is reflected in Fig.(8.39) because the RBE is different between 10% and 37% survival in the low dose region whilst becoming more pronounced in the high-LET region.

8.4.2 RBE as a Function of Depth

Experimental results from Section (8.3.1) and Chapter 5 show that the predicted RBE is strongly correlated with the position of the cells inside the Bragg Peak, which is inherently related to LET. Recall the study of Howard et al. [123], which considered responses from different cell lines of human and rodent origin in different regions of a 71 and 160MeV pristine proton Bragg Peak. Survival was measured when the cells were placed in the entrance, proximal, inter-peak and distal regions of the Bragg Peak. The LET range considered in Howard et al. [123] is small compared to our simulated depth study, however the results are similar in trend where the RBE remains relatively low before the Bragg Peak, then increases within and levels off in the distal region. This same behavior is reflected in our simulated results as depicted in Fig.(8.40). We cannot place experimental data in the figure for comparison as we do not possess any data for a pristine 20MeV Bragg Peak, however it is still reasonable to compare the trend with experiment.

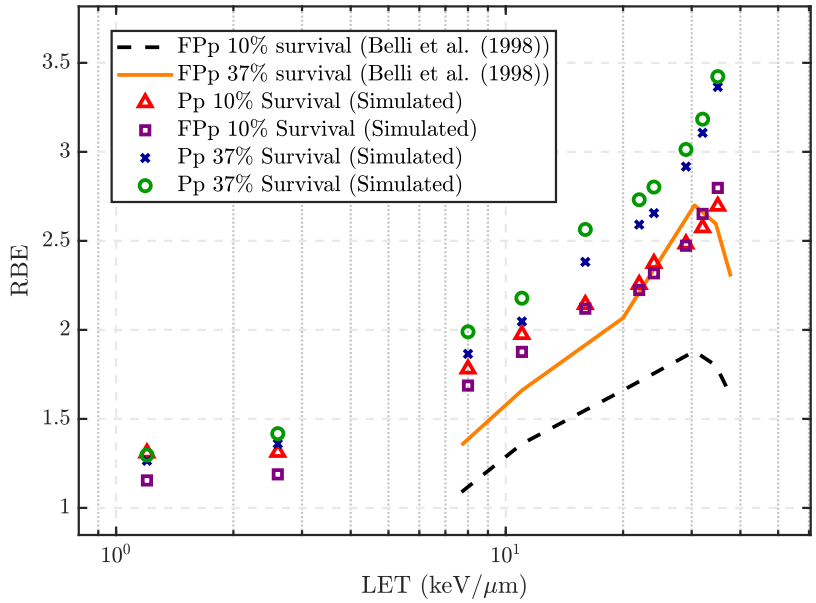


Figure 8.39: RBE as a function of LET from our simulated data with energy-varied LET and the RBE evolution of the data from Belli et al. [26] at survival fractions of 10% and 37%.

Discussions from in Sections (8.3.1.4) and (8.3.3) came to the conclusion that in the distal region of the Bragg Peak, contributions to DNA damage from primary protons is less frequent because most will not propagate to this region. The resultant effect on the RBE is a "leveling off" in the distal region of the Bragg Peak as we observe in our simulated results (see Fig.(8.40)).

8.4.3 RBE as a Function of Repair Time

Figure (8.41) depicts the behavior of RBE at repair times of 2, 16 and 40 hours in the high, intermediate and low-LET regions from our simulated data. We have previously discussed the trend with RBE as a function of LET, however the repair time also clearly has an influence. In the low-LET region there is little change in RBE after 2, 16 and 40 hours of repair time. The difference is more pronounced as LET increases. When we relate the survival fraction to the number of lethal lesions present at repair time t_{rep}, the survival fraction will decrease within the confines of the TLK model, as the lethal lesions accumulate (see Chapter 7 and Section (8.3.5)).

At low repair times the lethal lesion count is low compared with large repair times. Therefore, the biological effect is more enhanced with larger repair times due to the delayed effect of lesions being classified as lethal of mis-repaired and hence in survival fraction as more cells will die via apoptosis as a result of radiation effects. Recalling that the survival fraction undergoes a phase transition to its plateau re-

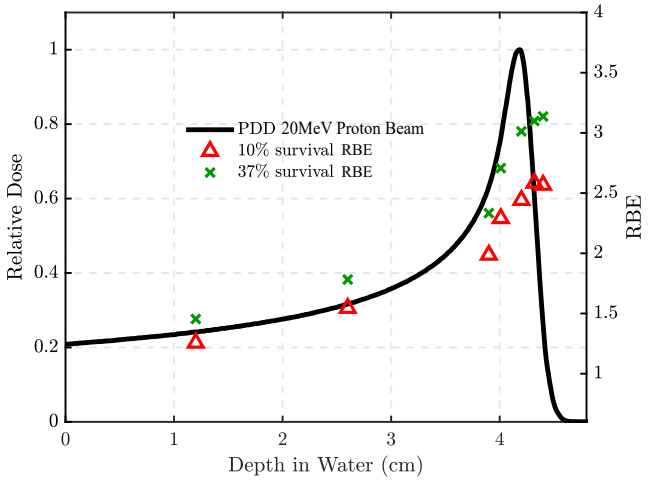


Figure 8.40: The RBE as a function of percentage depth dose (PDD)/Bragg Peak position using our simulated depth study results.

gion (equilibrium) at ≈ 16 hours, the difference in RBE is less pronounced between 16 and 40 hours compared to 2 hours of repair time. In the realistic physical picture, however, one would expect the RBE to decrease with repair time because less time has been allowed for the damage to repair. It would be likely that fast-repairing DNA damages are still present in the system within two hours after irradiation. Whilst at large repair times one would expect slowly-repairing damage only to be present. The discrepancy is due to the TLK model and how it classifies a lesion as lethal for increasing repair time.

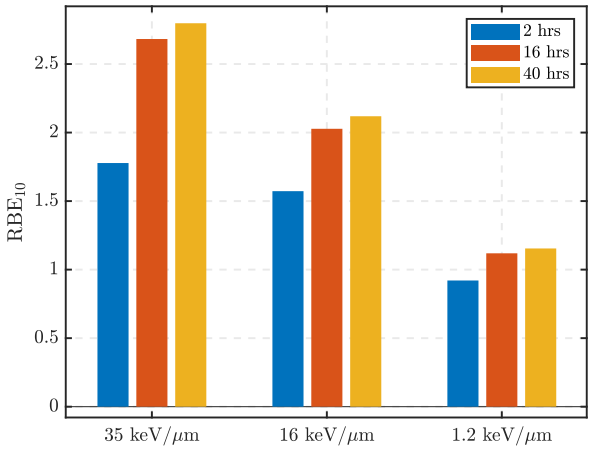


Figure 8.41: The trend in RBE when the survival fraction is measured at different repair times.

Overall, we can conclude that the fPp model is capable of predicting RBE accurately. The trend with other physical parameters such as LET, depth in the Bragg

Peak and repair time are consistent with experiment or can be explained by cell response theory. The fPp and Pp model yield slightly different RBE predictions which is to be expected from the vastly different fit. The fPp model consistently yields a similar or improved goodness-of-fit over the Pp model and hence can be deemed to give the most accurate RBE predictions.

8.5 Evidences of Fractionality

In this chapter, we have explored the fractional Poisson process count model and its ability to describe radiation-induced cell death. Here we will bring together the three key evidences of the fractional Poisson process to describe the lethal damages in the cells caused by radiation. Namely :

I - The Fractional Poisson Process (The Big Picture) : The fractional probabilities for cells containing $k = 1, 2, \cdots$ lethal lesions are given by

$$P_k(\Lambda) = \frac{(-\Lambda)^k}{k!} \sum_{j=0}^{\infty} \frac{(k+j)!}{j!} \frac{(\Lambda)^j}{\Gamma(\gamma(k+j)+1)} \quad .$$

This fractional Poisson count model explains the distribution of cells that contain k lethal lesions significantly better than the Poisson count model, Figs.(8.8) and (8.7), in all LET regions, for all absorbed doses. By considering the full picture, we observe the true behavior of the system in terms of DNA damage repair propagation at equilibrium.

II - The Convex Argument of the Fractional Poisson Process : $\Lambda \equiv \lambda t^{\gamma}$ written in terms of the effective dose-time relation $t(D) = aD + bD^2$ as discussed in Chapter 6 and is calculated from the distribution of lethal damages which includes all probabilities, not only the zero count ones (survival fraction). As required, the calculated Λ is found to be a convex function of effective dose (see Fig.(8.42)). From this data the extracted Mittag-Leffler shape parameter, γ, is 0.82 at LET 35keV/μm which is in excellent agreement with an independent calculation of the dose-response curve, $\gamma = 0.82$ (± 0.01) in Fig.(8.27).

III - The Survival Probability (Zero State ($k = 0$)) of the FPp : is given by the Mittag-Leffler distribution which describes the data with excellent accuracy for all LET ranges, different cell lines and radiation types.

$$\begin{aligned} P_0(t(D)) &= E_{\gamma}\left[-\lambda t^{\gamma}\right] = E_{\gamma}[-\Lambda] \\ &= E_{\gamma}[-(aD + bD^2)^{\gamma}] \quad , \end{aligned}$$

where $\Lambda = -(aD + bD^2)$ and λ is absorbed into constants a and b.

From evidence I, we uncovered the reason for the discrepancies observed in the LQ model in Chapter 5. When a cell is exposed to high-LET radiation, especially high doses, the cell will be saturated with lethal lesions. The effect of lesion saturation on the lethal lesion distribution is an increase in the variance. When the variance exceeds the mean, the Poisson process is no longer valid. The fPp, on the other hand, accounts for situations of overdispersion and underdispersion. In the low-LET region, the Poisson process is still valid and therefore the fPp becomes equivalent with it. For high-LET, especially high absorbed dose, overdispersion is present in the data and therefore the fPp fits much better, as depicted in Figs.(8.8) and (8.7).

From evidence II, we expect the mean lethal lesion count per cell, Λ, to be a monotonic increasing function as explained in Chapter 6. This function is our dose-time relation $\Lambda \equiv \lambda t^{\gamma} \equiv (aD + bD^2)^{\gamma}$ for $E_{\gamma}(-\Lambda)$. We can perform a fit of Λ as a function of absorbed dose where the Λ values are derived from the fPp fits in Fig.(8.7) to our dose-time relation. The resultant fit is presented in Fig.(8.42) with parameters a, b and γ. The γ parameter from the fit is 0.82, implying that we are in the fractional Poisson region which is close to that predicted by the Mittag-Leffler fit to the dose-response measurements of 0.81 as shown in Table (L.3) of Appendix J. In Fig.(8.7), γ is $\sim$0.6 for all doses which is much lower than evidences II and III. However, evidences II and II utilise all the doses to derive γ, whereas evidence I evaluates the fractional probabilities for each dose separately. This shows that our predictions of whether the system is in the fractional or non-fractional Poisson region is consistent across evidences I to III.

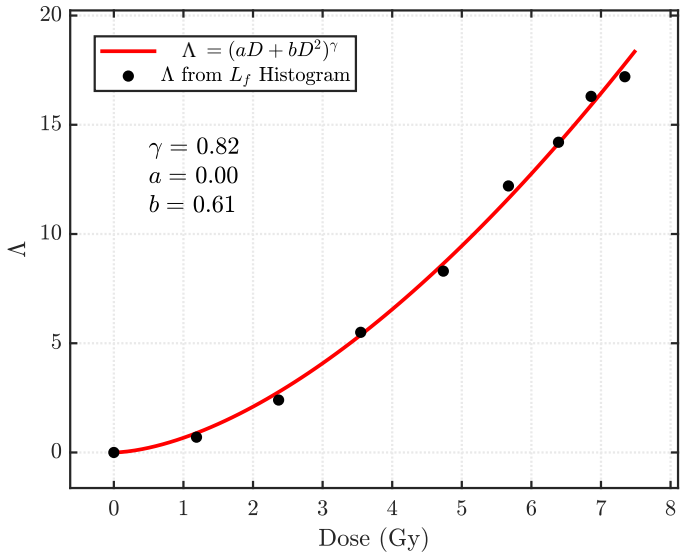


Figure 8.42: A fit to Λ as a function of dose D determined from the fPp distributions of L_f with increasing dose.

Next we combine the results presented throughout this chapter by comparing the cell survival fraction calculations using our fPp model and its zeroth state ($k = 0$) probabilities under the fractional and standard Poisson processes for one set of our simulated dose-response data and experimental data under similar conditions. Namely, our simulated 35keV/μm dose-response data is compared with similar experimental data of Belli et al. [26] (0.64MeV, 34.6keV/μm, V79 cells). We compare fPp model predictions of the survival fraction against the simulated and experimental dose-response curves making use of the following methods ;

1. By fitting the Mittag-Leffler distribution to our simulated cell survival dose-response curve after 40 hours of repair time using the NLLS method and the standard deviation of our measurements as weights, which was performed in Section (8.3.2) (the purple line in Fig.(8.43)).

2. Fitting the Mittag-Leffler distribution to the cell survival data of Belli et al. [26] in a similar LET range to the simulated data (35keV/μm) using a NLLS fit weighted by the SEM, as performed in Section (8.3.1) (the red line in Fig.(8.43)).

3. Using the model parameters predicted by the fPp and Pp models from Fig.(8.7) (the orange and green lines in Fig.(8.43), respectively).

4. By evaluating the Mittag-Leffler function with argument $-\Lambda$ ($E_\gamma(-\Lambda)$), which is taken from fits of the fPp to the normalised histograms of lethal damages (L_f) in Fig.(8.7) (the blue line in Fig.(8.43)).

Figure (8.43) depicts each survival fraction prediction using each of the methods listed above to compare with the experimental and simulated data sets, which are also provided in the figure. Firstly, we observe a very poor fit of the Poisson probabilities from the normalised lethal lesion histogram where most of the cell survival fractions are far below the simulated and experimental data. As shown in previous sections, the Mittag-Leffler distribution yields an excellent fit to high-LET data such as what we consider here. The simulated and experimental ML fits pass through all the points and describe the data very well. When we take the zeroth order probabilities from the fPp ML distribution fits to the histograms of Fig.(8.7) and plot them in Fig.(8.43), the resulting dose-response curve is extremely close to both the simulated and experimental survival fraction measurements. The same behavior is observed by taking the zeroth order probability from the L_f histogram itself. We expect this because of the excellent fit of the ML distribution to the data in Fig.(8.43).

In conclusion, this section has brought the results of testing the fPp model at different endpoints of cell response together to test their ability to describe the measured zeroth order (survival) probabilities for the experimental and simulated cell survival dose-response data. Figure (8.43) shows that the result of fitting the fPp model to the distribution of L_f for each dose D is reflected in the dose-response curve

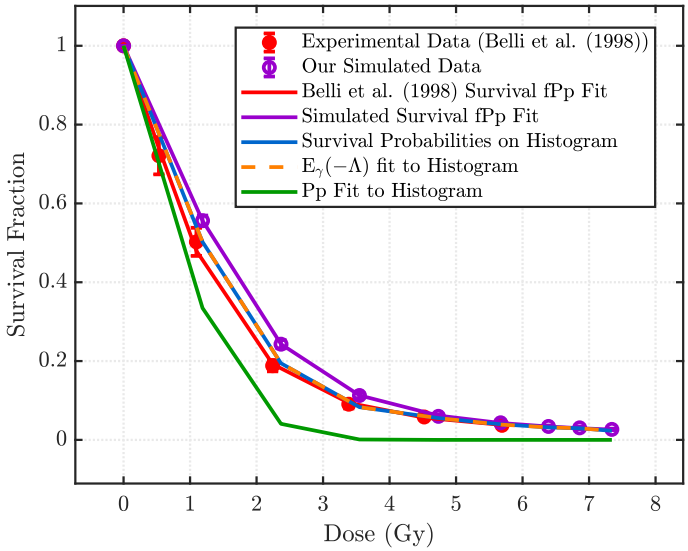


Figure 8.43: Comparison of fPp fits to survival data with the zeroth order probabilities derived from Fig.(8.7).

under simulated and experimental conditions. We expect this from the agreement with the fPp model to the distribution of lethal damages (L_f) and the dose-response curve as shown in Sections (8.2), (8.3.1) and (8.3.3). Figure (8.42) shows that the time-dose arguments are able to accurately predict the fractional parameter Λ from the L_f distribution.

By bringing together the survival probability predictions made at different dose-response endpoints throughout this chapter, there is further evidence of the model's effectiveness. We have demonstrated that the fPp model is able to predict survival and DNA damage probabilities with similar or higher accuracy than the Pp model. The relative performance of the two models is dependent on the LET considered. In the low-LET region, the two models converge to a Poisson process, yielding similar goodness-of-fit. When LET increases, the Poisson process is no longer valid and the fPp model yields the best fit. We have shown that the model is effective at describing select experimental and simulated data, as we could not expand our investigation due to time constraints. Future work will aim to improve that statistics involved by collecting more data on which to test the fPp model. Despite this, what we have observed thus far strongly supports the notion that radiation-induced cell death obeys a fractional Poisson process in the high-LET region where the Poisson process is no longer valid (Chapter 5).

8.6 Conclusions

In this chapter, we tested our new proposed stochastic radiobiological model introduced in Chapter 6 on select experimental and simulated dose-response data. Further evidence that our model is superior in describing the data is provided by considering the fPp at different stages of dose-response, namely the DNA lethal lesion yield and cell survival. From this we observe three key evidences of the fractional Poisson process in describing radiation-induced cell death. We then showed that an improved fit can change the resultant RBE predictions whilst demonstrating trends with LET, depth and dose that are consistent with our expectations from radiation response theory of cells.

Our findings that the fPp model consistently yields superior fit to dose-response data, allowing us to draw the following conclusions :

- Our proposed model explains the data for multiple stages of dose-response in an excellent accuracy including the lethal lesion distributions in cells and survival fraction.
- The fPp model explains cell response in the high-LET region significantly better than the Pp, due to the presence of overdispersion in the high-LET data. Whereas in the low-LET region, the equidispersion condition of the Pp is satisfied. Therefore, in the low-LET region the data is adequately explained by the Pp.
- The flexibility of the fPp model means that it is able to explain data that exists in both the fractional and standard Poisson regions because they become equivalent as the fractionality parameter, γ, approaches 1.
- The model is also able to explain data under different biological conditions such as across different cell lines where the radiation response can differ from the V79 cell.
- The superior fit of the fPp model compared to the LQ model means that it is able to predict the RBE with more accuracy than the LQ model.

Due to constraints with time and computation power, we could not collect a large enough sample size in our simulated data to generate more meaningful errors. Not simulating the full physical picture from experiment resulted in lower γ values than we observed in our fPp fits to the experimental world data. Regardless of these limitations, fPp model testing on the experimental world data and our simulated data shows that there is strong evidence that a cell's response to radiation obeys a fractional Poisson process under certain conditions and our model is able to converge to a Poisson process under low-LET conditions. The result is more accurate modelling in the regions where the LQ model is no longer valid and alternative endpoints

such as the distribution of lethal lesions across a multi-cellular system and RBE can be calculated with more accuracy. Our results are not only important to radiation research and theory, but has clinical relevance to hadron therapy treatment planning for cancer patients.

9

Conclusions

In modern radiation research, hadron therapy has become a popular alternative to photon-based radiation treatments for the deep-seated tumours. It is the unique method of dose delivery that allows focused ionisation directly targeted to the tumour while the surrounding healthy and potentially radiation-sensitive tissues are avoided or receive negligible radiation. Photon-based treatments, on the other hand, ionise the surrounding healthy tissues when irradiating the targeted tumour. When unnecessary doses of radiation are absorbed in healthy tissue, patients can experience health complications and even develop secondary cancers, which can greatly impact their quality of life during and after treatment. The healthy tissue-sparing effect of hadron therapy minimises these complications and long-term health side-effects. It is through careful manipulation of the Bragg Peak that tumours of varying depth and shape can be targeted with high precision.

When irradiating a targeted tumour, it is perturbed by the induction of lesions on the DNA helix within its nucleus and this process is stochastic in nature. The primary mechanism of cell death occurs via breakages to the DNA segments within the nucleus and the cell's probability of survival is correlated with the complexity of the DNA damage. The number of DNA double strand breaks (two DNA breakages in close proximity) and complex DNA double strand breaks (more than two DNA breakages in close proximity) are the main contributors to cell death [11,24] because they can repair incorrectly or not be repairable. DNA lesions are considered to be the primary mechanism for radiation-induced cell death and are produced via direct and indirect effects. Direct effects are those that occur on or within the vicinity of a DNA segment and are primarily the result of ionisation and excitation interactions (physical processes). Indirect effects are induced by products that form away from the DNA helix and propagate to it. Indirect effects are the result of free radicals produced from radiolysis reactions with water (chemical processes). The cell is able to detect when DNA damage is present, at which time the system will try to equilibrate by repairing the DNA damage via a series of repair pathways (biological response). The physical, chemical and biological processes follow a timeline as

shown in Fig.(2.1).

The effect of radiation on the cell can be quantified using a dose-response curve. It describes the probability of cell survival after exposure to a given absorbed dose of radiation. One can compare the effect of a given radiation type compared to another by calculating the relative biological effectiveness (RBE). As defined in Chapter 2, the RBE is defined as the ratio of the doses required for two radiation types to achieve the same biological effect, which is inclusive of survival fraction and DNA damage yield. RBE and survival measurements are clinically relevant because they are applied to treatment planning for cancer patients. The model used to derive these results should be able to adequately explain the data. If not, this could be detrimental to the treatment of cancer patients, leading to inaccurate dose calculations and treatment planning.

Many models have been developed and proposed since the first target theories were introduced in the 1970s [34, 173], including the Repair-Misrepair (RMR) [31], Lethal-Potentially Lethal (LPL) [32] and Two-Lesion Kinetic (TLK) [33] models. The Linear-Quadratic (LQ) model [23, 35] is the most widely used and accepted radiobiological model in the world and is based on the Poisson count process. The model contains two parameters, α and β, which parameterises the contributions of lethal lesions due to single and multiple "hits", respectively. It is an appealing model to researchers due to its simplicity, however it contains a small parameter space which "constricts" the model so that not all aspects of radiation-induced cell death can be explained using its parameters. In fact, the LQ model has very little predictive power in that the parameters cannot be chosen to make predictions of cell response under given conditions. It can be fit to the data and used to make inferences of dose-response under the same conditions, however the result of such a fit cannot be used to make predictions if any of these conditions are altered, i.e. α and β are not global. The common thread through all the established radiobiological models in the literature is the assumption that DNA damage production obeys a Poisson count process.

In this study, we explore the limits of the Linear-Quadratic model by fitting it to experimental world data of clonogenic cell survival from the literature. We obtained the data from Professor Cynthia Keppel and Dr. Pawel Ambrozewicz of the particle physics group at the Thomas Jefferson National Accelerator Facility in the United States and the Particle Irradiation Data Ensemble [46]. The data contains dose-response curves obtained experimentally under different radiation exposure and biological conditions including, but not limited to, incident proton energy, particle LET and cell line. We began by performing a *regression analysis* on the fits by testing against the key assumptions of regression modelling. To test the assumptions we made use of error (residual) analysis, goodness-of-fit analysis and hypothesis testing.

Our findings led to the following conclusions :

- The published fits in the literature suffer from several deficiencies under regression and goodness-of-fit analysis.
- Inaccurate fitting of the data can lead to incorrect conclusions being drawn from the results, including RBE predictions (see Fig.(5.17(b))).
- Making use of hypothesis testing, the residual errors of the Linear-Quadratic fits are not Normally distributed in most cases (see Tables (F.1) and (F.4) for the results).
- The regression fits reveal that the data is highly correlated (the variance of residual errors is not constant).
- The Linear-Quadratic model has a limited range of validity. It can adequately describe that data under low-LET exposure conditions but as the LET increases, the fits gradually worsen.
- The dense ionisation track of high-LET radiation cannot be suitably explained by the Linear-Quadratic model. It fails to adequately explain the dose-cell response relation in the high-LET region when lethal DNA lesion complexity and DNA damage interactions are raised. Therefore, the corresponding RBE value would be effected when the LQ model is used.

To understand the cell survival dose-response relations and to conduct a more accurate error analysis in a systematic way, we simulated cells irradiated with a proton beam by making use of the Monte Carlo software toolkits Geant4 [75] and TOPAS [47] to generate a collection of dose-response data under varied irradiation conditions. We chose the V79 Chinese Hamster cell line for our study due to its prevalence in radiobiological research and availability of cell-specific information in the literature [26, 91, 107]. The simulations consisted of 530 V79 cells placed inside a box, exposed to increasing doses of proton radiation at different beam energies (hence different LET). The cells were approximated to be spheres containing three organelles, a nucleus, nucleolus and cytoplasm each composed of liquid water. Due to time and computation constraints, we could only simulate direct effects inside the cells which were implemented using the Geant4-DNA [155] and Livermore low-energy physics models.

The output from our TOPAS simulation was used to cluster the ionisation events and characterise the DNA damage induced in each cell. The repair processes that the cells undergo after irradiation ceases were simulated using the TLK model [33] where each cell's survival probability is determined according to the number of lethal damages present. Both the clustering and repair models used to collect our results are based on those from [48,87] with some modifications. Due to the large processing time (1–3 days) per simulation and computation power required, we could perform

30 independent measurements (repetitions) per delivered dose for each beam energy.

We showed that our simulated results are consistent with the literature on DNA damage yields, Fig.(7.12) and dose-response theory, Fig.(7.16), giving us reasonable grounds to perform our investigation into the limits of the Linear-Quadratic model and also testing other dose-response models which are frequently used. Even though we did not simulate indirect DNA damage effects, our ionisation and DNA damage yields are within the experimental range according to particle LET, Fig.(7.12). Moreover, our cell survival fraction calculations are consistent with the world experimental data and demonstrate the expected trend with particle LET and repair time Figs.(7.16).

Our simulated cell survival dose-response data for irradiation of the V79 Chinese Hamster cell line showed that the distribution of lethal DNA lesions per cell violates the equidispersion condition[1] of the standard Poisson process and instead demonstrates overdispersion[2], Fig.(6.4), particularly for the mid to high-LET region. We observed that in the low-LET region, the equidispersion requirement of the Poisson count process is reinstated. Concluding that the underlying count process for lethal damages is only consistent with the Poisson distribution in the low-LET region and in the high-LET region they cannot be described by the standard Poisson process.

We proposed a new stochastic model to explain cell response to radiation for all LET regions, whilst overcoming these discrepancies. The model we propose substitutes the Poisson count process for a fractional Poisson count process for describing significantly better lethal damage counts. The fPp is a generalisation of the standard Poisson process and when the fractionality parameter, γ goes to 1 ($\gamma \to 1$) it approaches to a standard Poisson process. The fractional Poisson process is much more flexible and capable of explaining overdispersed and underdispersed data in addition to equidispersion.

Our proposed model is tested on both the experimental world cell survival data and our simulated data exhaustively, using rigorous statistical methods. Since the cell survival fraction is the zero count probability of the fractional Poisson process of lethal damages, naturally the cell survival probability will be defined by the Mittag-Leffler distribution with a convex argument of absorbed dose. We directly compare the performance of our new model against the Linear-Quadratic model to assess whether the fractional Poisson process better explains the data, particularly in the mid to high-LET region where the Linear-Quadratic model is problematic.

[1]The mean and variance are equal.
[2]The variance is greater than the mean.

Our findings can be summarised as :

- The relative complexity of each DNA break cluster inside the cells (i.e. damage clusters that contain $k = 1, 2, \cdots, n$ DNA breakages) can be explained by exponential decay Fig.(8.6). The relative number of damage clusters with k damages increases, however the decay rate remains the same ($\sim$0.63) regardless of the absorbed dose.

- The distribution of lethal lesions per cell is extremely overdispersed in the high-LET region and hence is described remarkably better by the fractional Poisson process in comparison to the Poisson distribution, Fig.(8.7). Furthermore, overdispersion increases with absorbed dose which is shown in Fig.(8.7), where the fractional and standard Poisson processes are in closer agreement at low doses but the disagreement grows bigger for high doses. Whereas in the low-LET region, the fractional and standard Poisson processes agree the most.

- The convex argument ($\Lambda = -\lambda t^{\gamma} = -(aD + bD^2)^{\gamma}$) of the fPp is calculated from the distribution of lethal damages which includes all probabilities, not only the zero count ones (survival fraction), Fig.(8.42). From this quantity, the extracted γ parameter is found to be $\sim$0.82 at LET 35keV/μm which is in excellent agreement with independent calculations of $\gamma = 0.82$ ($\pm$0.01) in Fig.(8.27).

- The Mittag-Leffler distribution, which corresponds to the fractional Poisson process, gives superior agreement in all LET regions for both the experimental and simulated data, as shown in Figs.(8.10)–(8.12) and (8.26)–(8.31). Moreover, it gives consistent results with the Poisson distribution in the low-LET region. An example of the remarkable fit in the high-LET region is shown in Fig.(9.1) to our simulated data with experimental data present for comparison.

- The fractional Poisson process approaches to the standard Poisson process in the low-LET region, where the fractionality parameter of the fPp model (γ) was shown to approach 1. Whilst the fractional parameter in the high-LET region consistently moves away from 1, implying that the fractional Poisson process is dominant. In the final state of the system (in equilibrium), we observed that the fPp is dominant for high-LET radiation at multiple endpoints of dose-response (lethal DNA lesion distributions and survival). The survival probabilities obtained from the outcome of the lethal lesion distribution fits and ones from the dose-response curve fits are in superb agreement when plotted against dose-response data, Fig.(8.43).

- Our new model is able to predict expected trends in RBE versus LET, depth in tissue and absorbed dose. This superior fit of our model to experimental and simulated dose-response data implies that it can detect RBE trends with higher accuracy than the Linear-Quadratic model, especially in the high-LET region.

- To model the cell survival fraction as a function of repair time, we tested different relations and assessed the goodness-of-fit. We observed that an exponential-power law relation, Eq.(8.10), yields the best fit to the data for the entire LET range, for all absorbed doses. The exponential relationship is dominant for small repair times where fast-repairing damages (simple DSB) are still present in the cells and the cell survival fraction decreases very quickly. As the cells approach their equilibrium state for larger repair times, only slow-repairing damages are present which results in the cell survival fraction trend flattening, this is where the power-law is dominant.

.

By demonstrating that our model accurately predicts the dose-response relations including survival fraction, DNA damage yield and RBE, we have highlighted the importance of accurate radiobiological modelling in hadron therapy. We have shown the implications of using an un-suitable model of dose-response such as the incorrect prediction of RBE. Our results are not only of importance to radiation research but are also of clinical relevance. The use of dose-response and RBE modelling in treatment planning for cancer patients is detrimental to the success or failure of their treatment. Accurate modelling in hadron therapy is required to justify the costs of building and maintaining the equipment to make the treatment available to cancer patients worldwide.

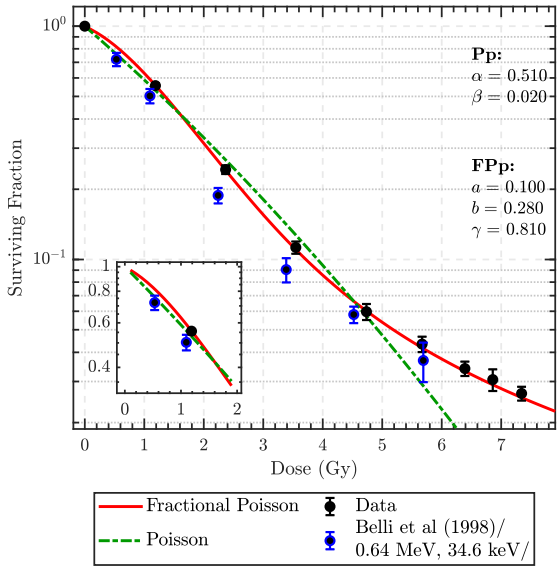


Figure 9.1: A comparison of the Linear-Quadratic model under the standard (Pp) and fractional Poisson processes (FPp) fitted to our simulated high-LET data (35keV/μm) with experimental data from Belli et al. [26] under similar LET conditions plotted for comparison.

9.1 Future Work

Our results have shown that there is strong evidence of the fractional Poisson process in the system in the high-LET region, however there are still many parameters we were unable to consider in our investigations due to time constraints. We have acknowledged the areas of improvement throughout the thesis, all of which will be the focus of future efforts and are summarised below.

- **Increase measurement repetitions and sample size :**
  A maximum of 30 independent measurements (repetitions) could be performed for a single data point due to time constraints. As a result adequate statistics could be achieved in our results. This is evidenced by some of the fluctuations in our data throughout Chapters 7 and 8. The small sample size with respect to hypothesis testing made the pass rates of said test very high. To gain better insight into the performance of each model, the sample size should be drastically increased - this is perhaps not as achievable experimentally by via Monte Carlo methods.

- **Test our model under different physical, chemical and biological conditions:**
  In order to achieve more conclusive results, data where all conditions are known and variable should be used in future model testing. This may be achievable by further optimising or developing a new Monte Carlo simulation where more parameters can be tested. The conditions of particular interest are;

  - different radiation types (such as alpha particles and heavy ions).
  - different cell lines and species of origin (human and animal).
  - homogeneous cell cycle phases.
  - varied oxygen concentrations.
  - wider LET and energy range (achievable using other radiation types).
  - with indirect effects implemented (free charged radicals).
  - realistic organelle materials in cells (achievable when Geant4-DNA is available in non-liquid water phantoms).
  - methods of beam delivery (pencil beam and Spread-Out Bragg Peak).

- **Extend the model to more endpoints :**
  Thus far we have considered survival and lethal lesion yield as fractional Poisson processes. Future efforts should focus on using the model to predict additional stages of dose-response such as;

  - DNA damage complexity yield
  - repair time beginning from the start of irradiation.

  - lesion repair processes.

- **Improve simulation time scale :**
  In the experiment, the repair process commences during irradiation. In our simulations we assume that the cells are intact and don't begin to repair until irradiation has ceased. Future Monte Carlo simulations should aim to perform clustering during beam delivery at select times to make measurements more consistent with experiment.

- **Realistic geometries :**
  Our Monte Carlo simulations approximate the cellular geometry as a sphere, whilst the reality is far more complicated. The DNA segments inside of the cell are also not explicitly simulated, but a scaling is performed to determine the number of strand breaks. Future efforts should focus on generating results on a realistic DNA segment which is achievable in Geant4-DNA.

- **Interpret the parameters in light of radiobiology :**
  The vast number of parameters involved in radiation-induced cell death makes it difficult to account for everything in a single model. However, the parameters involved in describing DNA damage and survival should be given radiobiologically relevant definitions and weightings.

- **Analyse inter-arrival time distributions :**
  A property of the fractional Poisson process is that the inter-arrival times between DNA breaks are non-exponentially distributed but Mittag-Leffler distributed. We plan to simulate the inter-arrival times to test alternative probability distributions such as the Weibull distribution.

- **Try alternative dose-time relations :**
  We have commenced testing of the fPp model using dose-time relations alternative to the linear-quadratic expression, which were not ready to be presented in this thesis. Publishing our findings from this study is underway, in which alternative dose-time relations will be discussed.

Appendix A

TOPAS-nBio Cell Culture Code

The file TsCellCulture.cc in the TOPAS-nBio [84] source code had to be edited to output the positions of the randomly placed cells into a CSV file. The changes made to do this are written in red. The changes made to add the Nucleolus organelle to each cell is written in blue.

```
1         // Component for TsCellCulture
2         // ********************************************************
3         // This file is part of the TOPAS-nBio extensions to the
4         // TOPAS Simulation Toolkit.
5         // The TOPAS-nBio extensions are freely available under the
              license
6         // agreement set forth at: https://topas-nbio.readthedocs.
              io/
7         // ********************************************************
8
9         // A simple cell culture consisting of random spherical
              cells.
10
11        #include "TsCellCulture.hh"
12        #include "TsParameterManager.hh"
13        #include "G4VPhysicalVolume.hh"
14        #include "G4Orb.hh"
15        #include "G4Box.hh"
16        #include "G4Ellipsoid.hh"
17        #include "G4SystemOfUnits.hh"
18        #include "G4PhysicalConstants.hh"
19        #include "Randomize.hh"
20        #include <iostream>
              #include <fstream>
              #include <string>
              #define G4endl std::endl
21
22        TsCellCulture::TsCellCulture(TsParameterManager* pM,
              TsExtensionManager* eM, TsMaterialManager* mM,
              TsGeometryManager* gM,
23        TsVGeometryComponent* parentComponent, G4VPhysicalVolume*
              parentVolume, G4String& name) :
```

```
24          TsVGeometryComponent(pM, eM, mM, gM, parentComponent,
               parentVolume, name)
25          {
26          ResolveParameters();
27          }
28
29          TsCellCulture::~TsCellCulture()
30          {;}
31
32          void TsCellCulture::ResolveParameters() {
33          HLX = fPm->GetDoubleParameter(GetFullParmName("
               Container_HLX"), "Length");
34          HLY = fPm->GetDoubleParameter(GetFullParmName("
               Container_HLY"), "Length");
35          HLZ = fPm->GetDoubleParameter(GetFullParmName("
               Container_HLZ"), "Length");
36
37          CellRadius = fPm->GetDoubleParameter(GetFullParmName("
               CellRadius"), "Length");
38          NbOfCells  = fPm->GetIntegerParameter(GetFullParmName("
               NumberOfCells"));
39          NuclRadius = fPm->GetDoubleParameter(GetFullParmName("
               NucleusRadius"), "Length");
40          NucleolRadius =
               fPm->GetDoubleParameter(GetFullParmName("NucleolRadius"),
               "Length");
41          }
42
43          G4VPhysicalVolume* TsCellCulture::Construct()
44          {
45          BeginConstruction();
46
47          //*******************************************************
48          //          Envelope Geometry : Rectanglar container
49          //*******************************************************
50
51          G4Box* gBox = new G4Box(fName, HLX, HLY, HLZ);
52          fEnvelopeLog = CreateLogicalVolume(gBox);
53          fEnvelopePhys = CreatePhysicalVolume(fEnvelopeLog);
54
55          //*******************************************************
56          //          Cell geometry : spherical
57          //*******************************************************
58          //Cell geometry
59          G4Orb* gCell = new G4Orb("cell", CellRadius);
60          G4LogicalVolume* lCell = CreateLogicalVolume(gCell);
61
62          //*******************************************************
63          // Optional : include a organelles in the cell
64          //*******************************************************
65
66          // Nucleolus
```

```
67          G4String subComponentName2 = "Nucleolus";
               G4Orb* gNucleolus = new G4Orb("gNucleolus", NucleolRadius);
               G4LogicalVolume* lNucleolus =
               CreateLogicalVolume(subComponentName2, gNucleolus);

68          // Nucleus
69          G4String subComponentName1 = "Nucleus";
70          G4Orb* gNucleus = new G4Orb("gNucleus", NuclRadius);
71          G4LogicalVolume* lNucleus = CreateLogicalVolume(
               subComponentName1, gNucleus);
72
73          std::ofstream myfile("cells.csv");
74
75          // Randomly place cells in the volume
76          for (int j = 0; j < NbOfCells; j++){
77
78          G4bool Overlap = true;
79          while (Overlap == true){
80
81          G4double phi = 0;
82          G4double psi = 0;
83          G4double x = 0.0;
84          G4double y = 0.0;
85          G4double z = 0.0;
86
87          x = (2*G4UniformRand()-1)*(HLX-CellRadius) ;
88          y = (2*G4UniformRand()-1)*(HLY-CellRadius) ;
89          z = (2*G4UniformRand()-1)*(HLZ-CellRadius) ;
90
91          G4ThreeVector* position = new G4ThreeVector(x,y,z);
92          G4ThreeVector* posNucl = new G4ThreeVector(0*mm,0*mm,0*mm);
93          G4ThreeVector* posNucleol = new G4ThreeVector(0*mm,0*mm,0*mm);
94
95          G4RotationMatrix* rotm = new G4RotationMatrix();
96
97          rotm->rotateX(psi);
98          rotm->rotateY(phi);
99
100         G4VPhysicalVolume* pCell = CreatePhysicalVolume("Cell", j,
               true, lCell, rotm, position, fEnvelopePhys);
101         G4VPhysicalVolume* pNucleus = CreatePhysicalVolume("Nucleus
               ", j, true, lNucleus, rotm, posNucl, pCell);
102         G4VPhysicalVolume* pNucleol = CreatePhysicalVolume("
               Nucleolus", j, true, lNucleolus, rotm, posNucleol,
               pNucleus);
103
104         G4bool OverlapCheck = pCell->CheckOverlaps();
105
106         if (OverlapCheck == false){
107         myfile << x << "," << y << "," << z << "\n";
108         break;}
109         if (OverlapCheck == true){
```

```
110         pCell = NULL;
111         pNucleus = NULL;
112         pNucleol = NULL;
113         G4cout << "**** Finding new position for volume Cell : " <<
                j <<  " ****" << G4endl;
114                         }
115                 }
116         }
117         myfile.close();
118         InstantiateChildren(fEnvelopePhys);
119
120         return fEnvelopePhys;
121         }
```

Appendix B

TOPAS Parameter File Used in the Simulations

An overview of the TOPAS [47] parameter file used in our simulations.

```
#=================================
#       GEOMETRY & MATERIALS
#=================================

s:Ma/G4_WATER_MODIFIED/CloneFromMaterial = "G4_WATER"
d:Ma/G4_WATER_MODIFIED/CloneWithDensity  = 1.1 g/cm3

#---------------WORLD---------------
d:Ge/World/HLX      = 10000. um
d:Ge/World/HLY      = 10000. um
d:Ge/World/HLZ      = 10000. um
s:Ge/World/Material = "Air"

#-----------WATER MEDIUM------------
s:Ge/GrowthMedium/Type     = "TsBox"
s:Ge/GrowthMedium/Material = "G4_WATER_MODIFIED"
s:Ge/GrowthMedium/Parent   = "World"
d:Ge/GrowthMedium/HLX      = 305. um
d:Ge/GrowthMedium/HLY      = 305. um
d:Ge/GrowthMedium/HLZ      = 25. um

#-----------CELL CULTURE-------------
s:Ge/MyCulture/Type                = "TsCellCulture"
s:Ge/MyCulture/Material            = "G4_WATER"
s:Ge/MyCulture/Parent              = "GrowthMedium"
d:Ge/MyCulture/Container_HLX       = 300. um
d:Ge/MyCulture/Container_HLY       = 300. um
d:Ge/MyCulture/Container_HLZ       = 20. um
i:Ge/MyCulture/NumberOfCells       = 530
#d:Ge/MyCulture/TransZ             = -1100. um # Change
   position of culture in world - used in depth study.
d:Ge/MyCulture/CellRadius          = 8.5 um
```

```
32          s:Ge/MyCulture/AssignToRegionNamed = "
               DefaultRegionForTheWorld" # Assign Livermore physics
               model to cytoplasm only
33
34          d:Ge/MyCulture/NucleusRadius       = 7 um
35          s:Ge/MyCulture/Nucleus/Material    = "G4_WATER"
36          s:Ge/MyCulture/Nucleus/AssignToRegionNamed = "G4DNA" #
               Assign Geant4-DNA physics model to nucleus only
37
38          d:Ge/MyCulture/NucleolRadius      = 5 um
39          s:Ge/MyCulture/Nucleolus/Material = "G4_WATER_MODIFIED" #
               Water with scaled-density
40          s:Ge/MyCulture/Nucleolus/AssignToRegionNamed = "
               DefaultRegionForTheWorld" # Assign Livermore physics
               model to nucleolus only
41
42          #=================================
43          #  SOURCE - BEAM CHARACTERISTICS
44          #=================================
45
46          s:So/BeamSource/Type                        = "Beam"
47          s:So/BeamSource/Component                   = "BeamPosition"
48          s:So/BeamSource/BeamParticle                = "Proton"
49          d:So/BeamSource/BeamEnergy                  = 20 MeV
50          u:So/BeamSource/BeamEnergySpread            = 0.25
51          s:So/BeamSource/BeamShape                   = "Rectangle"
52          s:So/BeamSource/BeamPositionDistribution = "Flat"
53          s:So/BeamSource/BeamAngularDistribution  = "Flat"
54          d:So/BeamSource/BeamAngularCutoffX          = 10 deg
55          d:So/BeamSource/BeamAngularCutoffY          = 10 deg
56          s:So/BeamSource/BeamPositionCutoffShape  = "Rectangle"
57          d:So/BeamSource/BeamPositionCutoffX         = 410 um
58          d:So/BeamSource/BeamPositionCutoffY         = 410 um
59          d:So/BeamSource/BeamAngularSpreadX          = 0.01 deg
60          d:So/BeamSource/BeamAngularSpreadY          = 0.01 deg
61          i:So/BeamSource/NumberOfHistoriesInRun   = 50000
62
63          #=================================
64          #        PHYSICS & CHEMISTRY
65          #=================================
66
67          d:Ph/Default/SetProductionCutLowerEdge = 250 eV
68          d:Ph/Default/EMRangeMin                = 10. eV          #
               Minimum for EM tables
69          d:Ph/Default/EMRangeMax                = 500. MeV        #
               Maximum for EM tables
70          i:Ph/Default/EMBins                    = 77              #
               Number of bins for EM tables
71          i:Ph/Default/EMBinsPerDecade           = 7               #
               Number of bins per decade for EM tables
72          b:Ph/Default/Fluorescence              = "True"
73          b:Ph/Default/Auger                     = "True"
```

```
74          b:Ph/Default/AugerCascade               = "True"
75          b:Ph/Default/DeexcitationIgnoreCut      = "True"
76          b:Ph/Default/PIXE                       = "True"
77
78          sv:Ph/Default/Modules = 1 "g4em-livermore" # "g4em-dna-
               chemistry"    # Enable Livermore in world
79          s:Ph/Default/ForRegion/G4DNA/ActiveG4EmModelFromModule = "
               g4em-dna"  # Enable Geant4-DNA physics model for region
80
81          #====================================
82          #               SCORERS
83          #====================================
84
85          #-------------------pTuple Scorer--------------------
86
87          #------------------General Settings------------------
88          s:Sc/pTuple/Quantity                  = "Tuple"
89          s:Sc/pTuple/Component                 = "MyCulture"
90          s:Sc/pTuple/OutputType                = "ASCII"
91          b:Sc/pTuple/OutputToConsole           = "True"
92          s:Sc/pTuple/IfOutputFileAlreadyExists = "Overwrite"
93          s:Sc/pTuple/OutputFile                = "pTuple"
94          b:Sc/pTuple/PropagateToChildren       = "True"
95          b:Sc/pTuple/OutputAfterRun            = "True"
96
97          #-------------Track Information-------------
98          b:Sc/pTuple/IncludeChemicalTrack        = "False"
99          b:Sc/pTuple/IncludeParticleName         = "True"
100         b:Sc/pTuple/IncludePhysicalTrack        = "True"
101         b:Sc/pTuple/IncludeEventID              = "False"
102         b:Sc/pTuple/IncludeTrackID              = "False"
103         b:Sc/pTuple/IncludeParentID             = "False"
104         b:Sc/pTuple/IncludeStepNumber           = "False"
105         b:Sc/pTuple/IncludeGlobalTime           = "False"
106
107         #-------------Process Information-------------
108         b:Sc/pTuple/IncludeEnergyDeposited = "True"
109         b:Sc/pTuple/IncludeKineticEnergy   = "False"
110         b:Sc/pTuple/IncludePhysicalProcessName  = "True"
111
112         #-------------Geometry Information-------------
113         b:Sc/pTuple/IncludeVolumeName       = "True"
114         b:Sc/pTuple/IncludeVolumeCopyNumber = "True"
115         b:Sc/pTuple/IncludeVertexPosition   = "False"
116
117         #====================================
118         #           MISCELLANEOUS
119         #====================================
120         b:Ts/PauseBeforeQuit = "False"
121
122         #--------------Verbosity--------------
123         i:Tf/Verbosity = 0
```

```
124         i:Ts/ChemistryVerbosity = 0
125
126         #------------Optimisation-------------
127         b:Ts/ShowCPUTime                = "True"
128         i:Ts/NumberOfThreads            = 10
129         b:Ts/BufferThreadOutput         = "False"
130         i:Ts/ShowHistoryCountAtInterval = 10000
131
132         #-----------Time of Irradiation--------
133         d:Tf/TimelineStart           = 0. s # defaults to zero
134         d:Tf/TimelineEnd             = 330. s # must be larger than
                TimelineStart
135         i:Tf/NumberOfSequentialTimes = 20 # Forces beam delivery
               from TimelineStart to TimeLineEnd
136
137         #------------Random Seed---------------
138         i:Ts/Seed/100
```

Appendix C

Useful Bash Scripting Commands

The size of the files generated in this thesis are extremely large, meaning that post-processing could not be performed entirely with MATLAB. Post-processing of the data generated in this thesis is cumbersome and requires large amounts of computing power with the entirety of our data occupying 30TB of storage space. Bash scripting is a method used to process and derive information from file-systems such as those in this thesis. It is made up of a set of commands that can be applied in a Unix environment allowing formatting and processing of files to be performed in the Unix terminal environment. This Appendix provides a series of useful commands used in this thesis during the post-processing of data generated by the TOPAS toolkit. Note: in this section a generic command will be written as ≪COMMAND≫ and a file will be indicated as ≪filename≫.

- **Writing Shell scripts:** Multiple Bash commands can be entered into a single file to form a small "program" called a Shell script (extension `.sh`). We can turn this into an executable by entering the command that enables executable permissions

  `chmod +x` ≪filename≫.

  The first line of a Shell script should always tell the Unix environment which shell the script will run in, commonly this is the Bourne shell activated by `#/bin/bash`. To run the executable, simply enter `./filename.sh`.

- **Appending the output of a command to the end of a file:**

  ≪COMMAND≫ `>>` ≪filename≫

  To overwrite the file, should it already exist, use > in place of >>.

- **Search for a string within a file:**

```
grep "string" ≪filename≫.
```

- **Print select columns of ≪filename≫:**

```
awk '{print $[COLUMN NUMBER]}' ≪filename≫.
```

- **Concatenate a file to the end another file:**

```
cat ≪filename≫ >> ≪another filename≫
```

- **Print string to the end of a file:**

```
echo "string" >> ≪filename≫
```

- **Select n lines of a file using random permutations:**

```
shuf -n [NUMBER] ≪filename≫
```

Appendix D

Experimental World Dose-Response Data Overview

This appendix contains a summary of the experimental results considered in Chapters 5, 7 and 8 in this thesis. We were granted access to these results by Professor Cynthia Keppel and Dr. Pawel Ambrozewicz of JLab group in the preliminary stages of the project and additional information of these studies were provided in the PIDE [46]. The table includes the primary author, radiation exposure conditions and cellular conditions. **Note:** Human cell lines are denoted by a dagger (†), whilst the remaining are animal cell lines. All of these experiments are assumed to have been performed with asynchronous cells (i.e. the cultured cells are in different stages of the cell cycle) according to [46].

Table D.1: The experimental data considered in this thesis.

Study	Cell Line	Energy (MeV)	LET (keV/μm)	Particle	Depth (mm)	Comments
Ando et al. 2001 [174]	NB1RGB†	235	-	Proton	-	Peak of SOBP in water
Ando et al. 2001 [174]	NB1RGB†	235	-	Proton	-	Plateau of SOBP in water
Ando et al. 2001 [174]	SCC61†	235	-	Proton	-	Peak of SOBP in water
Ando et al. 2001 [174]	SCC61†	235	-	Proton	-	Plateau of SOBP in water
Ando et al. 2001 [174]	V79	235	-	Proton	-	Peak of SOBP in water
Ando et al. 2001 [174]	V79	235	-	Proton	-	Plateau of SOBP in water
Antoccia et al. 2009 [175]	HFFF2†	28.5	0.8	Proton	-	
Belli et al. 1989 [127]	V79	23	1.16	Proton	-	
Belli et al. 1989 [127]	V79	17.5	1.70	Proton	-	
Belli et al. 1989 [127]	V79	10.5	3.36	Proton	-	
Belli et al. 1998 [26]	V79	5.01	7.7	Proton	-	Incident energy at the cell surface
Belli et al. 1998 [26]	V79	3.2	11	Proton	-	Incident energy at the cell surface
Belli et al. 1998 [26]	V79	1.41	20	Proton	-	Incident energy at the cell surface
Belli et al. 1998 [26]	V79	0.76	30.5	Proton	-	Incident energy at the cell surface
Belli et al. 1998 [26]	V79	0.64	34.6	Proton	-	Incident energy at the cell surface
Belli et al. 1998 [26]	V79	0.57	37.8	Proton	-	Incident energy at the cell surface
Belli et al. 2000 [60]	SQ20B†	5.04	7.7	Proton	-	Incident energy at the cell surface
Belli et al. 2000 [60]	SQ20B†	1.49	19.8	Proton	-	Incident energy at the cell surface
Belli et al. 2000 [60]	SQ20B†	0.88	30	Proton	-	Incident energy at the cell surface
Belli et al. 2000 [60]	SCC25†	5.04	7.7	Proton	-	Incident energy at the cell surface
Belli et al. 2000 [60]	SCC25†	1.49	19.7	Proton	-	Incident energy at the cell surface
Belli et al. 2000 [60]	SCC25†	0.88	29.5	Proton	-	Incident energy at the cell surface
Belli et al. 2000 [60]	M10†	4.05	9.11	Proton	-	Incident energy at the cell surface
Belli et al. 2000 [60]	M10†	1.35	21.4	Proton	-	Incident energy at the cell surface
Belli et al. 2000 [60]	M10†	0.79	33	Proton	-	Incident energy at the cell surface

Study	Cell Line	Energy (MeV)	LET (keV/μm)	Particle	Depth (mm)	Comments
Belli et al. 2000 [60]	HF19†	5.04	7.7	Proton	-	Incident energy at the cell surface
Belli et al. 2000 [60]	HF19†	1.49	19.5	Proton	-	Incident energy at the cell surface
Belli et al. 2000 [60]	HF19†	0.88	30	Proton	-	Incident energy at the cell surface
Bettega et al. 2000 [176]	SCC25†	65	-	Proton	2	Depth in plexiglass
Bettega et al. 2000 [176]	SCC25†	65	-	Proton	15.6	Depth in plexiglass
Bettega et al. 2000 [176]	SCC25†	65	-	Proton	25	Depth in plexiglass
Bettega et al. 2000 [176]	SCC25†	65	-	Proton	27.2	Depth in plexiglass
Bettega et al. 2000 [176]	SCC25†	65	-	Proton	27.8	Depth in plexiglass
Bettega et al. 1998 [61]	C3H10T1/2	0.72	33.2	Proton	-	
Bettega et al. 1998 [61]	C3H10T1/2	0.74	32.5	Proton	-	
Bettega et al. 1998 [61]	C3H10T1/2	0.77	31.6	Proton	-	
Bettega et al. 1998 [61]	C3H10T1/2	0.87	19.7	Proton	-	
Bettega et al. 1998 [61]	C3H10T1/2	3.18	11	Proton	-	
Britten et al. 2013 [168]	V79	87	5.3	Proton	33.9	Depth in water/Pristine Bragg Peak
Britten et al. 2013 [168]	V79	87	20.5	Proton	58.6	Depth in water/Pristine Bragg Peak
Britten et al. 2013 [168]	V79	87	28.8	Proton	60.9	Depth in water/Pristine Bragg Peak
Calugaru et al. 2011 [177]	HeLa†	76	-	Proton	-	
Calugaru et al. 2011 [177]	HeLa†	201	-	Proton	-	
Calugaru et al. 2011 [177]	SQ20B†	76	-	Proton	-	
Folkard et al. 1989 [96]	V79	32	0.76	Proton	-	
Folkard et al. 1989 [96]	V79	24	1.15	Proton	-	
Folkard et al. 1989 [96]	V79	17	1.90	Proton	-	
Folkard et al. 1996 [126]	V79	10	3.66	Proton	-	
Folkard et al. 1996 [126]	V79	18	1.85	Proton	-	
Folkard et al. 1996 [126]	V79	28	1.07	Proton	-	

Study	Cell Line	Energy (MeV)	LET (keV/μm)	Particle	Depth (mm)	Comments
Guan et al. 2015 [27]	H1437†	79.7	10.8	Proton	-	
Guan et al. 2015 [27]	H1437 †	79.7	15.2	Proton	-	
Guan et al. 2015 [27]	H1437†	79.7	17.7	Proton	-	
Guan et al. 2015 [27]	H1437†	79.7	19	Proton	-	
Guan et al. 2015 [27]	H460†	79.7	10.8	Proton	-	Depth in water
Guan et al. 2015 [27]	H460†	79.7	15.2	Proton	-	Depth in water
Guan et al. 2015 [27]	H460†	79.7	17.7	Proton	-	Depth in water
Guan et al. 2015 [27]	H460†	79.7	19	Proton	-	Depth in water
Gueulette et al. 1996 [125]	CHO	85	-	Proton	10	Depth in water/Pristine Bragg Peak
Gueulette et al. 1996 [125]	CHO	85	-	Proton	47	Depth in water/Mid-5mm SOBP
Gueulette et al. 1996 [125]	CHO	85	-	Proton	25	Depth in water/Proximal 3cm SOBP
Gueulette et al. 1996 [125]	CHO	85	-	Proton	45	Depth in water/Distal 3cm SOBP
Howard et al. 2018 [123]	CHO	71	1.78	Proton	27	Approximate depths in water
Howard et al. 2018 [123]	CHO	71	3.36	Proton	37	Approximate depths in water
Howard et al. 2018 [123]	CHO	71	4.35	Proton	38	Approximate depths in water
Howard et al. 2018 [123]	CHO	71	7.34	Proton	40	Approximate depths in water
Howard et al. 2018 [123]	CHO	160	0.99	Proton	100	Approximate depths in water
Howard et al. 2018 [123]	CHO	160	2.26	Proton	168	Approximate depths in water
Howard et al. 2018 [123]	CHO	160	4.19	Proton	170	Approximate depths in water
Howard et al. 2018 [123]	CHO	160	7.29	Proton	172	Approximate depths in water
Howard et al. 2018 [123]	A549†	71	1.78	Proton	27	Approximate depths in water
Howard et al. 2018 [123]	A549†	71	3.36	Proton	37	Approximate depths in water
Howard et al. 2018 [123]	A549†	71	4.35	Proton	38	Approximate depths in water
Howard et al. 2018 [123]	A549†	71	7.29	Proton	40	Approximate depths in water
Howard et al. 2018 [123]	A549†	160	1.78	Proton	100	Approximate depths in water

Study	Cell Line	Energy (MeV)	LET (keV/μm)	Particle	Depth (mm)	Comments
Howard et al. 2018 [123]	A549†	160	2.26	Proton	168	Approximate depths in water
Howard et al. 2018 [123]	A549†	160	4.19	Proton	170	Approximate depths in water
Howard et al. 2018 [123]	A549†	160	7.29	Proton	172	Approximate depths in water
Matsumoto et al. 2018 [178]	HSG†	190	-	Proton	150	Depth in water/Fig. A/5cm SOBP
Matsumoto et al. 2018 [178]	HSG†	190	-	Proton	159	Depth in water/Fig. B/5cm SOBP
Matsumoto et al. 2018 [178]	HSG†	190	-	Proton	165	Depth in water/Fig. C/5cm SOBP
Matsumoto et al. 2018 [178]	HSG†	190	-	Proton	168	Depth in water/Fig. D/5cm SOBP
Matsumoto et al. 2018 [178]	HSG†	190	-	Proton	171	Depth in water/Fig. E/5cm SOBP
Matsumoto et al. 2018 [178]	HSG†	190	-	Proton	174	Depth in water/Fig. F/5cm SOBP
Matsumoto et al. 2018 [178]	HSG†	190	-	Proton	177	Depth in water Fig. G/5cm SOBP
Matsumoto et al. 2018 [178]	HSG†	190	-	Proton	180	Depth in water/Fig. H/5cm SOBP
Matsumoto et al. 2018 [178]	HSG†	190	-	Proton	183	Depth in water/Fig. I/5cm SOBP
Miller et al. 1995 [179]	C3H10T1/2	25.8	4	Proton	-	
Miller et al. 1995 [179]	C3H10T1/2	2.25	15	Proton	-	
Prise et al. 1990 [29]	V79	1.9	17	Proton	-	Hypoxic Cellular Conditions
Prise et al. 1990 [29]	V79	1.15	24	Proton	-	Hypoxic Cellular Conditions
Prise et al. 1990 [29]	V79	0.76	32	Proton	-	Hypoxic Cellular Conditions
Prise et al. 1990 [29]	V79	1.9	17	Proton	-	Oxic Cellular Conditions
Prise et al. 1990 [29]	V79	1.15	24	Proton	-	Oxic Cellular Conditions
Prise et al. 1990 [29]	V79	0.76	32	Proton	-	Oxic Cellular Conditions
Raju et al. 1979 [180]	V79	160	-	Proton	-	Hypoxic cells in peak region.
Raju et al. 1979 [180]	V79	160	-	Proton	-	Hypoxic cells in plateau region.
Raju et al. 1979 [180]	V79	160	-	Proton	-	Oxic cells in plateau region.
Robertson et al. 1994 [181]	V79	200	-	Proton	-	Depth in water not reported.
Robertson et al. 1994 [181]	V79	200	-	Proton	-	Depth in water not reported.
Robertson et al. 1994 [181]	V79	200	-	Proton	-	Depth in water not reported.

Study	Cell Line	Energy (MeV)	LET (keV/μm)	Particle	Depth (mm)	Comments
Robertson et al. 1994 [181]	V70	200	-	Proton	-	Depth in water not reported.
Robertson et al. 1994 [181]	V79	200	-	Proton	-	Depth in water not reported.
Tang et al. 1997 [130]	CHO	65	-	Proton	2	Depth in nylon/18mm SOBP
Tang et al. 1997 [130]	CHO	65	-	Proton	10	Depth in nylon/18mm SOBP
Tang et al. 1997 [130]	CHO	65	-	Proton	18	Depth in nylon/18mm SOBP
Tang et al. 1997 [130]	CHO	65	-	Proton	23	Depth in nylon/18mm SOBP
Wouters et al. 1996 [113]	V79	70	-	Proton	-	Depth in water/20mm SOBP/ average over proximal and distal
Wouters et al. 2015 [38]	V79	160	1.1	Proton	27.7	Depth in water
Wouters et al. 2015 [38]	V79	160	2.06	Proton	103.3	Depth in water
Wouters et al. 2015 [38]	V79	160	2.41	Proton	125.4	Depth in water
Wouters et al. 2015 [38]	V79	160	3.2	Proton	143.2	Depth in water
Wouters et al. 2015 [38]	V79	160	4.74	Proton	155.8	Depth in water
Wouters et al. 2015 [38]	V79	230	1.03	Proton	97.5	Depth in water
Wouters et al. 2015 [38]	V79	230	1.95	Proton	230.9	Depth in water
Wouters et al. 2015 [38]	V79	230	2.28	Proton	248.7	Depth in water
Wouters et al. 2015 [38]	V79	230	2.95	Proton	268.9	Depth in water
Wouters et al. 2015 [38]	V79	230	4.02	Proton	282.3	Depth in water

Appendix E

Regression Analysis Methods

This appendix contains an in-depth description of the methods employed in Chapter 5. We begin with a description of the assumptions considered in our study and the methods used to test them.

E.1 The Assumptions of Regression Analysis

The four assumptions are listed below are mandatory for any regression model. They are the minimum number of requirements that must be satisfied. If any of these assumptions are violated the forecasting and confidence intervals and the scientific findings by the model can be inefficient or seriously biased and misleading.

- **Linearity/Non-linearity** : The regression model is linear in parameters (α, β) of Eq.(5.12), suggesting that expected value of the response variable changes linearly with the independent variable. The slope of the resulting line must be independent of external variables. Finally, in the case of multiple effects from different independent variables, the resulting effects should be additive.

  In the case of Eq.(5.10), linearity is clearly not a requirement. Both of these properties are often easily assessed by plotting various results. If a linear model is fitted to non-linear data then the predictions can be seriously in error.

  This assumption will be tested using the following methods.

  (i) A plot of the fit residual errors[1] against the independent variable. If the residuals are not symmetrically distributed about a horizontal line, the assumption is violated.
  (ii) By observing the goodness-of-fit metrics[2].

[1] Residual errors are defined as the deviation between the observations $\mathbf{Y}$ and the model $f(\mathbf{X})$ for independent variable $\mathbf{X}$

[2] These are introduced in Section (E.1.1).

- **Homoscedasticity (Constant Variance of Residuals)** : The fit residuals must demonstrate *homoscedasticity*, meaning that they have a mean of zero and constant variance

$$\text{var}(\epsilon_\mathbf{i}|\mathbf{X_i}) = \sigma^2,$$

for i observations.

If this assumption is satisfied, each observation contributes equally to the total response modelled. If the assumption is violated, the data demonstrates *heteroscedasticity*,

$$\text{var}(\epsilon_\mathbf{i}|\mathbf{X_i}) = \sigma_i^2,$$

for i observations.

When the residual errors are not constant (homoscedasticity), the standard errors and estimated parameters are not reliable [114].

This assumption will be tested using the following methods.

(i) A plot of the residual errors against the fitted values or the independent variable. Similar to the former assumption, the plots should show symmetric scattering of the points about a horizontal line.
(ii) The Breusch-Pagan test for homoscedasticity [115].
(iii) The White test for homoscedasticity [116].

- **Autocorrelation** : Successive residual errors should be completely independent identically distributed (i.i.d.) of one another. Additionally, this assumption is analogous with homoscedasticity in that a plot of the residual errors against the independent variable should not show any correlation about the mean. If this assumption is violated, the ordinary least squares (OLS) procedure will no longer yield valid results [117]. In such a case, OLS is likely to underestimate the variance and the process for determining confidence intervals loses its validity.

This assumption will be tested using the following method.

(i) Durbin-Watson Statistical Test for residual correlation [117].

- **Normality** : The residual errors should be normally distributed by

$$\epsilon \sim \mathcal{N}(0, \sigma^2) \quad . \tag{E.1}$$

Non-normality can effect the determination of confidence intervals and the outcome of significance testing. Typically, the residual data can be used to generate a histogram to which a normal distribution can be fit. However, this method is not appropriate for the data considered here due to the small sample

sizes.

This assumption will be tested using the following methods.

(i) Kolmogorov-Smirnov test (KS test) [118, 119].
(ii) Anderson-Darling test (AD test) [121].

The following two assumptions are not directly applicable to the current investigation, however we feel that it is constructive to acknowledge.

- **No Multicollinearity** : This assumption will not be considered in this investigation, however we felt it would be important to discuss. Regression analysis requires that two or more variables in the model are not linearly related. That is the say that the variables in the regression model do not display multicollinearity.

- **Exogeneity** : This property implies that the response variable $\mathbf{Y}$ is determined as a direct result of the independent variable $\mathbf{X}$. If this assumption were violated, there would be a degree of causality between variables within the model disregarding $\mathbf{Y}$ being a consequence of $\mathbf{X}$.

E.1.1 Hypothesis Testing and Goodness-of-Fit Metrics

Many of the tests listed above are known as a *hypothesis test* which is a statistical technique employed to determine whether a null hypothesis (H_0) is true for a given system based on a set of n observations $\mathbf{x} = (x_1, x_2, ..., x_n)$. The test allows us to determine whether the null should be rejected (false) or not rejected (true) based on the observations $\mathbf{x}$ of the system. In the case that H_0 is rejected, we cannot reject the alternative hypothesis H_A.

Hypothesis testing always has a margin of error in that the null can be incorrectly rejected/accepted. Following the flowchart provided in Fig.(E.1), in the event that the null hypothesis correctly describes the system (i.e. H_0 is true) but the hypothesis test result rejects it given $\mathbf{x}$, a type I error is present. A type I error has a probability α of occurring where α is the significance level. It is common to use a significance level of 5% meaning that there is a 5% probability that the null will be incorrectly rejected and a 95% it will be correctly accepted. The same logic applies to the opposite case where the null is false but the hypothesis test does not reject it. This is known as a type II error and can occur with a probability of $1-\beta$ where β is the statistical power. When α and $1-\beta$ are small, the hypothesis testing result will be the most accurate.

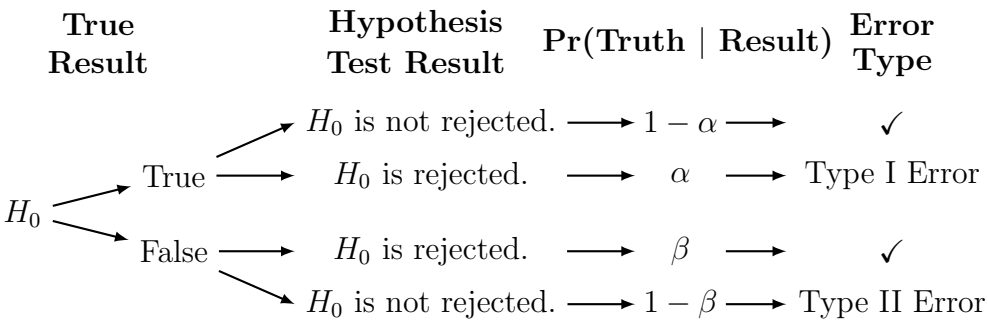


Figure E.1: An overview of hypothesis testing results.

The acceptance/rejection of the null hypothesis is determined quantitatively using a *test statistic*. The method used to compute the test statistic is unique to the hypothesis test but it is generally used in the same way to accept or reject the null. We determine the result of the hypothesis test by comparing it with a *critical value* that is known previously. Noting that the test statistic obeys a probability distribution (the exact distribution depends on the test), if the computed test statistic exists in the shaded region of Fig.(E.2) the null is accepted. The result of a hypothesis test is determined using the test statistic T as follows :

$$\begin{array}{lll} H_0 & \text{is not rejected if} & T < \text{critical value}, \\ \text{and } H_0 & \text{is rejected if} & T \geq \text{critical value}. \end{array}$$

E.1.1.1 More on the Critical Value and the p-Value

Using the correct critical value in hypothesis testing is clearly vital to giving a reliable result. The significance level α defines the sensitivity of the test. For example, a significance level of 95% means that the critical value is chosen so that 95% of the test statistics generated by the test - i.e. $\alpha = 1 - 0.95$. Figure (E.2) depicts an arbitrary distribution of test statistics with a critical value corresponding to a significance level α indicated as z_α. The area of the shaded region is given by Pr(test statistic $<$ critical value) = 1 - α and the remainder being α.

The p-value is the probability of the test statistic being in the rejection region of area α. Considering this, the result of the hypothesis test can also be given by the following.

$$\begin{array}{lll} H_0 & \text{is rejected if} & p < \alpha, \\ \text{and } H_0 & \text{is not rejected if} & p \geq \alpha. \end{array}$$

Upon observation of Fig.(E.2), the hypothesis test has a higher pass rate for smaller α due to the corresponding critical value being larger. At this point it is worth noting that the rejection or non-rejection of a null hypothesis does not infer as such with 100% certainty. We can only say that the conclusion is made with a

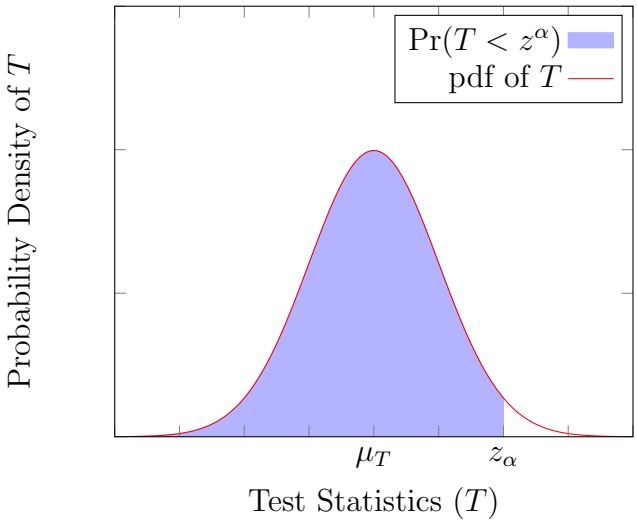


Figure E.2: Critical value z_α of a hypothesis test indicated on an arbitrary distribution of test statistics T for significance level α.

degree of certainty analogous to the significance level.

Additionally, the critical value can differ according to the size and nature of the data. Generally, the critical value is given as a function of sample size n. In fact, some hypothesis tests can only be applied over a specified range of sample sizes.

A common application of a hypothesis test is to determine whether a set of observations follow a given parameterised probability distribution. In this instance, the critical values differ depending on how the distribution parameters are obtained. This means that the critical values are obtained according to two categories.

1. **Out-of-sample :** The parameter space of the probability distribution that the data is being tested against, θ, are known in advance.

2. **In-sample :** The parameter space of the probability distribution that the data is being tested against, θ, is estimated from the data.

In the current context, our investigation falls under the latter category - *in-sample*. For example when testing the assumption of normality, we will be estimating the mean and variance of the normal distribution by minimising the difference between the expected and measured residual errors. This typically results in smaller test statistics that are not widely distributed[3]. When the parameters of the distribution are estimated the critical value will depend on the number of parameters to estimate, sample size and significance level.

The hypothesis tests employed in the study from Chapter 5 are now described in detail below.

[3]In this case, Figure (E.2) would be narrower than when the parameters are known *a priori*.

E.1.1.2 Breusch-Pagan Test for Homoscedasticity

The Breusch-Pagan (BP) test was developed by Breusch and Pagan 1979 [115] and is based on the Lagrange Multiplier Test. Consider

$$\mathbf{Y_i} = \mathbf{f}(\mathbf{X_i}|\beta) + \epsilon_\mathbf{i}, \tag{E.2}$$

where β is a $(k \times 1)$ vector of coefficient parameters and errors ϵ_i are normally and independently distributed with mean zero, $\mathbf{E}(\epsilon_i) = 0$ and variance, $\text{var}(\epsilon_i) = \sigma_i^2 = h(z'\alpha)$.

The null and alternative hypotheses can be written as :

H_0: $\alpha_2 = ... = \alpha_p = 0$ for z'$\alpha = \alpha_1$ so that $\sigma_i^2 = h(\alpha_1) = \sigma^2$ in constant (the residual errors of regression fit $\mathbf{Y_i} = \mathbf{f}(\mathbf{X_i}|\hat{\beta}) + \epsilon$ demonstrate homoscedasticity).
H_A: The residual errors of regression fit $\mathbf{Y_i} = \mathbf{f}(\mathbf{X_i}|\hat{\beta}) + \epsilon$ demonstrate heteroscedasticity.

The test statistic is given by the Lagrange Multiplier :

$$LM = nR^2 \sim \chi^2_{k-1},$$

where n is the sample size, R^2 is the coefficient of determination[4] and k is the number of degrees of freedom, $\chi^2(LM, k)$.

The test is performed in the following steps.

1. Perform a least squares fitting with the regression model $\mathbf{Y_i} = \mathbf{f}(\mathbf{X_i}|\hat{\beta})$ to the data and determine the residual errors.

2. Based on the assumption that $\hat{\epsilon} = 0$, the variance is computed by taking the mean square of the residuals, $\widehat{\epsilon^2}$.

3. Perform a regression of the residuals-squared against the independent variable from which we determine a coefficient of determination R^2.

4. Compute the test statistic, $LM = nR^2$ and the p-value using $1-\chi^2(LM, k)$ distribution where k is the number of degrees of freedom. If the p-value is greater than the α (we are testing at the 95% significance level, therefore $\alpha = 0.05$, i.e. p-value> 0.05) we cannot reject the null hypothesis of homoscedasticity.

The objective of the BP test is to determine if the regression coefficients of this fit are zero - i.e. the residuals are homoscedastic. This test is very common when quantifying homoscedasticity and was chosen for use in our investigation for this reason.

[4] This is defined in the next section.

E.1.1.3 White Test for Homoscedasticity

The White Test was developed in 1981 [116] and is similar to the Breusch-Pagan Test in nature in that it test for constant variance of errors with the difference that non-linear correlations in the residual errors. It tests for the variance of errors in a regression model fit of the residuals to the independent variable. It also accounts for relationships between the residual errors and the square or cross product of the independent variables. This is a situation where the Breusch-Pagan Test fails to give reliable results. Alike to the BP test, the White test assumes that the errors are i.i.d.

The null and alternative hypothesis, test statistic, and p-value are computed identically to the Breusch-Pagan Test. The fundamental difference is when the second regression is performed, the equation to which the residuals are fit is given by

$$\hat{\epsilon} = \gamma_0 + \gamma_1 x + ... + \gamma_n x_n + \gamma_{n+1} x_1^2 + \gamma_1 \gamma_{k+2} x_1 x_2 + e.$$

This test was chosen in the current investigation to account for residual correlations that are non-linear, unlike the BP test which only accounts for linear relationships between residuals and the independent variable.

E.1.1.4 Durbin-Watson Statistics for Autocorrelation

The Durbin-Watson (DW) test for autocorrelation was developed by Durbin and Watson in 1950 [117]. We start by defining the first-order autocorrelation of n residual errors, ϵ, as follows,

$$\epsilon_i = \rho \epsilon_{i-1} + u_i$$

where $i = \{1, 2, ..., n\}$, $|\rho| < 1$ and $u_i \sim \mathcal{N}(0, \sigma^2)$. ρ represents the degree of correlation between the current and proceeding residual error, therefore our null and alternative hypotheses are as follows.

$$H_0\text{: } \rho = 0 \text{ (ϵ are not autocorrelated.)}$$
$$H_A\text{: } \rho > 0 \text{ or } \rho < 0 \text{ (ϵ are autocorrelated.)}$$

The test statistic is computed using

$$D = \frac{\sum_{i=2}^{n} (\epsilon_i - \epsilon_{i-1})^2}{\sum_{i=1}^{n} \epsilon_i^2},$$

where $D \in [0, 4]$. The result is interpreted as follows,

- $0 \leq D < 2$ indicates *positive* autocorrelation,
- $D = 2$ indicates *zero* autocorrelation,
- and $2 \leq D \leq 4$ indicates *negative* autocorrelation.

Note: This test only considers a lag value of one, meaning that a given residual error is only tested for correlation against the preceding and following residuals. For a given sample size n, number of model parameters and significance level, the critical values are presented as a range with an upper and lower bound, denoted as d_U and d_L respectively. H_0 can then be rejected or not rejected according to the following criteria.

To test for positive autocorrelation

- if $D < d_{L,\alpha}$ - positive autocorrelation,
- if $D > d_{U,\alpha}$ - no statistical evidence of autocorrelation,
- if $d_{L,\alpha} < D < d_{U,\alpha}$
- if $(4 - D) < d_{L,\alpha}$ - negative autocorrelation,
- if $(4 - D) > d_{U,\alpha}$ - no statistical evidence of autocorrelation,
- $d_{L,\alpha} < (4 - D) < d_{U,\alpha}$ - inconclusive.

E.1.1.5 Kolmogorov-Smirnov Test

The Kolmogorov-Smirnov (KS) Test was formulated from seperate studies by Kolmogorov [118] and Smirnov [119] in 1933 and 1945, respectively. This test is the most common of all edf tests. The test can be used to test a sample against any theoretical distribution. In the current context, we will apply the KS test to the residual errors to determine if they are normally distributed. Our null and alternative hypotheses are

H_0: The residual errors are normally distributed,
H_A: The residual errors are not normally distributed.

To test if a sample of n observations $\mathbf{x} = (x_1, x_2, ..., x_n)$ follows a chosen distribution. This is achieved by comparing the empirical distribution function (edf) of the sample against the cumulative distribution function (cdf) that the sample is being tested against. We define the edf of x as

$$\hat{F}_n(x) = \frac{\text{Number of observations in sample } \leq x}{n} = \frac{1}{n}\sum_{i=1}^{n} I(x_i \leq x), \qquad \text{(E.3)}$$

$$\text{where } I = \begin{cases} 0 \text{ for } x_i > x \\ 1 \text{ for } x_i \leq x, \end{cases}$$

and the cdf is denoted as $F(x|\theta)$ for all the parameters in the parameter space θ. For example $F(x|\theta)$ = Normal Distribution (for our purpose). The test is performed

by taking the maximum difference between the edf and theoretical cdf to determine D^+ and D^- respectively (Figure E.3). The D^+ and D^- values are expressed as

$$\begin{aligned} D^+ &= \max_{i=1,\dots,n} \left[\frac{i}{n} - F(x|\theta) \right], \\ D^- &= \max_{i=1,\dots,n} \left[F(x|\theta) - \frac{i-1}{n} \right]. \end{aligned} \tag{E.4}$$

This test requires prior knowledge of the theoretical cdf parameters θ, therefore in the case that these are unknown, the distribution should be fit before performing the test.

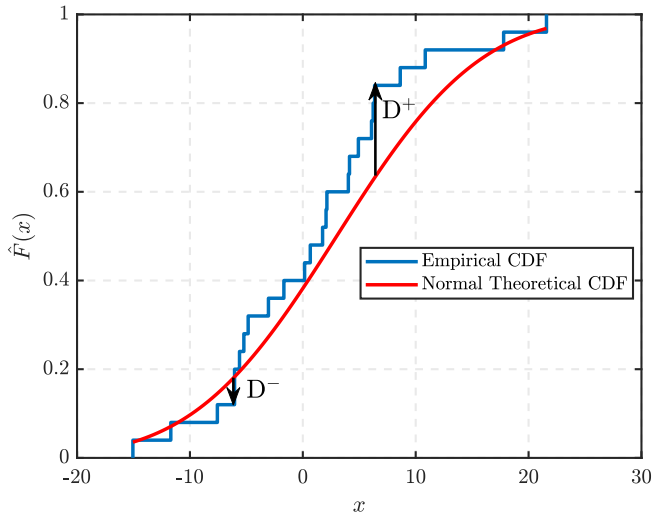


Figure E.3: The edf of 25 observations with the normal theoretical cdf - $\mathcal{N}(3, 10)$. The maximum differences D^+ and D^- are indicated.

The test statistic D is determined from the maximum difference between the edf and cdf

$$D = \max(D^+, D^-) \quad . \tag{E.5}$$

The critical values of the KS Test are dependent on whether the parameters of our theoretical cdf are estimated from the sample or previously known. In the current context, we will be applying a method to estimate the parameters of the theoretical cdf for the normal distribution, μ and σ. Therefore, we must ensure that the *in-sample* critical values are used. Such critical values for the normal distribution are available in the literature [120] for a given sample size and significance level.

The null hypothesis is rejected or not rejected based on the following criteria,

$$\begin{aligned} H_0 \text{ is rejected if } D > \mathrm{CV}_{n,\alpha}, \\ H_0 \text{ is not rejected if } D < \mathrm{CV}_{n,\alpha} \end{aligned}$$

where $\text{CV}_{n,alpha}$ is the critical value for a given sample size n and significance level α.

E.1.1.6 Anderson-Darling Test

The Anderson-Darling Test was developed in 1952 by T.W. Anderson and D.A. Darling [121]. It is another common edf test used to determine is a sample is derived from a given probability distribution. In the current context, we compare our residual errors against a normal distribution. Therefore our null and alternative hypotheses are:

$$H_0\text{: The residual errors are normally distributed,}$$
$$H_A\text{: The residual errors are not normally distributed.}$$

The test statistic A^2 is computed using,

$$A^2 = n \int_{-\infty}^{\infty} \frac{[F_n(x) - F(x)]^2}{F(x)[1 - F(x)]} dF(x),$$

which can be simplified to,

$$A^2 = -n - \frac{1}{n} \sum_{i=1}^{n} \{2i - 1\}\{\log[F(x_i)] + \log[1 - F(x_{n-1+1})]\}. \tag{E.6}$$

The critical values are unique for sample size n and significance level α:

$$CV_{n,\alpha} = \frac{0.752}{1 + \frac{0.75}{n} + \frac{2.25}{n^2}}. \tag{E.7}$$

The null hypothesis is rejected or not rejected based on the following criteria,

$$H_0 \text{ is rejected if } A^2 > \text{CV}_{n,\alpha},$$
$$H_0 \text{ is not rejected if } A^2 < \text{CV}_{n,\alpha}$$

where $\text{CV}_{n,alpha}$ is the critical value for a given sample size n and significance level α.

E.1.2 Goodness-of-fit Metrics

In addition to testing our model fits against the assumptions of regression analysis, it is important to evaluate the model's effectiveness by comparing the goodness-of-fit metrics [114]. Below we introduce each of the metrics to be used in this investigation with a discussion of how to interpret the results in the context of the LQ model performance.

First, we consider n observations $\mathbf{X} = \{x_1, x_2, ..., x_n\}$ with corresponding responses $\mathbf{Y} = \{y_1, y_2, ..., y_n\}$ on which a model, $\mathbf{Y} = f(\mathbf{X})$, will be evaluated for goodness-of-fit.

E.1.2.1 Sum of Squared Errors

The sum of squared errors (SSE) is computed by squaring the difference between the observations y_i and predicted values $\hat{y}_i$ and is written as

$$SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2. \tag{E.8}$$

The result is a numeric interpretation of how well the model fits the sample or alternatively the variance in the difference between the observed and predicted values. Trivially, a smaller SSE implies smaller residual errors, thus resulting in a better fit for forecasting. Although a more profound method of interpreting the SSE is to take the sample size into consideration. For larger sample sizes, our SSE will naturally increase, thus giving the illusion of the model fit being worse when this is not necessarily the case. This must be taken into account, specifically when comparing SSE for different sample sizes.

E.1.2.2 Root Mean Square Error

The root mean square error (RMSE) is the square root of the averaged squared error in the model fit.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i)^2} \tag{E.9}$$

Heuristically, this metric gives us insight into the mean distance between the predicted ($\hat{y}_i$) and observed values (y_i). Unlike the SSE metric, the RMSE is scales the difference by $\sqrt{1/n}$. This allows direct comparison of this metric between the model fits to data of varying sample size. The rule of thumb when interpreting this metric is that a RMSE close to zero implies that our model is an effective predictor of our observations. Additionally, a large RMSE implies that there may be additional factors contributing to the data that we have not accounted for in the model.

E.1.2.3 Coefficient of Determination

The coefficient of determination or the R-squared value is defined as the proportion of the SSE that is derived from the independent variable [114]. The metric ranges from 0 to 1 and is defined as

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} y_i^2} = 1 - \frac{SSE}{\sum_{i=1}^{n} y_i^2}. \tag{E.10}$$

If the R^2 coefficient is close to 1, we can conclude that our model accurately described the observations on which it is being tested. It is possible that R^2 can increase as a result of adding more model parameters without adding value to the

model thus resulting in '"over-fitting"[5] of the model to the data.

For example, if an R^2 metric of 0.90 is achieved for an arbitrary dose response curve. Then it can be concluded that 90% of the variance in the survival fraction is the result of its relationship with the absorbed dose.

E.1.2.4 Adjusted Coefficient of Determination

Unlike the regular coefficient of determination, the adjusted coefficient of determination (R^2_{Adj}) takes into account the number of parameters in the model. It is possible to "overfit" a model by using unnecessary parameters. The adjusted R-squared value allows models with different parameter spaces to be compared directly. Of course a model with five parameters would be a better fit to a data set than a one parameter model. The adjusted R-squared values allows these two models to be compared directly. It is evaluated using

$$R^2_{Adj} = 1 - \frac{(1 - R^2)(n - 1)}{n - p} \tag{E.11}$$

for p model parameter, sample size n and the regular coefficient of determination R^2.

The result is often smaller than the R^2 and will only increase with the addition of a parameter that improves the overall performance of the model by a significant amount. Additionally, the adjusted R^2 value will decrease following the introduction of a irrelevant model parameter.

E.1.2.5 Pearson's χ^2 Test

The χ^2 goodness-of-fit test was developed by Pearson in 1900 [182] and is among the most common used in model testing and fitting. The current study is no exception. The statistic is calculated using

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} \tag{E.12}$$

where O_i is the measured value, E_i is the model predicted value and n is the sample size. A χ^2 statistic close to zero indicates that the model is a good fit to the data.

When weighted by errors we determine the χ^2 statistic using the expected (y_i) and predicted ($\hat{y}_i$)

$$\chi^2 = \frac{1}{N - d} \sum_{i=1}^{n} \frac{(y_i - \hat{y}_i^2)}{\hat{y}_i^2} \tag{E.13}$$

[5]Models that contain unnecessary parameters decrease the degrees of freedom when it is not warranted and should be penalised statistically.

where σ_i are the measurement errors, N is the sample size and d are the number of parameters in the fit equation. This is the χ^2 statistic used in the current study for a given error (standard error of the mean, standard deviation, etc.).

The result can also be scaled by the number of degrees of freedom as we have done for this study. This allows direct comparison of fits with different sample size n which is useful in the current context because n ranges between 5 and 24 in the experimental data.

Appendix F

Regression Analysis Results: Linear-Quadratic Model Fits to Experimental Data

This appendix consists of the regression analysis results performed on fits using the linear and non-linear form of the LQ model for the experimental data in Appendix D.

Table F.1: The fit parameters and goodness-of-fit metrics for linearised LQ model fits to selected experimental data . Cell lines of human origin are denoted by a dagger (†).

Study	LET (keV/μm)	Energy (MeV)	Cell Line	$\alpha_{Orig.}$ (±SE)	$\beta_{Orig.}$ (±SE)	α (±SE)	β (±SE)	$\chi^2_{\mathbf{Orig.}}$	χ^2	RMSE	R^2
Ando et al. [174]	-	235	SCC61†	0.470±0.114	0.085±0.019	0.418±0.062	0.092±0.012	17.583	15.481	0.977	0.776
Ando et al. [174]	-	235	SCC61†	0.245±0.079	0.097±0.013	0.273±0.04	0.092±0.008	6.915	6.636	3.935	0.824
Ando et al. [174]	-	235	NB1RGB†	0.390±0.050	0.041±0.008	0.528±0.035	0.017±0.008	3.257	1.284	2.576	0.915
Ando et al. [174]	-	235	NB1RGB†	0.401±0.071	0.040±0.012	0.586±0.049	-0.003±0.013	7.233	3.184	1.133	0.255
Ando et al. [174]	-	235	NB1RGB†	0.390±0.050	0.041±0.008	0.528±0.035	0.017±0.008	3.257	1.284	1.784	0.003
Ando et al. [174]	-	235	V79	0.166±0.022	0.024±0.002	0.151±0.010	0.025±0.001	1.796	0.632	1.133	0.255
Ando et al. [174]	-	235	V79	0.214±0.021	0.021±0.002	0.228±0.011	0.020±0.001	0.836	0.604	0.795	0.978
Antoccia et al. [175]	28.5	0.8	HFFF2†	0.910±0.030	0.000±0.000	0.922±0.038	-0.038±0.011	12.214	3.034	0.777	0.957
Baek et al. [124]	-	190	HSG†	0.291±0.000	0.041±0.000	0.349±0.017	0.030±0.004	0.221	0.070	1.742	0.267
Belli et al. [60]	7.7	5.04	SQ20B†	0.150±0.050	0.011±0.012	0.154±0.010	0.011±0.002	0.179	0.167	0.409	0.642
Belli et al. [60]	19.8	1.49	SQ20B†	0.230±0.040	0.004±0.007	0.226±0.018	0.004±0.003	0.771	0.755	0.869	0.128
Belli et al. [60]	7.7	5.01	SCC25†	0.410±0.011	0.092±0.036	0.434±0.051	0.084±0.016	0.789	0.748	0.865	0.834
Belli et al. [60]	19.7	1.49	SCC25†	0.870±0.050	0.000±0.000	0.897±0.026	-0.009±0.009	0.139	0.116	0.340	0.161
Belli et al. [60]	29.5	0.88	SCC25†	0.810±0.030	0.000±0.000	0.900±0.048	-0.026±0.012	0.646	0.361	0.601	0.429
Belli et al. [60]	9.11	4.05	M10†	0.500±0.030	0.000±0.000	0.530±0.022	-0.005±0.004	0.451	0.391	0.625	0.114
Belli et al. [60]	21.4	1.35	M10†	0.490±0.020	0.000±0.000	0.467±0.038	0.003±0.007	1.517	1.482	1.218	0.011
Belli et al. [60]	33	0.79	M10†	0.930±0.050	0.000±0.000	1.047±0.050	-0.032±0.013	0.774	0.521	0.722	0.325
Belli et al. [60]	7.7	5.04	HF19†	0.550±0.010	0.000±0.000	0.514±0.048	0.001±0.012	7.627	4.267	2.066	0.001
Belli et al. [60]	19.5	1.49	HF19†	0.540±0.020	0.000±0.000	0.538±0.043	-0.002±0.011	1.866	1.776	1.333	0.002
Belli et al. [60]	29	0.88	HF19†	0.520±0.010	0.000±0.000	0.662±0.047	-0.03±0.009	5.627	2.464	1.570	0.476
Belli et al. [26]	7.7	5.01	V79	0.289±0.023	0.024±0.006	0.302±0.012	0.022±0.003	0.507	0.459	0.677	0.716
Belli et al. [26]	11	3.2	V79	0.372±0.032	0.036±0.009	0.376±0.016	0.033±0.004	0.988	0.926	0.962	0.77
Belli et al. [26]	20	1.41	V79	0.469±0.029	0.043±0.009	0.498±0.017	0.035±0.005	0.498	0.399	0.631	0.818
Belli et al. [26]	30.5	0.76	V79	0.721±0.017	0.000±0.000	0.720±0.027	0.001±0.008	1.807	1.797	1.341	0.001
Belli et al. [26]	34.6	0.65	V79	0.653±0.025	0.000±0.000	0.779±0.035	-0.032±0.008	4.622	1.999	1.414	0.565
Belli et al. [26]	37.8	0.57	V79	0.589±0.019	0.000±0.000	0.690±0.060	-0.023±0.003	3.624	2.090	1.446	0.529
Bettega et al. [61]	11	3.18	C3H10T12	0.470±0.060	0.019±0.014	0.469±0.038	0.023±0.010	1.609	1.429	1.196	0.312
Bettega et al. [61]	19.7	1.46	C3H10T12	0.430±0.060	0.038±0.013	0.424±0.027	0.039±0.006	0.759	0.747	0.864	0.769
Bettega et al. [61]	31.6	0.77	C3H10T12	0.670±0.040	0.000±0.000	0.729±0.056	-0.024±0.017	0.946	0.639	0.799	0.263
Bettega et al. [61]	32.5	0.74	C3H10T12	0.750±0.040	0.000±0.000	0.835±0.054	-0.027±0.017	0.525	0.364	0.604	0.302
Bettega et al. [61]	33.2	0.720	C3H10T12	1.020±0.060	0.000±0.000	1.002±0.066	0.001±0.029	0.480	0.448	0.669	<0.001
Bettega et al. [176]	-	65	SCC25†	0.700±0.050	0.018±0.012	0.748±0.032	0.009±0.007	1.917	1.609	1.714	0.206
Bettega et al. [176]	-	65	SCC25†	0.610±0.050	0.010±0.008	0.610±0.034	0.011±0.005	1.620	1.497	0.751	0.961
Bettega et al. [176]	-	65	SCC25†	0.830±0.080	0.001±0.014	0.832±0.040	0.001±0.009	1.767	1.766	1.269	0.124
Calugaru et al. [177]	-	76	HeLa†	0.293±0.026	0.020±0.012	0.287±0.015	0.022±0.008	1.405	1.35	1.224	0.279

Study	**LET (keV/μm)**	**Energy (MeV)**	**Cell Line**	$\alpha_{Orig.}$ **(±SE)**	$\beta_{Orig.}$ **(±SE)**	α **(±SE)**	β **(±SE)**	$\chi^2_{\mathbf{Orig.}}$	χ^2	**RMSE**	R^2
Calugaru et al. [177]	-	201	HeLa†	0.213±0.016	0.054±0.007	0.177±0.002	0.055±0.001	4.479	0.085	1.329	0.001
Calugaru et al. [177]	-	76	SQ20B†	0.029±0.013	0.012±0.002	0.025±0.003	0.013±0.000	2.451	1.767	1.162	0.410
Folkard et al. [96]	17	1.9	V79	0.130±0.036	0.078±0.007	0.183±0.031	0.07±0.006	3.106	2.821	1.68	0.818
Folkard et al. [96]	24	1.15	V79	0.330±0.038	0.066±0.005	0.335±0.012	0.067±0.002	0.672	0.586	0.765	0.969
Folkard et al. [96]	32	0.76	V79	1.030±0.016	0.000±0.000	1.069±0.029	-0.006±0.004	3.349	2.621	1.619	0.083
Folkard et al. [126]	10	3.66	V79	0.320±0.058	0.039±0.011	0.235±0.022	0.055±0.004	1.024	0.463	0.291	0.99
Folkard et al. [126]	18	1.83	V79	0.450±0.035	0.028±0.006	0.462±0.024	0.027±0.004	2.597	2.522	1.329	0.96
Folkard et al. [126]	28	1.07	V79	0.740±0.025	0.011±0.004	0.741±0.012	0.014±0.003	1.391	0.823	0.681	0.928
Goodhead et al. [169]	23	1.16	V79	0.300±0.060	0.052±0.008	0.232±0.012	0.059±0.002	15.94	9.997	1.588	0.544
Goodhead et al. [169]	20	1.38	V79	0.420±0.050	0.019±0.008	0.408±0.007	0.019±0.001	10.653	10.162	0.907	0.591
Goodhead et al. [169]	23	1.16	HeLa†	0.530±0.12	0.084±0.025	0.485±0.034	0.087±0.007	12.827	11.506	3.162	0.803
Goodhead et al. [169]	20	1.83	HeLa†	0.850±0.16	0.037±0.028	0.923±0.048	0.012±0.010	19.090	16.868	3.188	0.307
Guan et al. [27]	10.8	79.7	H460†	0.318±0.000	0.154±0.000	0.341±0.038	0.143±0.013	4.918	4.399	3.392	0.679
Guan et al. [27]	15.2	79.7	H460†	0.446±0.000	0.341±0.000	0.461±0.076	0.333±0.044	4.221	4.122	4.107	0.016
Guan et al. [27]	17.7	79.7	H460†	0.596±0.000	0.662±0.000	0.591±0.057	0.655±0.045	3.306	3.078	2.097	0.951
Guan et al. [27]	19	79.7	H460†	0.883±0.000	0.956±0.000	0.939±0.111	0.894±0.105	8.501	8.229	2.030	0.971
Gueulette et al. [125]	-	85	CHO	0.071±0.044	0.085±0.007	0.023±0.011	0.083±0.002	5.315	1.847	1.755	0.974
Howard et al. [123]	1.78	71	CHO	0.220±0.040	0.030±0.010	0.224±0.008	0.033±0.001	1.604	0.293	1.324	0.808
Howard et al. [123]	3.36	71	CHO	0.240±0.060	0.050±0.010	0.236±0.014	0.053±0.003	0.276	0.21	0.541	0.982
Howard et al. [123]	4.35	71	CHO	0.340±0.100	0.050±0.020	0.228±0.036	0.078±0.009	1.025	0.359	0.458	0.969
Howard et al. [123]	7.34	71	CHO	0.470±0.110	0.040±0.030	0.411±0.048	0.054±0.012	1.211	0.954	0.599	0.936
Howard et al. [123]	0.99	160	CHO	0.220±0.030	0.030±0.010	0.216±0.011	0.034±0.002	2.069	0.567	1.755	0.974
Howard et al. [123]	2.26	160	CHO	0.230±0.040	0.040±0.010	0.203±0.022	0.042±0.005	1.207	0.779	2.869	0.859
Howard et al. [123]	4.19	160	CHO	0.290±0.070	0.050±0.010	0.261±0.020	0.052±0.004	0.559	0.340	1.359	0.928
Howard et al. [123]	7.29	160	CHO	0.310±0.100	0.060±0.020	0.287±0.047	0.069±0.011	0.731	0.598	0.753	0.961
Howard et al. [123]	1.78	71	A549†	0.280±0.050	0.030±0.010	0.264±0.010	0.033±0.001	0.14	0.095	0.753	0.961
Howard et al. [123]	3.36	71	A549†	0.320±0.070	0.030±0.010	0.234±0.029	0.052±0.006	0.951	0.295	0.883	0.875
Howard et al. [123]	7.29	71	A549†	0.590±0.080	0.010±0.010	0.480±0.056	0.029±0.012	1.201	0.689	0.883	0.875
Howard et al. [123]	0.99	160	A549†	0.250±0.040	0.030±0.010	0.247±0.008	0.027±0.001	2.052	0.085	0.583	0.937
Howard et al. [123]	2.26	160	A549†	0.230±0.040	0.020±0.010	0.232±0.020	0.023±0.003	4.124	0.178	0.773	0.879
Howard et al. [123]	4.19	160	A549†	0.190±0.050	0.040±0.010	0.155±0.018	0.042±0.003	1.161	0.544	0.308	0.976
Howard et al. [123]	7.29	160	A549†	0.260±0.070	0.040±0.010	0.265±0.007	0.042±0.002	0.162	0.019	0.543	0.920
Mastumoto et al. [178]	-	190	HSG†	0.220±0.070	0.050±0.010	0.193±0.009	0.054±0.002	0.347	0.231	0.543	0.920
Mastumoto et al. [178]	-	190	HSG†	0.280±0.070	0.050±0.010	0.267±0.012	0.050±0.002	1.456	0.437	0.83	0.527
Mastumoto et al. [178]	-	190	HSG†	0.250±0.070	0.050±0.010	0.247±0.007	0.051±0.001	0.517	0.395	0.583	0.937
Mastumoto et al. [178]	-	190	HSG†	0.260±0.030	0.050±0.010	0.236±0.008	0.057±0.001	3.062	0.436	0.773	0.879
Mastumoto et al. [178]	-	190	HSG†	0.420±0.010	0.050±0.010	0.419±0.011	0.046±0.003	3.103	1.040	0.291	0.986
Mastumoto et al. [178]	-	190	HSG†	0.410±0.040	0.050±0.010	0.403±0.010	0.052±0.002	0.953	0.910	0.422	0.977

Study	**LET (keV/μm)**	**Energy (MeV)**	**Cell Line**	$\alpha_{Orig.}$ **(±SE)**	$\beta_{Orig.}$ **(±SE)**	α **(±SE)**	β **(±SE)**	$\chi^2_{\mathbf{Orig.}}$	χ^2	**RMSE**	R^2
Mastumoto et al. [178]	-	190	HSG†	0.380±0.030	0.050±0.010	0.380±0.014	0.053±0.003	1.825	1.211	0.738	0.947
Mastumoto et al. [178]	-	190	HSG†	0.440±0.060	0.030±0.010	0.421±0.014	0.035±0.003	0.473	0.386	0.137	0.993
Mastumoto et al. [178]	-	190	HSG†	0.420±0.070	0.030±0.010	0.428±0.006	0.028±0.001	0.332	0.232	0.481	0.985
Tang et al. [130]	-	65	CHO	0.200±0.035	0.022±0.007	0.215±0.011	0.020±0.001	0.538	0.462	0.661	0.956
Tang et al. [130]	-	65	CHO	0.210±0.044	0.022±0.007	0.219±0.011	0.023±0.001	1.106	0.334	0.628	0.987
Tang et al. [130]	-	65	CHO	0.220±0.036	0.026±0.007	0.215±0.008	0.026±0.001	0.150	0.093	0.660	0.987
Tang et al. [130]	-	65	CHO	0.249±0.045	0.025±0.008	0.259±0.008	0.020±0.001	11.967	0.658	1.020	0.932
Wouters et al. [38]	1.1	160	V79	0.114±0.027	0.048±0.007	0.122±0.000	0.047±0.000	0.014	<0.001	0.954	0.961
Wouters et al. [38]	2.06	160	V79	0.104±0.017	0.055±0.003	0.112±0.001	0.054±0.000	0.01	0.001	1.101	0.955
Wouters et al. [38]	2.41	160	V79	0.113±0.014	0.054±0.003	0.124±0.001	0.053±0.000	0.026	<0.001	0.622	0.897
Wouters et al. [38]	3.2	160	V79	0.133±0.014	0.055±0.003	0.141±0.000	0.054±0.000	0.007	<0.001	0.482	0.972
Wouters et al. [38]	4.74	160	V79	0.141±0.007	0.059±0.001	0.153±0.000	0.057±0.000	0.014	0.002	0.68	0.954
Wouters et al. [38]	1.03	160	V79	0.103±0.016	0.051±0.003	0.103±0.000	0.05±0	0.038	<0.001	0.578	0.954
Wouters et al. [38]	1.95	230	V79	0.103±0.008	0.054±0.002	0.101±0.001	0.054±0.000	0.002	<0.001	0.306	0.985
Wouters et al. [38]	2.28	230	V79	0.102±0.008	0.057±0.002	0.103±0.001	0.057±0.000	0.001	<0.001	0.811	0.951
Wouters et al. [38]	2.95	230	V79	0.123±0.011	0.057±0.002	0.122±0.000	0.057±0.000	0.004	<0.001	0.016	0.999
Wouters et al. [38]	4.02	230	V79	0.121±0.013	0.063±0.002	0.122±0.000	0.062±0.000	0.010	<0.001	0.035	0.999

Table F.4: The fit parameters and goodness-of-fit metrics for non-linear LQ model fits to selected experimental data .

Study	**LET** (keV/μm)	**Energy** (MeV)	**Cell Line**	$\alpha_{Orig.}$ (±**SE**)	$\beta_{Orig.}$ (±**SE**)	α (±**SE**)	β (±**SE**)	$\chi^2_{\mathbf{Orig.}}$	χ^2	**RMSE**	R^2	**RBE** 10%
Ando et al. [174]	-	235	SCC61†	0.470 ± 0.114	0.085 ± 0.019	0.481 ± 0.069	0.082 ± 0.014	14.931	14.860	3.855	0.941	1.102
Ando et al. [174]	-	235	SCC61†	0.245 ± 0.079	0.097 ± 0.013	0.298 ± 0.042	0.088 ± 0.008	7.276	6.343	2.518	0.977	0.931
Ando et al. [174]	-	235	NB1RGB†	0.401 ± 0.071	0.040 ± 0.012	0.578 ± 0.045	0.003 ± 0.011	8.699	3.249	1.803	0.97	1.103
Ando et al. [174]	-	235	NB1RGB†	0.390 ± 0.050	0.041 ± 0.008	0.544 ± 0.036	0.015 ± 0.008	3.665	1.195	1.093	0.988	1.121
Ando et al. [174]	-	235	V79	0.166 ± 0.022	0.024 ± 0.002	0.153 ± 0.009	0.024 ± 0.001	1.461	0.519	0.721	0.996	0.999
Ando et al. [174]	-	235	V79	0.214 ± 0.021	0.021 ± 0.002	0.232 ± 0.011	0.02 ± 0.001	0.796	0.503	0.709	0.994	1.099
Antoccia et al. [175]	28.5	0.8	HFFF2†	0.910 ± 0.030	0.000 ± 0.000	0.939 ± 0.034	-0.042 ± 0.010	7.417	2.178	1.476	0.978	1.598
Baek et al. [124]	-	190	HSG†	0.291 ± 0.000	0.041 ± 0.000	0.348 ± 0.016	0.031 ± 0.004	0.174	0.052	0.229	0.998	1.033
Belli et al. [60]	7.7	5.04	SQ20B†	0.150 ± 0.050	0.011 ± 0.012	0.155 ± 0.011	0.011 ± 0.002	0.164	0.147	0.384	0.994	0.898
Belli et al. [60]	19.8	1.49	SQ20B†	0.230 ± 0.040	0.004 ± 0.007	0.234 ± 0.018	0.003 ± 0.003	0.69	0.683	0.826	0.985	0.926
Belli et al. [60]	30	0.88	SQ20B†	0.450 ± 0.050	0.000 ± 0.000	0.567 ± 0.018	-0.033 ± 0.003	9.909	0.762	0.873	0.986	1.246
Belli et al. [60]	7.7	5.01	SCC25†	0.410 ± 0.011	0.092 ± 0.036	0.440 ± 0.050	0.082 ± 0.016	0.676	0.634	0.797	0.991	1.035
Belli et al. [60]	19.7	1.49	SCC25†	0.870 ± 0.050	0.000 ± 0.000	0.894 ± 0.026	-0.007 ± 0.009	0.108	0.094	0.306	0.998	1.281
Belli et al. [60]	29.5	29.5	SCC25†	0.810 ± 0.030	0.000 ± 0.000	0.915 ± 0.044	-0.029 ± 0.011	0.566	0.259	0.509	0.995	1.220
Belli et al. [60]	9.11	4.05	M10†	0.500 ± 0.030	0.000 ± 0.000	0.539 ± 0.020	-0.006 ± 0.004	0.381	0.288	0.536	0.996	0.908
Belli et al. [60]	21.4	1.35	M10†	0.490 ± 0.020	0.000 ± 0.000	0.503 ± 0.036	-0.001 ± 0.006	1.209	1.182	1.087	0.954	0.888
Belli et al. [60]	33	0.79	M10†	0.930 ± 0.050	0.000 ± 0.000	1.079 ± 0.047	-0.037 ± 0.012	0.745	0.382	0.618	0.980	1.769
Belli et al. [60]	7.7	5.04	HF19†	0.550 ± 0.010	0.000 ± 0.000	0.526 ± 0.047	0.000 ± 0.011	6.075	4.269	2.066	0.974	0.830
Belli et al. [60]	19.5	1.49	HF19†	0.540 ± 0.020	0.000 ± 0.000	0.559 ± 0.043	-0.006 ± 0.011	1.608	1.552	1.246	0.975	0.843
Belli et al. [60]	29	0.88	HF19†	0.520 ± 0.010	0.000 ± 0.000	0.687 ± 0.043	-0.035 ± 0.008	5.907	1.927	1.388	0.963	0.854
Belli et al. [26]	7.7	5.01	V79	0.289 ± 0.023	0.024 ± 0.006	0.304 ± 0.012	0.022 ± 0.003	0.487	0.420	0.648	0.995	1.048
Belli et al. [26]	11	3.2	V79	0.372 ± 0.032	0.036 ± 0.009	0.375 ± 0.017	0.035 ± 0.004	0.871	0.863	0.929	0.994	1.307
Belli et al. [26]	20	1.41	V79	0.469 ± 0.029	0.043 ± 0.009	0.502 ± 0.017	0.034 ± 0.005	0.484	0.366	0.605	0.997	1.559
Belli et al. [26]	30.5	0.76	V79	0.721 ± 0.017	0.000 ± 0.000	0.730 ± 0.027	-0.001 ± 0.008	1.785	1.719	1.311	0.990	1.808
Belli et al. [26]	34.6	0.65	V79	0.653 ± 0.025	0.000 ± 0.000	0.791 ± 0.035	-0.034 ± 0.008	4.268	1.766	1.329	0.984	1.672
Belli et al. [26]	37.8	0.57	V79	0.580 ± 0.019	0.000 ± 0.000	0.701 ± 0.055	-0.025 ± 0.011	3.887	1.746	1.321	0.985	1.488
Bettega et al. [61]	11	3.18	C3H10T12	0.470 ± 0.060	0.019 ± 0.014	0.487 ± 0.038	0.019 ± 0.010	1.518	1.242	1.115	0.978	1.575
Bettega et al. [61]	19.7	1.460	C3H10T12	0.430 ± 0.060	0.038 ± 0.013	0.432 ± 0.028	0.037 ± 0.006	0.701	0.697	0.835	0.989	1.619
Bettega et al. [61]	28.8	0.870	C3H10T12	0.550 ± 0.012	0.053 ± 0.031	0.578 ± 0.034	0.047 ± 0.009	0.233	0.178	0.422	0.995	2.033
Bettega et al. [61]	31.6	0.770	C3H10T12	0.670 ± 0.040	0.000 ± 0.000	0.752 ± 0.053	-0.029 ± 0.016	0.815	0.497	0.705	0.988	1.814
Bettega et al. [61]	32.5	0.740	C3H10T12	0.750 ± 0.040	0.000 ± 0.000	0.836 ± 0.058	-0.025 ± 0.019	0.445	0.314	0.560	0.990	2.128
Bettega et al. [61]	33.2	0.720	C3H10T12	1.020 ± 0.060	0.000 ± 0.000	0.998 ± 0.067	0.007 ± 0.032	0.384	0.375	0.612	0.993	2.835
Bettega et al. [176]	-	65	SCC25†	0.570 ± 0.060	0.012 ± 0.011	0.596 ± 0.038	0.008 ± 0.007	2.776	2.640	1.625	0.975	0.999
Bettega et al. [176]	-	65	SCC25†	1.230 ± 0.120	0.000 ± 0.000	1.392 ± 0.024	-0.065 ± 0.004	8.951	0.511	0.715	0.992	1.058
Bettega et al. [176]	-	65	SCC25†	0.700 ± 0.050	0.018 ± 0.012	0.732 ± 0.037	0.015 ± 0.008	1.756	1.420	1.192	0.982	1.237
Bettega et al. [176]	-	65	SCC25†	0.610 ± 0.050	0.010 ± 0.008	0.630 ± 0.033	0.008 ± 0.005	1.533	1.344	1.159	0.983	1.051
Bettega et al. [176]	-	65	SCC25†	0.830 ± 0.080	0.001 ± 0.014	0.855 ± 0.043	-0.002 ± 0.010	1.686	1.606	1.267	0.981	1.355

Study	**LET** (keV/μm)	**Energy** (MeV)	**Cell Line**	$\alpha_{Orig.}$ (±SE)	$\beta_{Orig.}$ (±SE)	α (±SE)	β (±SE)	$\chi^2_{\mathbf{Orig.}}$	χ^2	**RMSE**	R^2	**RBE** 10%
Calugaru et al. [177]	-	76	HeLa†	0.293 ± 0.026	0.020 ± 0.012	0.288 ± 0.015	0.021 ± 0.008	1.387	1.348	1.161	0.995	0.99
Calugaru et al. [177]	-	201	HeLa†	0.213 ± 0.016	0.054 ± 0.007	0.177 ± 0.002	0.055 ± 0.001	4.164	0.08	0.282	1.000	1.105
Calugaru et al. [177]	-	76	SQ20B†	0.029 ± 0.013	0.012 ± 0.002	0.027 ± 0.004	0.013 ± 0.000	2.468	1.759	1.326	0.997	1.099
Folkard et al. [96]	17	0.72	V79	0.130 ± 0.036	0.078 ± 0.007	0.132 ± 0.028	0.085 ± 0.006	2.273	1.506	1.227	0.978	1.261
Folkard et al. [96]	24	1.15	V79	0.330 ± 0.038	0.066 ± 0.005	0.343 ± 0.013	0.065 ± 0.002	0.667	0.565	0.752	0.993	1.463
Folkard et al. [96]	32	0.76	V79	1.030 ± 0.016	0.000 ± 0.000	1.098 ± 0.026	-0.010 ± 0.003	3.009	2.039	1.428	0.962	2.645
Folkard et al. [126]	10	3.66	V79	0.320 ± 0.058	0.039 ± 0.011	0.231 ± 0.023	0.057 ± 0.005	0.809	0.352	0.593	0.99	1.219
Folkard et al. [126]	18	1.83	V79	0.450 ± 0.035	0.028 ± 0.006	0.468 ± 0.024	0.029 ± 0.005	2.255	1.767	1.329	0.969	1.429
Folkard et al. [126]	28	1.07	V79	0.740 ± 0.025	0.011 ± 0.004	0.740 ± 0.013	0.015 ± 0.003	1.351	0.658	0.811	0.995	1.923
Goodhead et al. [169]	-	1.16	V79	0.300 ± 0.060	0.052 ± 0.008	0.247 ± 0.014	0.060 ± 0.003	14.293	12.946	3.598	0.948	-
Goodhead et al. [169]	20	1.38	V79	0.420 ± 0.050	0.019 ± 0.008	0.439 ± 0.007	0.014 ± 0.001	9.516	9.23	3.038	0.965	-
Goodhead et al. [169]	20	1.83	HeLa†	0.530 ± 0.120	0.084 ± 0.025	0.549 ± 0.036	0.083 ± 0.007	10.355	10.107	3.179	0.887	-
Goodhead et al. [169]	23	1.16	HeLa†	0.850 ± 0.160	0.037 ± 0.028	1.001 ± 0.047	0.022 ± 0.009	16.160	11.767	3.430	0.764	-
Guan et al. [27]	10.8	79.7	H460†	0.318 ± 0.000	0.154 ± 0.000	0.336 ± 0.040	0.147 ± 0.016	4.305	4.165	2.041	0.995	1.316
Guan et al. [27]	10.8	79.7	H460†	0.446 ± 0.000	0.341 ± 0.000	0.459 ± 0.079	0.336 ± 0.048	3.905	3.780	1.944	0.997	1.934
Guan et al. [27]	17.7	79.7	H460†	0.596 ± 0.000	0.662 ± 0.000	0.585 ± 0.062	0.665 ± 0.050	3.165	3.100	1.761	0.997	2.654
Guan et al. [27]	19	79.7	H460†	0.883 ± 0.000	0.956 ± 0.000	0.854 ± 0.106	1.002 ± 0.106	6.278	6.031	2.456	0.988	3.397
Gueulette et al. [125]	-	85	CHO	0.071 ± 0.044	0.085 ± 0.007	0.021 ± 0.012	0.086 ± 0.002	2.819	1.62	1.273	0.973	1.005
Howard et al. [123]	1.78	71	CHO	0.220 ± 0.040	0.030 ± 0.010	0.225 ± 0.008	0.033 ± 0.001	1.69	0.255	0.505	0.998	1.083
Howard et al. [123]	3.36	71	CHO	0.240 ± 0.060	0.050 ± 0.010	0.236 ± 0.015	0.053 ± 0.003	0.251	0.177	0.421	0.996	1.283
Howard et al. [123]	4.35	71	CHO	0.340 ± 0.100	0.050 ± 0.020	0.228 ± 0.036	0.079 ± 0.009	0.831	0.287	0.536	0.993	1.462
Howard et al. [123]	0.990	160	CHO	0.220 ± 0.030	0.030 ± 0.010	0.222 ± 0.011	0.034 ± 0.002	2.295	0.499	0.706	0.993	1.095
Howard et al. [123]	2.26	160	CHO	0.230 ± 0.040	0.040 ± 0.010	0.209 ± 0.022	0.041 ± 0.005	0.902	0.625	0.79	0.991	1.141
Howard et al. [123]	4.19	160	CHO	0.290 ± 0.070	0.050 ± 0.010	0.261 ± 0.019	0.052 ± 0.004	0.348	0.239	0.489	0.995	1.337
Howard et al. [123]	7.29	160	CHO	0.310 ± 0.100	0.060 ± 0.020	0.284 ± 0.048	0.071 ± 0.012	0.646	0.477	0.69	0.988	1.523
Howard et al. [123]	1.78	71	A549†	0.280 ± 0.050	0.030 ± 0.010	0.266 ± 0.010	0.033 ± 0.002	0.119	0.079	0.282	0.997	1.234
Howard et al. [123]	3.36	71	A549†	0.320 ± 0.070	0.030 ± 0.010	0.236 ± 0.029	0.051 ± 0.006	0.881	0.249	0.499	0.994	1.355
Howard et al. [123]	4.35	71	A549†	0.460 ± 0.070	0.030 ± 0.010	0.416 ± 0.044	0.040 ± 0.01	0.563	0.472	0.687	0.990	1.622
Howard et al. [123]	7.29	71	A549†	0.590 ± 0.080	0.010 ± 0.010	0.486 ± 0.056	0.030 ± 0.012	0.855	0.535	0.732	0.982	1.689
Howard et al. [123]	0.99	160	A549†	0.250 ± 0.040	0.030 ± 0.010	0.247 ± 0.009	0.027 ± 0.001	1.48	0.075	0.274	0.999	1.125
Howard et al. [123]	2.26	160	A549†	0.230 ± 0.040	0.020 ± 0.010	0.231 ± 0.021	0.024 ± 0.003	4.583	0.146	0.382	0.998	1.057
Howard et al. [123]	4.19	160	A549†	0.190 ± 0.050	0.040 ± 0.010	0.154 ± 0.019	0.043 ± 0.003	0.883	0.488	0.698	0.993	1.129
Howard et al. [123]	7.29	160	A549†	0.260 ± 0.070	0.040 ± 0.010	0.266 ± 0.008	0.042 ± 0.002	0.141	0.016	0.127	1.000	1.325
Mastumoto et al. [178]	-	190	HSG†	0.220 ± 0.070	0.050 ± 0.010	0.192 ± 0.009	0.054 ± 0.002	0.296	0.195	0.442	0.998	1.241
Mastumoto et al. [178]	-	190	HSG†	0.280 ± 0.070	0.050 ± 0.010	0.268 ± 0.012	0.05 ± 0.002	1.249	0.410	0.641	0.997	1.334
Mastumoto et al. [178]	-	190	HSG†	0.250 ± 0.070	0.050 ± 0.010	0.247 ± 0.007	0.051 ± 0.001	0.511	0.376	0.613	0.998	1.311
Mastumoto et al. [178]	-	190	HSG†	0.260 ± 0.030	0.050 ± 0.010	0.238 ± 0.007	0.057 ± 0.001	3.218	0.409	0.639	0.998	1.341
Mastumoto et al. [178]	-	190	HSG†	0.420 ± 0.010	0.050 ± 0.010	0.422 ± 0.012	0.046 ± 0.003	2.780	1.039	1.019	0.998	1.602
Mastumoto et al. [178]	-	190	HSG†	0.410 ± 0.040	0.050 ± 0.010	0.405 ± 0.009	0.052 ± 0.002	0.851	0.814	0.902	0.998	1.616

Study	**LET** (keV/μm)	**Energy** (MeV)	**Cell Line**	$\alpha_{Orig.}$ (±SE)	$\beta_{Orig.}$ (±SE)	α (±SE)	β (±SE)	$\chi^2_{Orig.}$	χ^2	**RMSE**	R^2	**RBE** 10%
Mastumoto et al. [178]	-	190	HSG†	0.380 ± 0.030	0.050 ± 0.01	0.383 ± 0.014	0.052 ± 0.003	1.798	1.118	1.058	0.996	1.575
Mastumoto et al. [178]	-	190	HSG†	0.440 ± 0.060	0.030 ± 0.01	0.426 ± 0.014	0.034 ± 0.003	0.421	0.346	0.588	0.996	1.511
Mastumoto et al. [178]	-	190	HSG†	0.420 ± 0.070	0.030 ± 0.010	0.429 ± 0.006	0.028 ± 0.001	0.305	0.208	0.456	0.999	1.467
Tang et al. [130]	-	65	CHO	0.200 ± 0.035	0.022 ± 0.007	0.218 ± 0.011	0.020 ± 0.001	0.514	0.42	0.648	0.996	1.056
Tang et al. [130]	-	65	CHO	0.210 ± 0.044	0.022 ± 0.007	0.220 ± 0.011	0.023 ± 0.001	1.08	0.287	0.536	0.997	1.106
Tang et al. [130]	-	65	CHO	0.220 ± 0.036	0.026 ± 0.007	0.216 ± 0.007	0.026 ± 0.001	0.116	0.081	0.284	0.998	1.133
Tang et al. [130]	-	65	CHO	0.249 ± 0.045	0.025 ± 0.008	0.264 ± 0.009	0.019 ± 0.001	7.667	0.636	0.797	0.997	1.150
Wouters et al. [38]	1.1	160	V79	0.114 ± 0.027	0.048 ± 0.007	0.122 ± 0.001	0.047 ± 0.001	0.012	<0.001	0.015	0.999	1.047
Wouters et al. [38]	2.06	160	V79	0.104 ± 0.017	0.055 ± 0.003	0.112 ± 0.001	0.054 ± 0.001	0.007	0.001	0.033	0.999	1.086
Wouters et al. [38]	2.41	160	V79	0.113 ± 0.014	0.054 ± 0.003	0.123 ± 0.001	0.053 ± 0.001	0.022	<0.001	0.014	0.999	1.100
Wouters et al. [38]	3.2	160	V79	0.133 ± 0.014	0.055 ± 0.003	0.141 ± 0.001	0.054 ± 0.001	0.005	<0.001	0.02	0.999	1.132
Wouters et al. [38]	4.74	160	V79	0.141 ± 0.007	0.059 ± 0.001	0.154 ± 0.001	0.057 ± 0.001	0.012	0.002	0.044	0.999	1.180
Wouters et al. [38]	1.03	230	V79	0.103 ± 0.016	0.051 ± 0.003	0.103 ± 0.001	0.05 ± 0.001	0.029	<0.001	0.01	0.999	1.044
Wouters et al. [38]	1.95	230	V79	0.103 ± 0.008	0.054 ± 0.002	0.101 ± 0.001	0.054 ± 0.001	0.002	<0.001	0.015	0.999	1.073
Wouters et al. [38]	2.28	230	V79	0.102 ± 0.008	0.057 ± 0.002	0.103 ± 0.001	0.057 ± 0.001	<0.001	<0.001	0.019	0.999	1.100
Wouters et al. [38]	2.95	230	V79	0.123 ± 0.011	0.057 ± 0.002	0.122 ± 0.001	0.057 ± 0.001	0.003	<0.001	0.012	0.999	1.128
Wouters et al. [38]	4.02	230	V79	0.121 ± 0.013	0.063 ± 0.002	0.122 ± 0.001	0.062 ± 0.001	0.007	<0.001	0.015	0.999	1.173

Appendix G

Hypothesis Testing Results: Linear-Quadratic Model Fits to Experimental Data

This appendix consists of the hypothesis testing results performed on fits using the linear and non-linear forms of the LQ model to the experimental data in Appendix D. The details of the test listed below are provided in Appendix F. **Note:** The results of each test indicate the following.

- Kolmogorov-Smirnov (KS) Test: ✓= Residual errors distributed normally, ✗= Residual errors are not distributed normally.

- Anderson-Darling (AD) Test: ✓= Residual errors distributed normally, ✗= Residual errors are not distributed normally.

- Breusch-Pagan (BP) Test:
  ✓= residual errors are homoscedastic (have constant variance),
  ✗= residual errors are heteroscedastic (do not have constant variance).

- White Test: ✓= residual errors are homoscedastic (have constant variance), ✗= residual errors are heteroscedastic (do not have constant variance).

- Durbin-Watson (DW) Test: ✓= Residual errors do not have autocorrelation, ✗+/- = Residual errors have positive/negative autocorrelation,
  - = inconclusive.

If either the LET, proton energy, or cell line are not reported in the literature, they are indicated with a "-". Cell lines of human descent are indicated with a dagger

(†). All data with non-Normally distributed residual errors (i.e. that fail the KS or AD tests) are not eligible for the BP test and are therefore indicated with a "-".

Table G.1: Hypothesis and "regression assumption" testing results for LQ fits using linear regression analysis.

Data	LET (keV/μm)/ Energy (MeV)/ Cell Line	Sample Size	KS Test	AD Test	BP Test	White Test	DW Test
Ando et al. [174]	-/235/SCC61†	6	✓	✓	✗	✓	✓
Ando et al. [174]	-/235/SCC61†	6	✓	✓	✓	✓	✓
Ando et al. [174]	-/235/NB1RGB†	6	✓	✓	✓	✓	✓
Ando et al. [174]	-/235/NB1RGB†	6	✓	✓	✓	✓	✓
Ando et al. [174]	-/235/V79	6	✗	✗	-	✓	✓
Ando et al. [174]	-/235/V79	6	✓	✓	✓	✓	-
Antoccia et al. [175]	28.5/0.8/HFFF2†	8	✗	✗	-	✓	-
Belli et al. [60]	7.7/5.04/SQ20B†	6	✓	✓	✓	✓	✓
Belli et al. [60]	19.8/1.49/SQ20B†	6	✓	✓	✓	✓	-
Belli et al. [60]	7.7/5.01/SCC25†	6	✗	✗	-	✓	-
Belli et al. [60]	19.7/1.49/SCC25†	5	✓	✓	✓	✓	-
Belli et al. [60]	29.5/29.5/SCC25†	5	✗	✗	-	✓	-
Belli et al. [60]	9.11/4.05/M10†	6	✗	✗	-	✓	-
Belli et al. [60]	21.4/1.35/M10†	7	✗	✗	-	✓	-
Belli et al. [60]	33/0.79/M10†	6	✗	✗	-	✓	-
Belli et al. [60]	7.7/5.04/HF19†	6	✓	✓	✓	✓	-
Belli et al. [60]	19.5/1.49/HF19†	6	✓	✓	✗	✓	-
Belli et al. [60]	29/0.88/HF19†	6	✓	✗	-	✓	-
Belli et al. [26]	7.7/5.01/V79	7	✓	✓	✓	✓	✓
Belli et al. [26]	11/3.2/V79	6	✓	✓	✓	✓	✓
Belli et al. [26]	20/1.41/V79	6	✓	✓	✓	✓	✓
Belli et al. [26]	30.5/0.76/V79	6	✓	✓	✓	✓	✓
Belli et al. [26]	34.6/0.65/V79	6	✓	✓	✗	✓	-
Belli et al. [26]	37.8/0.57/V79	7	✗	✗	-	✓	-
Bettaga et al. [61]	11/3.18/C3H10T12†	6	✓	✓	✓	✓	-
Bettaga et al. [61]	19.7/1.46/C3H10T12†	6	✗	✗	-	✓	✓
Bettaga et al. [61]	31.6/0.77/C3H10T12†	5	✓	✓	✓	✓	-
Bettaga et al. [61]	32.5/0.74/C3H10T12†	5	✗	✗	-	✓	✓
Bettaga et al. [61]	33.2/0.72/C3H10T12†	6	✓	✓	✓	✓	-
Bettega et al. [176]	-/65/SCC25†	6	✓	✓	✓	✓	✓
Bettega et al. [176]	-/65/SCC25†	6	✗	✗	-	✓	-
Bettega et al. [176]	-/65/SCC25†	6	✓	✓	✓	✓	✓
Bettega et al. [176]	-/65/SCC25†	6	✗	✗	-	✓	✓
Bettega et al. [176]	-/65/SCC25†	6	✓	✓	✓	✓	✓
Calugaru et al. [177]	-/76/HeLa†	6	✗	✗	-	✓	-
Calugaru et al. [177]	-/201/HeLa†	7	✗	✗	-	✓	-
Calugaru et al. [177]	-/76/SQ20B†	9	✓	✓	✓	✓	✓
Folkard et al. [96]	17/0.72/V79	8	✗	✓	-	✓	✓
Folkard et al. [96]	24/1.15/V79	8	✗	✗	-	✓	-
Folkard et al. [96]	32/0.76/V79	8	✓	✗	-	✓	✗+
Folkard et al. [126]	10/3.66/V79	6	✓	✓	✓	✓	-
Folkard et al. [126]	18/1.83/V79	8	✓	✓	✓	✓	✓
Folkard et al. [126]	28/1.07/V79	7	✓	✓	✓	✓	✓
Goodhead et al. [169]	20/1.38/V79	23	✓	✓	✗	✗	-
Goodhead et al. [169]	20/1.83/HeLa†	11	✓	✓	✓	✗	✓
Goodhead et al. [169]	23/1.16/HeLa†	12	✗	✗	-	✓	✓
Guan et al. [27]	10.8/79.7/H460†	5	✓	✓	✓	✓	✓
Guan et al. [27]	15.2/79.7/H460†	4	✓	✓	✓	✓	✓
Guan et al. [27]	17.7/79.7/H460†	5	✗	✗	-	✓	✓
Guan et al. [27]	19/79.7/H460†	6	✓	✓	✓	✓	✓
Gueulette et al. [125]	-/85/CHO	14	✗	✗	-	✗	✗+
Howard et al. [123]	1.78/71/CHO	6	✓	✓	✓	✓	-
Howard et al. [123]	3.36/71/CHO	6	✓	✓	✓	✓	✓
Howard et al. [123]	4.35/71/CHO	5	✓	✓	✓	✓	✓
Howard et al. [123]	0.99/160/CHO	7	✗	✗	-	✓	✓
Howard et al. [123]	2.26/160/CHO	6	✗	✗	-	✓	-
Howard et al. [123]	4.19/160/CHO	6	✓	✓	✓	✓	-
Howard et al. [123]	7.29/160/CHO	5	✓	✓	✓	✓	-

Data	**LET (keV/μm)/ Energy (MeV)/ Cell Line**	**Sample Size**	**KS Test**	**AD Test**	**BP Test**	**White Test**	**DW Test**
Howard et al. [123]	1.78/71/A549†	6	✗	✗	-	✓	✓
Howard et al. [123]	3.36/71/A549†	5	✓	✓	✓	✓	✓
Howard et al. [123]	7.29/71/A549†	5	✓	✓	✓	✓	✓
Howard et al. [123]	0.99/160/A549†	5	✓	✓	✓	✓	✓
Howard et al. [123]	2.26/160/A549†	6	✓	✓	✓	✓	✓
Howard et al. [123]	4.19/160/A549†	6	✓	✓	✓	✓	-
Howard et al. [123]	7.29/160/A549†	5	✓	✓	✓	✓	✓
Matsumoto et al. [178]	-/190/HSG†	7	✗	✗	-	✓	-
Matsumoto et al. [178]	-/190/HSG†	7	✗	✗	-	✓	-
Matsumoto et al. [178]	-/190/HSG†	7	✗	✗	-	✓	✗+
Matsumoto et al. [178]	-/190/HSG†	7	✗	✗	-	✓	-
Matsumoto et al. [178]	-/190/HSG†	7	✗	✗	-	✓	-
Matsumoto et al. [178]	-/190/HSG†	7	✗	✗	-	✓	-
Matsumoto et al. [178]	-/190/HSG†	6	✗	✗	-	✓	-
Matsumoto et al. [178]	-/190/HSG†	7	✗	✗	-	✓	-
Matsumoto et al. [178]	-/190/HSG†	7	✗	✗	-	✓	-
Tang et al. [130]	-/65/CHO	6	✗	✗	-	✓	-
Tang et al. [130]	-/65/CHO	6	✗	✗	-	✓	-
Tang et al. [130]	-/65/CHO	6	✓	✓	✓	✓	-
Tang et al. [130]	-/65/CHO	7	✗	✗	-	✓	-
Wouters et al. [38]	1.1/160/V79	15	✓	✓	✗	✗	✓
Wouters et al. [38]	2.06/160/V79	19	✗	✗	-	✗	✓
Wouters et al. [38]	2.41/160/V79	14	✗	✗	-	✓	✗+
Wouters et al. [38]	3.2/160/V79	18	✗	✗	-	✗	✓
Wouters et al. [38]	4.74/160/V79	16	✗	✗	-	✗	✗+
Wouters et al. [38]	1.03/230/V79	9	✓	✓	✓	✓	-
Wouters et al. [38]	1.95/230/V79	14	✓	✗	-	✗	-
Wouters et al. [38]	2.28/230/V79	16	✓	✓		✓	✗+
Wouters et al. [38]	2.95/230/V79	16	✓	✓	✗	✗	✓
Wouters et al. [38]	4.02/230/V79	15	✓	✗	-	✗	✓

Table G.2: Hypothesis and "regression assumption" testing results for LQ fits using non-linear regression analysis. **Note:** The BP test is not valid for fits with non-Normally distributed errors. Therefore, BP test result of the fits that fail the KS and AD tests are denoted with a "-".

Data	LET (keV/μm) Energy (MeV)/ Cell Line	Sample Size	KS Test	AD Test	BP Test	White Test
Ando et al. [174]	-/235/SCC61†	6	✓	✓	✓	✓
Ando et al. [174]	-/235/SCC61†	6	✓	✓	✓	✓
Ando et al. [174]	-/235/NB1RGB†	6	✓	✓	✓	✓
Ando et al. [174]	-/235/NB1RGB†	6	✓	✓	✓	✓
Ando et al. [174]	-/235/V79	6	✗	✗	-	✓
Ando et al. [174]	-/235/V79	6	✓	✓	✓	✓
Antoccia et al. [175]	28.5/0.8/HFFF2†	8	✗	✗	-	✓
Belli et al. [60]	7.7/5.04/SQ20B†	6	✓	✓	✓	✓
Belli et al. [60]	19.8/1.49/SQ20B†	6	✓	✓	✓	✓
Belli et al. [60]	7.7/5.01/SCC25†	6	✗	✗	-	✓
Belli et al. [60]	19.7/1.49/SCC25†	5	✓	✓	✓	✓
Belli et al. [60]	29.5/29.5/SCC25†	5	✗	✗	-	✓
Belli et al. [60]	9.11/4.05/M10†	6	✓	✓	✓	✓
Belli et al. [60]	21.4/1.35/M10†	7	✓	✓	✓	✓
Belli et al. [60]	33/0.79/M10†	6	✓	✓	✓	✓
Belli et al. [60]	7.7/5.04/HF19†	6	✓	✓	✓	✓
Belli et al. [60]	19.5/1.49/HF19†	6	✓	✓	✓	✓
Belli et al. [60]	29/0.88/HF19†	6	✓	✗	✓	✓
Belli et al. [26]	7.7/5.01/V79	7	✓	✓	✓	✓
Belli et al. [26]	11/3.2/V79	6	✓	✓	✓	✓
Belli et al. [26]	20/1.41/V79	6	✓	✓	✓	✓
Belli et al. [26]	30.5/0.76/V79	6	✓	✓	✗	✓
Belli et al. [26]	34.6/0.65/V79	6	✗	✗	-	✓
Belli et al. [26]	37.8/0.57/V79	7	✗	✗	-	✓
Bettega et al. [61]	11/3.18/C3H10T12	6	✓	✓	✓	✓
Bettega et al. [61]	19.7/1.46/C3H10T12	6	✓	✓	✓	✓
Bettega et al. [61]	28.8/0.87/C3H10T12	5	✓	✓	✓	✗
Bettega et al. [61]	31.6/0.77/C3H10T12	5	✓	✓	✓	✓
Bettega et al. [61]	32.5/0.74/C3H10T12	5	✓	✓	✓	✓
Bettega et al. [61]	33.2/0.72/C3H10T12	6	✓	✓	✗	✓
Bettega et al. [176]	-/65/SCC25†	6	✓	✓	✗	✓
Bettega et al. [176]	-/65/SCC25†	6	✗	✗	-	✓
Bettega et al. [176]	-/65/SCC25†	6	✓	✓	✓	✓
Bettega et al. [176]	-/65/SCC25†	6	✓	✓	✓	✓
Bettega et al. [176]	-/65/SCC25†	6	✓	✓	✓	✓
Calugaru et al. [177]	-/76/HeLa†	6	✓	✓	✓	✓
Calugaru et al. [177]	-/201/HeLa†	7	✗	✗	-	✓
Calugaru et al. [177]	-/76/SQ20B†	9	✓	✓	✓	✓
Folkard et al. [96]	17/0.72/V79	8	✓	✓	✓	✓
Folkard et al. [96]	24/1.15/V79	8	✗	✗	-	✓
Folkard et al. [96]	32/0.76/V79	8	✗	✗	-	✓
Folkard et al. [126]	10/3.66/V79	6	✓	✓	✓	✓
Folkard et al. [126]	18/1.83/V79	8	✗	✗	-	✓
Folkard et al. [126]	28/1.07/V79	7	✗	✗	-	✓
Goodhead et al. [169]	20/1.38/V79	23	✓	✓	✓	✗
Goodhead et al. [169]	20/1.83/HeLa†	11	✗	✗	-	✓
Goodhead et al. [169]	23/1.16/HeLa†	12	✗	✗	-	✓
Guan et al. [27]	10.8/79.7/H460†	5	✓	✓	✓	✓
Guan et al. [27]	15.2/79.7/H460†	4	✓	✓	✓	✓
Guan et al. [27]	17.7/79.7/H460†	5	✗	✗	-	✓
Guan et al. [27]	19/79.7/H460†	6	✓	✓	✓	✓
Gueulette et al. [125]	Proximal/85/CHO	14	✓	✓	✓	✓
Howard et al. [123]	1.78/71/CHO	6	✓	✓	✓	✓
Howard et al. [123]	3.36/71/CHO	6	✓	✓	✗	✓
Howard et al. [123]	4.35/71/CHO	5	✓	✓	✓	✓

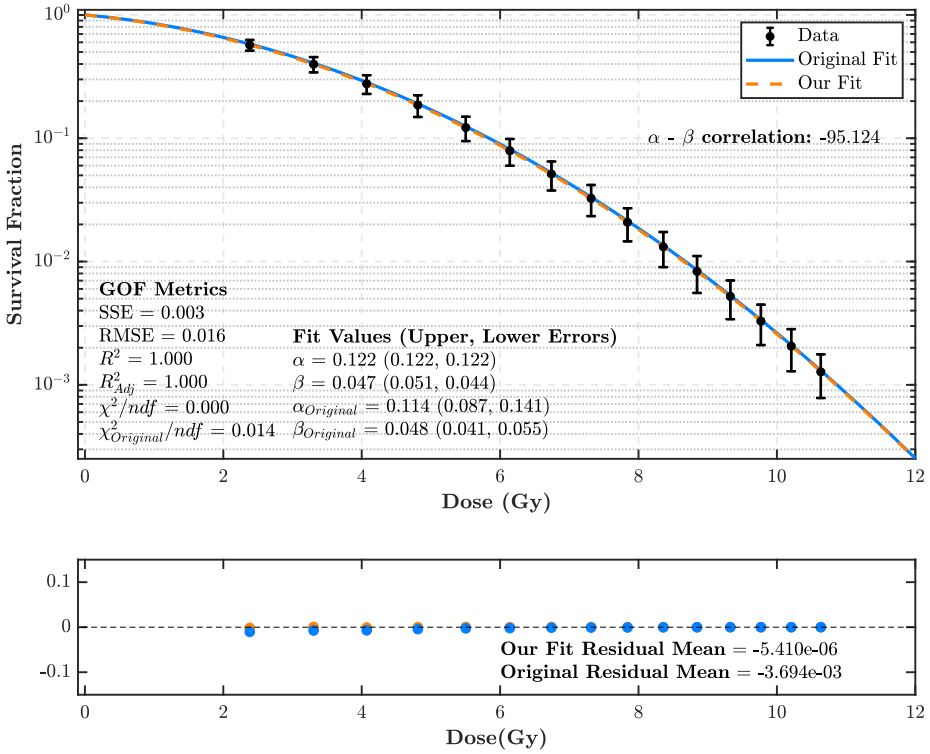
Data	**LET (keV/μm) Energy (MeV)/ Cell Line**	**Sample Size**	**KS Test**	**AD Test**	**BP Test**	**White Test**
Howard et al. [123]	0.99/160/CHO	7	✗	✗	-	✓
Howard et al. [123]	2.26/160/CHO	6	✗	✗	-	✓
Howard et al. [123]	4.19/160/CHO	6	✓	✓	✓	✓
Howard et al. [123]	7.29/160/CHO	5	✗	✗	-	✓
Howard et al. [123]	1.78/71/A549†	6	✗	✗	-	✓
Howard et al. [123]	3.36/71/A549†	5	✓	✓	✗	✓
Howard et al. [123]	7.29/71/A549†	5	✓	✓	✓	✓
Howard et al. [123]	0.99/160/A549†	5	✗	✗	-	✓
Howard et al. [123]	2.26/160/A549†	6	✓	✗	✓	✓
Howard et al. [123]	4.19/160/A549†	6	✗	✗	-	✓
Howard et al. [123]	7.29/160/A549†	5	✓	✓	✓	✓
Matsumoto et al. [178]	-/190/HSG†	7	✗	✗	-	✓
Matsumoto et al. [178]	-/190/HSG†	7	✗	✗	-	✓
Matsumoto et al. [178]	-/190/HSG†	7	✗	✗	-	✓
Matsumoto et al. [178]	-/190/HSG†	7	✗	✗	-	✓
Matsumoto et al. [178]	-/190/HSG†	7	✓	✓	✓	✓
Matsumoto et al. [178]	-/190/HSG†	7	✗	✗	-	✓
Matsumoto et al. [178]	-/190/HSG†	6	✓	✗	-	✓
Matsumoto et al. [178]	-/190/HSG†	7	✓	✓	✓	✓
Matsumoto et al. [178]	-/190/HSG†	7	✗	✗	-	✓
Tang et al. [130]	-/65/CHO	6	✗	✗	-	✓
Tang et al. [130]	-/65/CHO	6	✗	✗	-	✓
Tang et al. [130]	-/65/CHO	6	✓	✓	✗	✓
Tang et al. [130]	-/65/CHO	7	✓	✓	✓	✓
Wouters et al. [38]	1.1/160/V79	15	✗	✗	-	✗
Wouters et al. [38]	2.06/160/V79	19	✗	✗	-	✗
Wouters et al. [38]	2.41/160/V79	14	✗	✓	-	✓
Wouters et al. [38]	3.2/160/V79	18	✗	✗	-	✗
Wouters et al. [38]	4.74/160/V79	16	✗	✗	-	✗
Wouters et al. [38]	1.03/230/V79	9	✗	✗	-	✗
Wouters et al. [38]	1.95/230/V79	14	✗	✗	-	✗
Wouters et al. [38]	2.28/230/V79	16	✗	✗	-	✗
Wouters et al. [38]	2.95/230/V79	16	✗	✗	-	✗
Wouters et al. [38]	4.02/230/V79	15	✗	✗	-	✗

Appendix H

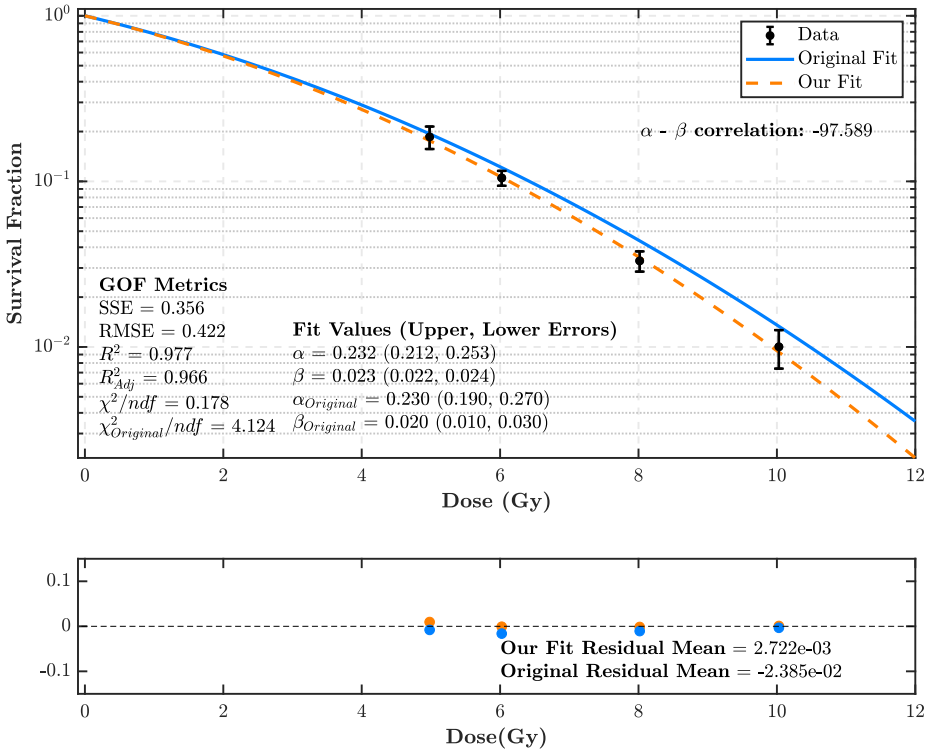
Additional Experimental Data Linear-Quadratic Model Regression Figures

This appendix contains additional linear and non-linear LQ model fit figures to aid discussion in Chapter 5. The data presented in the fits are from the world data outlined in Appendix D.

H.1 Linear Regression Fits

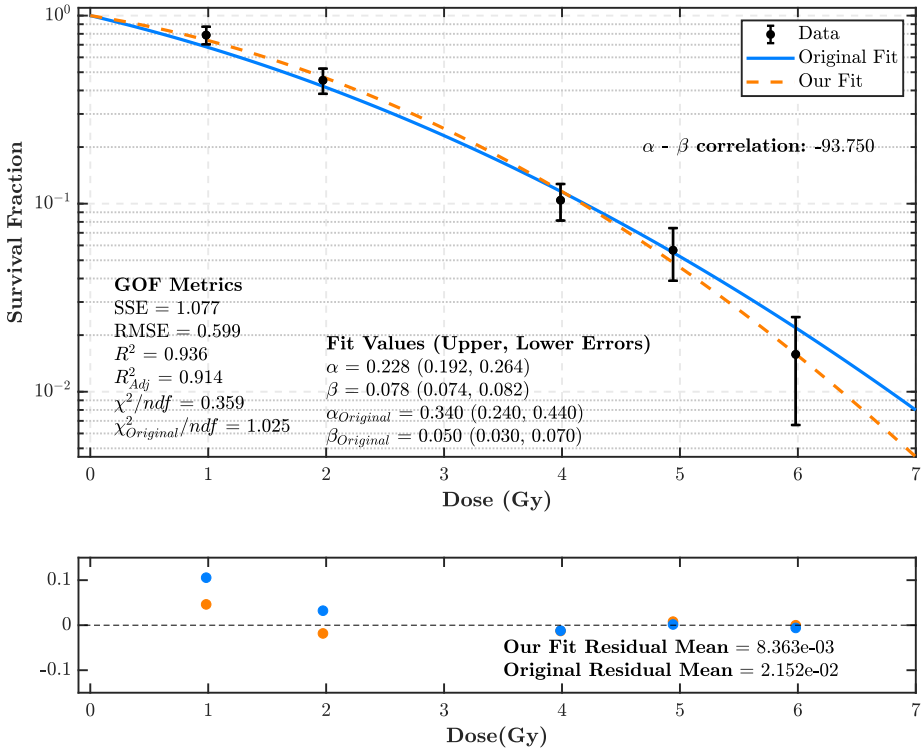


(a) 160MeV SOBP entrance region (1.1 keV/μm) V79 cells [38]

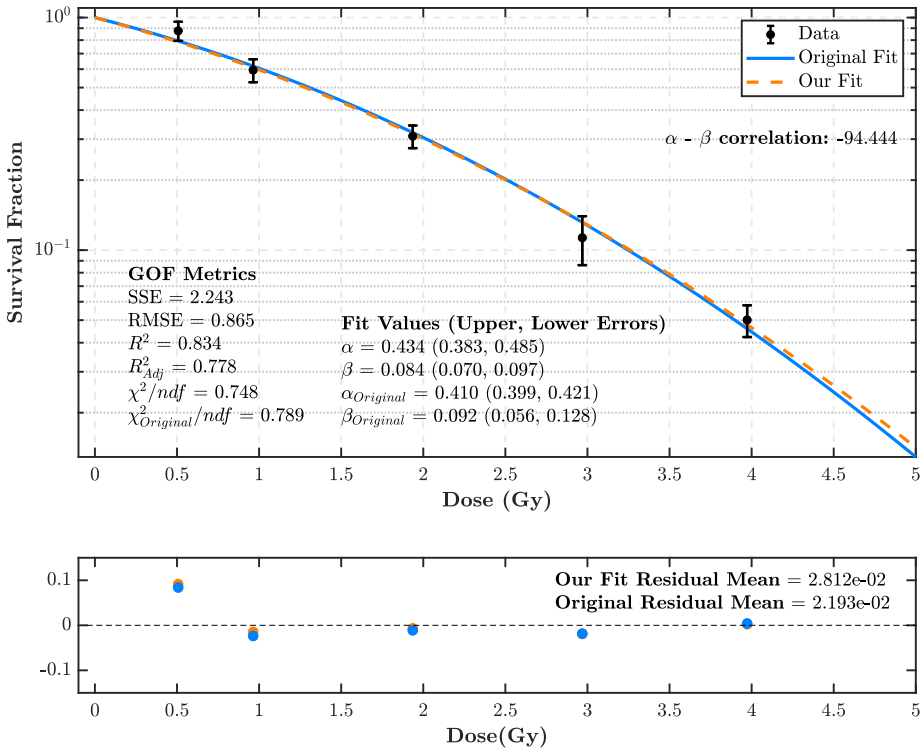


(b) 160MeV (2.26 keV/μm) protons incident on A549 cells from Howard et al. [123].

Figure H.1: Linearised LQ fits to experimental data. The second panel is a plot of the residuals between the observations and the linear LQ model fit.

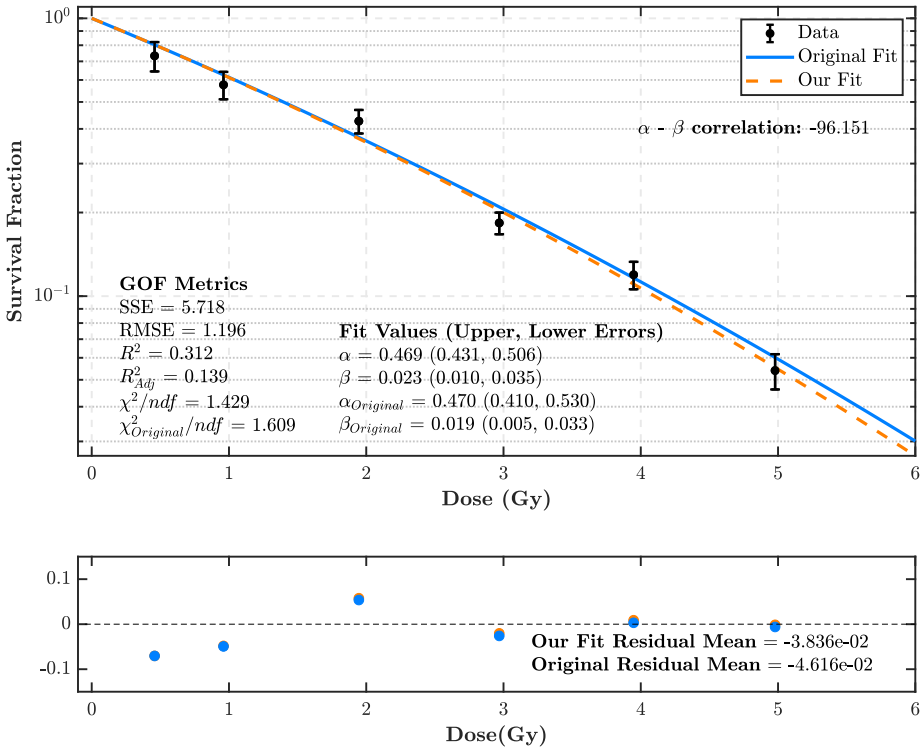


(a) 71MeV Pristine Bragg Peak (4.35 keV/μm) CHO cells from Howard et al. [123].

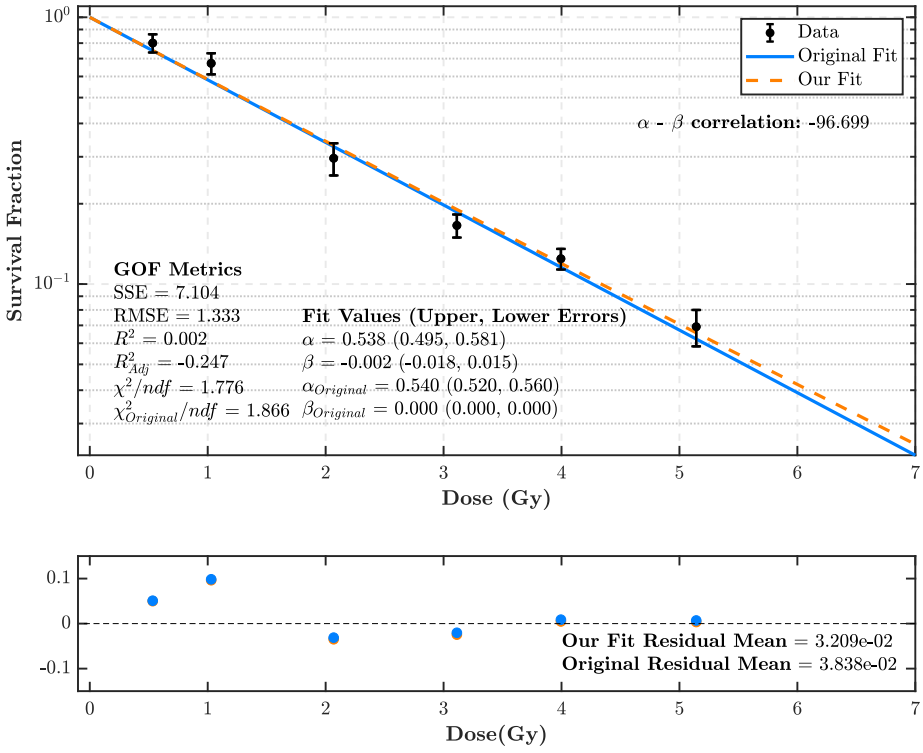


(b) 5.01MeV (7.7 keV/μm) protons incident on SCC25 cells from Belli et al. [60]

Figure H.2: Linearised LQ fits to experimental data. The second panel is a plot of the residuals between the observations and the linear LQ model fit.

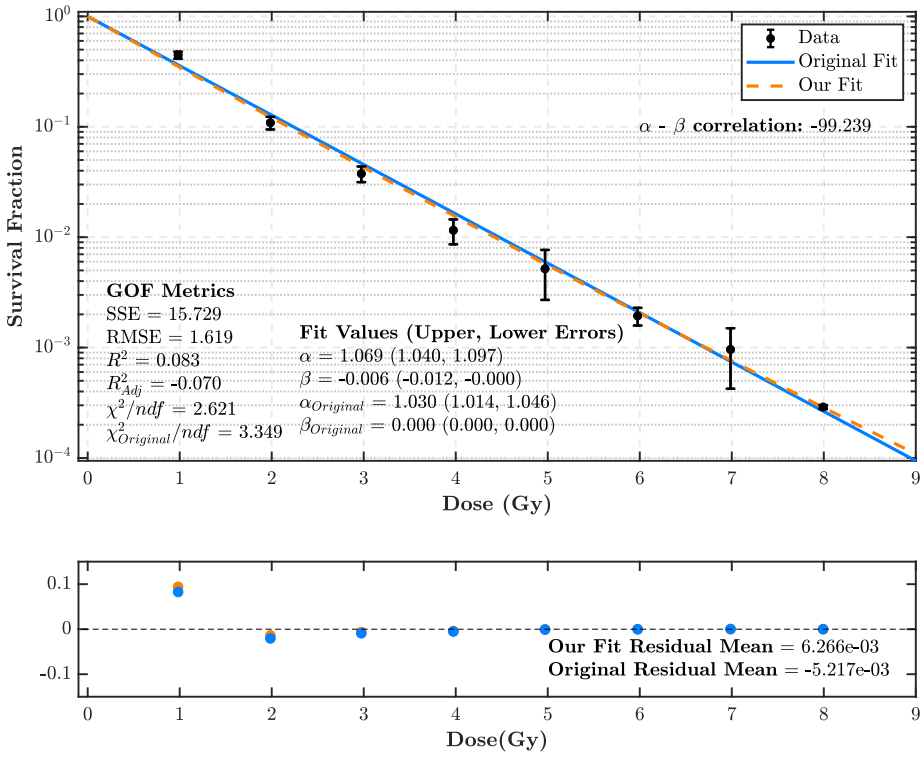


(a) 3.18MeV (11 keV/μm) protons incident on C3H10T12 cells from Bettaga et al. (1998) [61]. The second panel is a plot of the residuals between the observations and the linear LQ model.

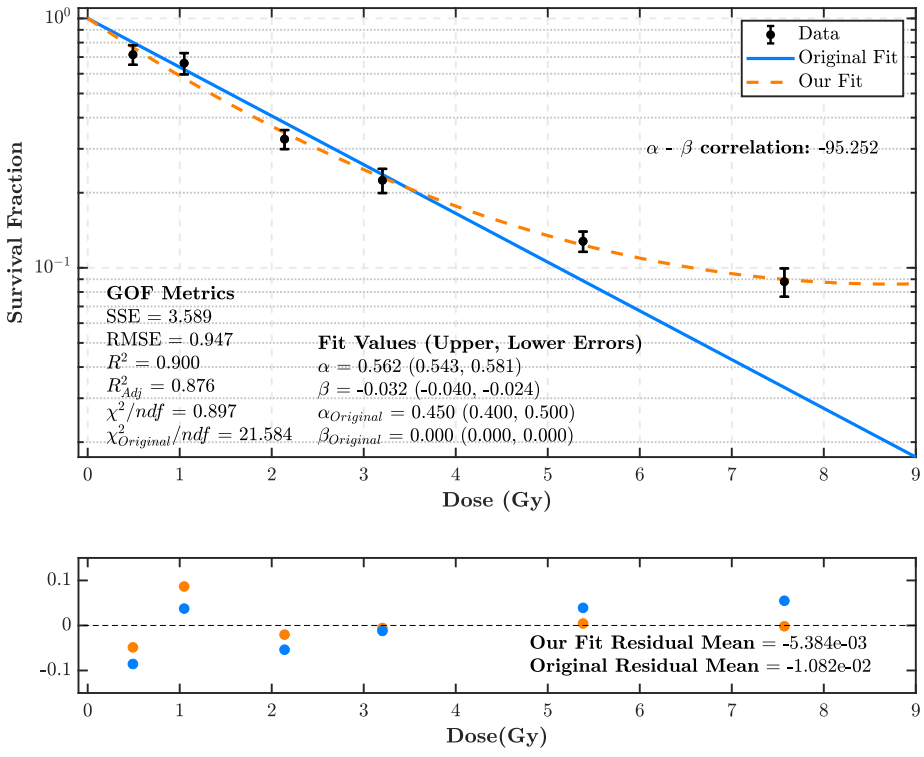


(b) 1.49MeV (19.5 keV/μm) protons incident on HF19 cells from Belli et al. [60].

Figure H.3: Linearised LQ fits to experimental data. The second panel is a plot of the residuals between the observations and the linear LQ model fit.



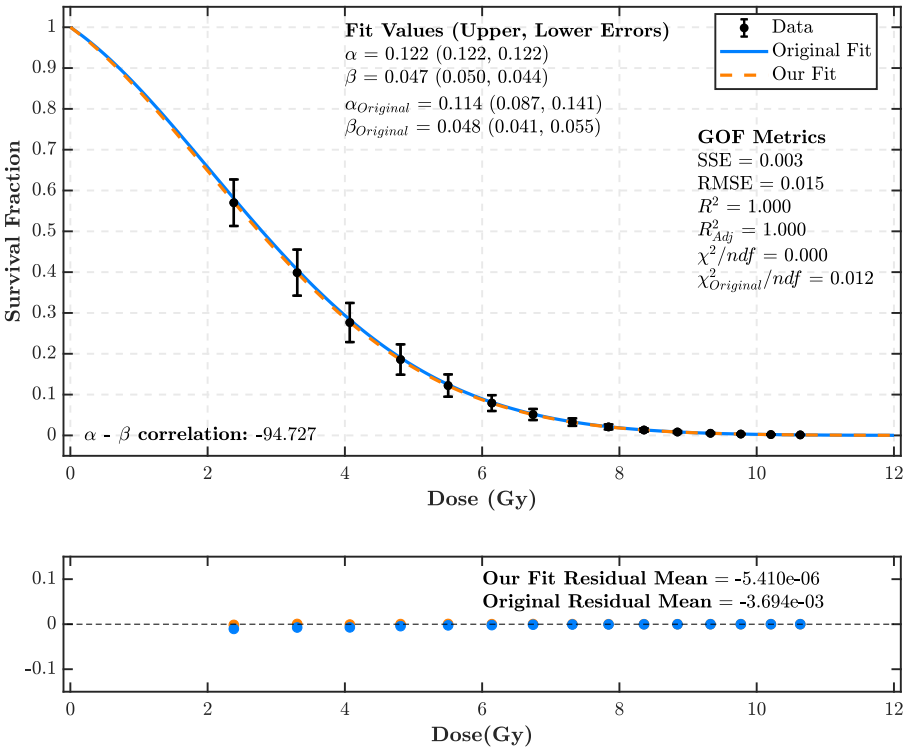
(a) 0.76MeV (32 keV/μm) protons incident on V79 cells on data from Folkard et al. (1989) [96].



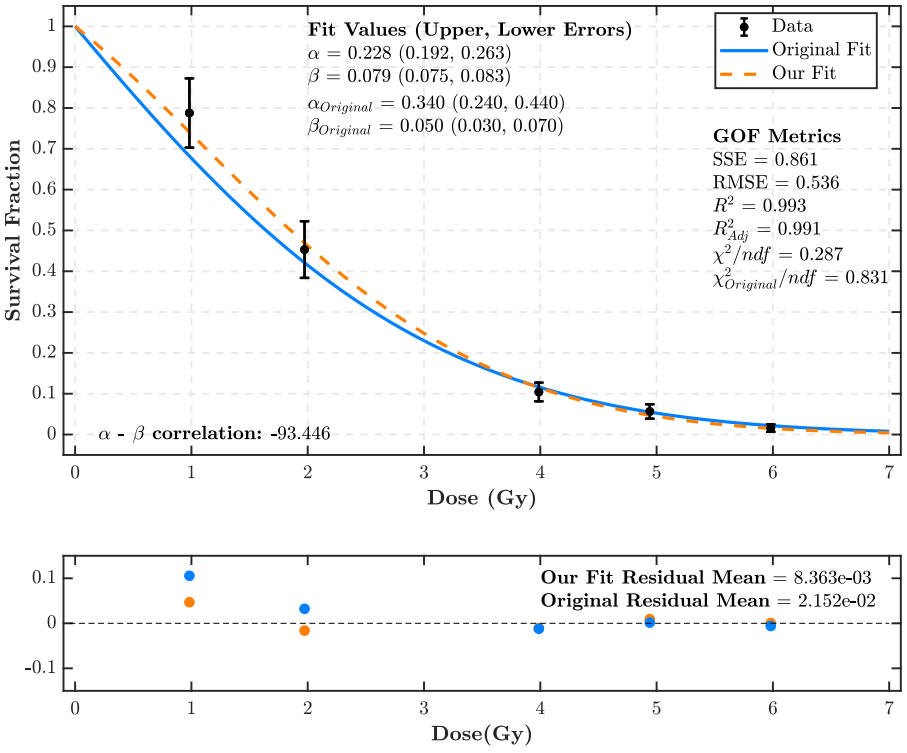
(b) 0.88MeV (30 keV/μm) protons incident on SQ20B cells from Belli et al. [60].

Figure H.4: Linearised LQ fits on experimental data. The second panel is a plot of the residuals between the observations and the linear LQ model.

H.2 Non-linear Regression Fits

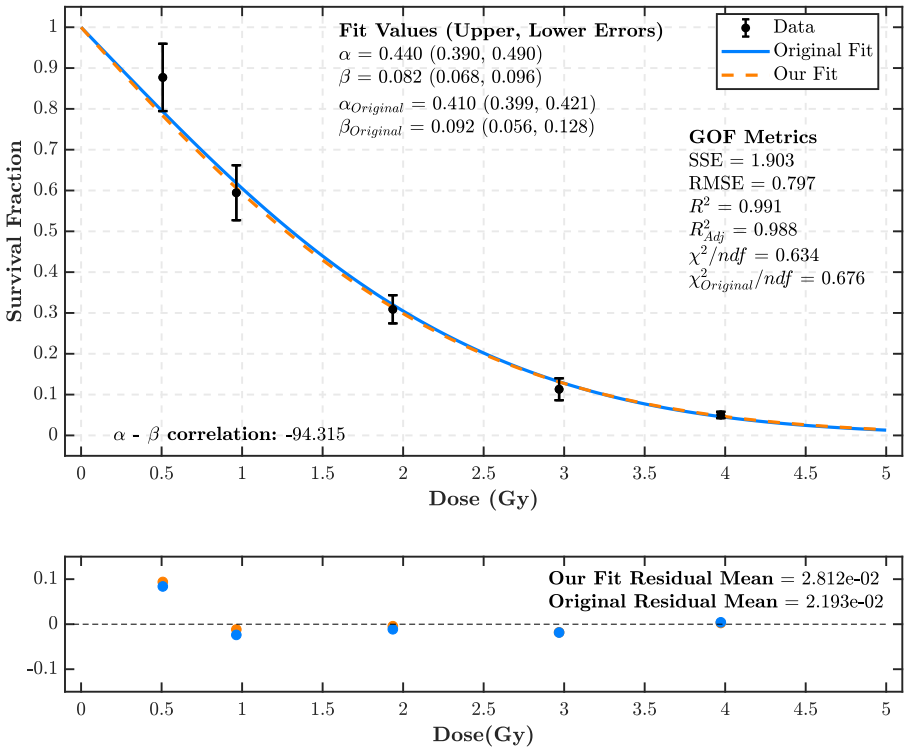


(a) 160MeV SOBP entrance region (1.1 keV/μm) V79 cells from Wouters et al. [38]

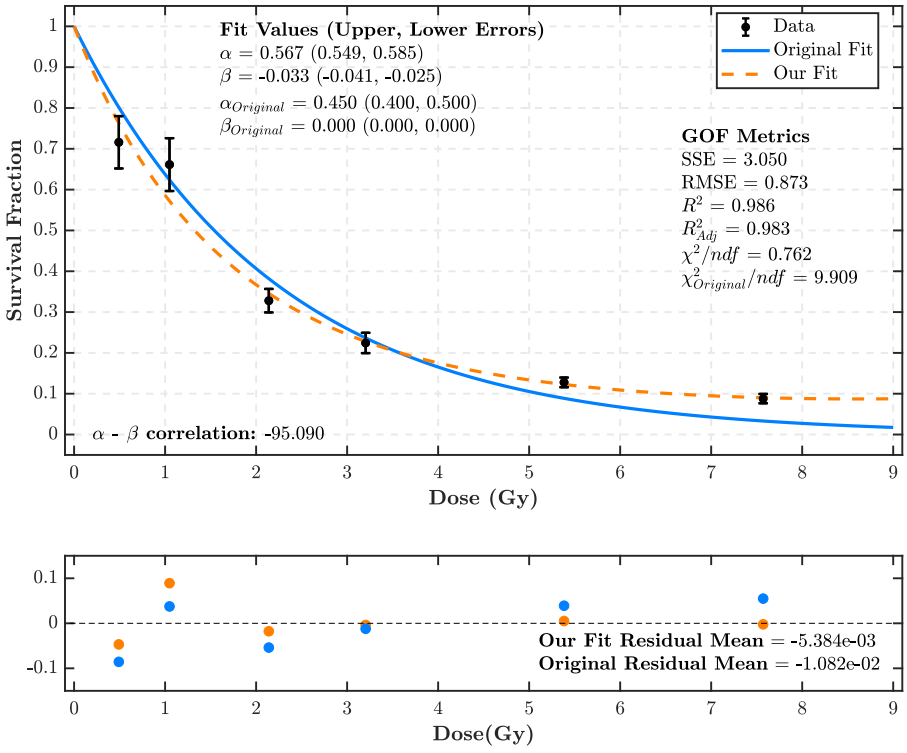


(b) 71MeV Pristine Bragg Peak (4.35 keV/μm) CHO cells from Howard et al. [123].

Figure H.5: Non-linear LQ fits to experimental data. The second panel is a plot of the residuals between the observations and the linear LQ model fit.

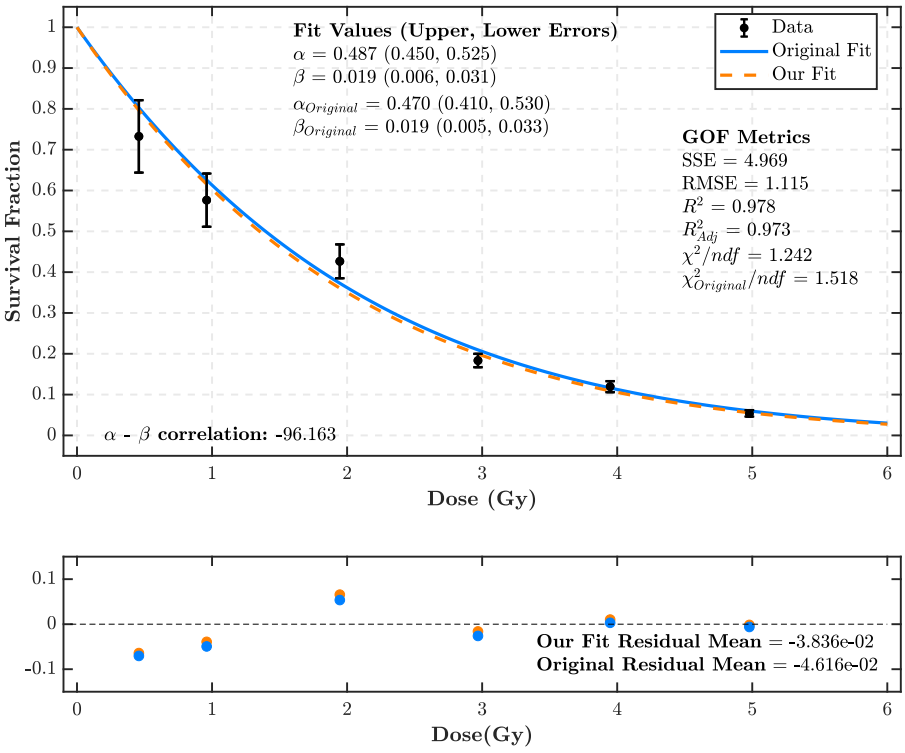


(a) 5.01MeV (7.7 keV/μm) protons incident on SCC25 cells from Belli et al. [60].

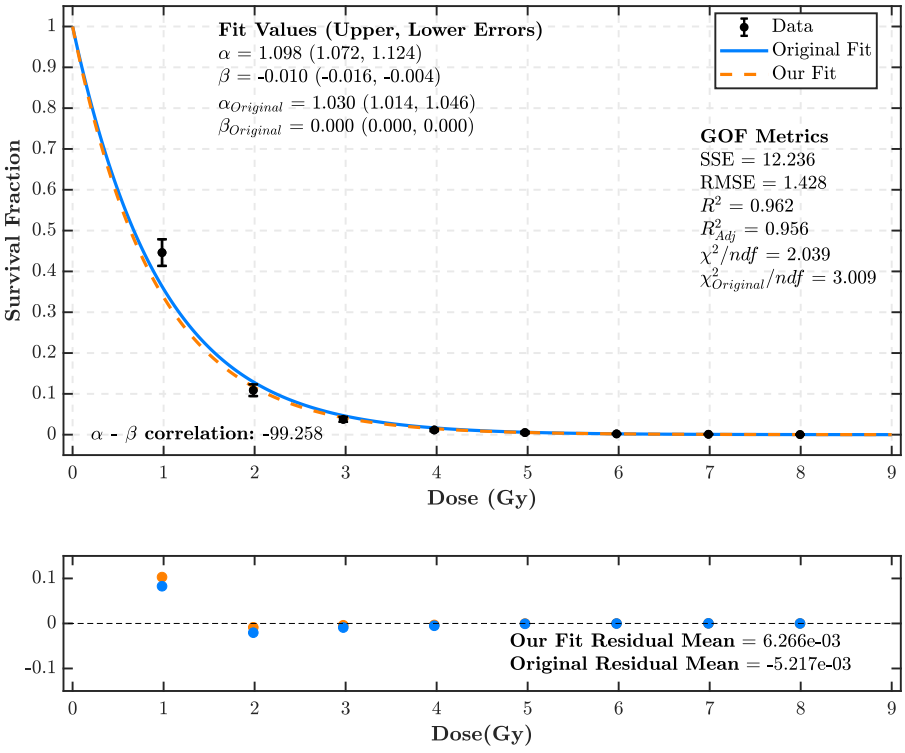


(b) 0.88MeV (30 keV/μm) protons incident on SQ20B cells from Belli et al. [60].

Figure H.6: Non-linear LQ fits to experimental data. The second panel is a plot of the residuals between the observations and the linear LQ model fit.



(a) 3.18MeV (11 keV/μm) protons incident on C3H10T12 cells from Bettega et al. (1998) [61].



(b) 0.76MeV (32 keV/μm) protons incident on V79 cells on data from Folkard et al. (1989) [96].

Figure H.7: Non-linear LQ fits to experimental data. The second panel is a plot of the residuals between the observations and the linear LQ model fit.

Appendix I

The Mittag-Leffler Function

The Mittag-Leffler function was first introduced in 1903 and in recent decades has found applications in the physical, chemical, biological and earth sciences. It can be written as a series representation

$$E_{\gamma,\delta}(z) = \sum_{k=0}^{\infty} \frac{z^k}{\Gamma(\gamma k + \delta)} \tag{I.1}$$

and in a special case where $\delta \to 1$

$$E_{\gamma,1}(z) = E_{\gamma}(z) = \sum_{k=0}^{\infty} \frac{z^k}{\Gamma(\gamma k + 1)}. \tag{I.2}$$

It is considered a fractional generalisation of the exponential function such that as $\gamma \to 1$ we recover $\exp(z)$ [143, 145]. It is a solution of fractional differential equations Eqs.(6.14) and (6.15).

If we take a random i.i.d. Mittag-Leffler distributed variable X, we can write the probability density function (PDF) as

$$f(t) = \lambda t^{\gamma-1} E_{\gamma,\gamma}(-\lambda t^{\gamma}) \quad , \tag{I.3}$$

and the cumulative density function of inter-arrival times (CDF),

$$F(t) = 1 - E_{\gamma}[-\lambda t^{\gamma}] \quad . \tag{I.4}$$

where for $\gamma = 1$ we recover the exponential distribution.

If the inter-arrival (waiting) times are i.i.d. as the Mittag-Leffler function, then the fractional Poisson count model probabilities are given by

$$P_0(t) = E_{\gamma,1}(-\lambda t^{\gamma}) \tag{I.5}$$

and

$$P_k(t) = \frac{(\lambda t^\gamma)^k}{k!} \sum_{j=0}^{\infty} \frac{(j+k)!}{j!} \frac{(-\lambda t^\gamma)^j}{\Gamma(\gamma(j+k)+1)} \text{ for } j = 0, 1, \cdots \tag{I.6}$$

where for $\gamma = 1$, $P_k(t)$ becomes the Poisson probability $P_k(t) = \frac{te^{-t}}{k!}$.

The mean and variance of the fPp are given by

$$E[X(t)] = \frac{\lambda t^\gamma}{\Gamma(\gamma+1)} \tag{I.7}$$

and

$$V[X(t)] = \frac{\lambda t^\gamma}{\Gamma(\gamma+1)} \left\{ 1 + \frac{\lambda t^\gamma}{\Gamma(\gamma+1)} \left[\frac{\gamma B(\gamma, 1/2)}{2^{2\gamma-1}} - 1 \right] \right\} \tag{I.8}$$

respectively, where $B(\gamma, 1/2) = \frac{\Gamma(\gamma)\Gamma(1/2)}{\Gamma(\gamma+1/2)}$.

I.1 Numerical Computation of Fractional Poisson Process Probabilities

I.1.1 Algorithm 1 (To Evaluate the Mainardi Function)

1. Compute for $\gamma \in (0, 1)$ the Mainardi function for an interval $[0, z_{max}]$ at discrete points $z_i = i\dot{\Delta}z$ with step size $\Delta z = 0.01$ and $i = 0, 1, \ldots z_{max}/\Delta Z$. For all $z_i \in [0, z_0)$ we use Eq.(6.20) and truncate the infinite series at $N \approx 140$. Up to $z_0 \in (1, z_{max})$ the truncated series will "converge" numerically. The number $z_0 > 1$ is defined by the fact that for this value the truncated series becomes numerically unstable. Since for all $z_i > 1$, $M_\gamma(z_i)$ declines, we can choose z_0 to be that value, where $M_\gamma(z_i) < M_\gamma(z_i + 1)$ due to numerical instability. From z_0 onward we make $M_\gamma(z_i) = 0$ for $z_i > z_0$. Each of these values is stored in an array which is named $M^{(1)}$.

2. Compute for $\gamma \in (0, 1)$ the Mainardi asymptotics in Eq.(6.21) for the same interval $[0, z_{max}]$ and store the results in an array $M^{(2)}$.

3. Blend both arrays $M^{(1)}$ and $M^{(2)}$ by defining $M_\gamma(z_i) = \max\{M^{(1)}, M^{(2)}\}$.

4. For fixed γ store $M_\gamma(z_i)$ for all further computations.

Note that due to the fact that for $\gamma = 1/2$ and $\gamma = 1/3$ exact values for the Mainardi function are known, we can use this information to judge the quality of our computations.

I.1.2 Algorithm 2 (Fractional Poisson Process Probabilities)

1. Fix $t > 0$, $\gamma \in (0, 1)$ and compute the Mainardi function using Algorithm 1.
2. Loop for $k = 0, 1, 2, \ldots, k_{max}$ and compute for each k integral Eq.(6.19) numerically, e.g. using the trapezoidal rule for a known integrand function.
3. Store fractional probabilities $p_k^{(\gamma)}(t)$ in array for further analysis.

I.1.3 Algorithm 3 (Fitting Fractional Poisson Process Probabilities)

1. Generate normalised histogram from the observed data of lethal damages with bin size 1.
2. Set for simplicity, $\Lambda = \lambda \cdot t^{\gamma}$.
3. Let $(\gamma_i, \Lambda_j) \in K$, then compute in this rectangular range K, using Algorithm 2, the fractional probabilities $p_k^{\gamma_i}(\Lambda_j)$.
4. Compute the mean-square difference of the normalised histogram and fractional probabilities. Try to find $(\gamma_i, \Lambda_j) \in K$, which minimises the mean-square differences.

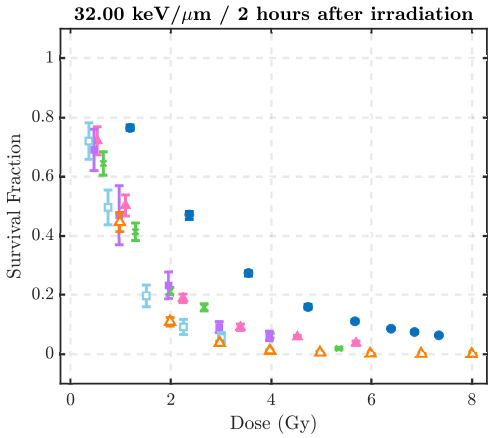
Appendix J

Simulated Dose Response Data Comparison with Experimental Data

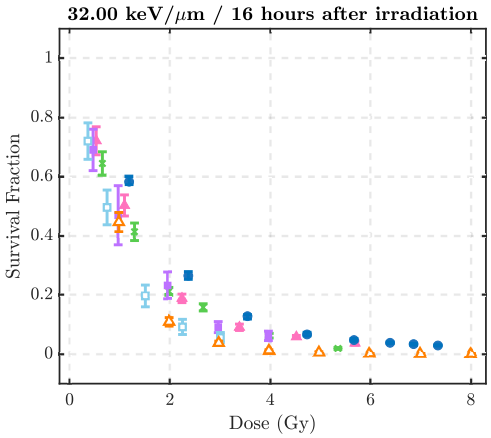
This Appendix contains additional comparisons of our simulated survival data for proton radiation (energy 1.6–50 MeV, corresponding to an LET range of 35 to 1.2keV/μm) incident on the V79 Chinese Hamster cell line. Each figure shows the survival at a different repair time. Basic information on the nature of the experimental data in the figures are provided in the legends :

- proton energy and LET,
- the cell line,
- radiation type (M = monoenergetic and S = modulated/SOBP)
- and cellular conditions (oxygen content and cell cycle phase distribution).

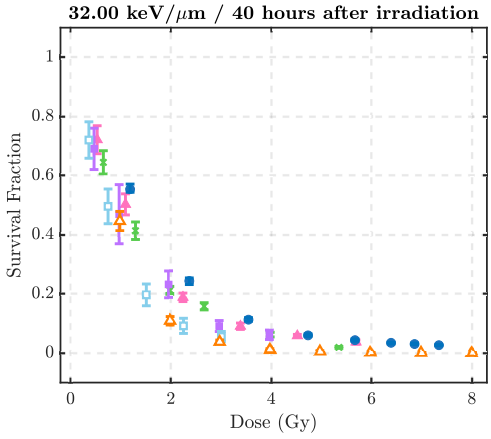
More information is provided on the experimental data in Appendix D.



(a) $t_{pr} = 2$ hours.



(b) $t_{pr} = 16$ hours.



(c) $t_{pr} = 40$ hours.

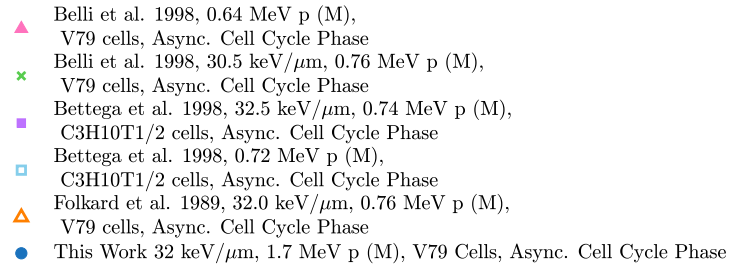
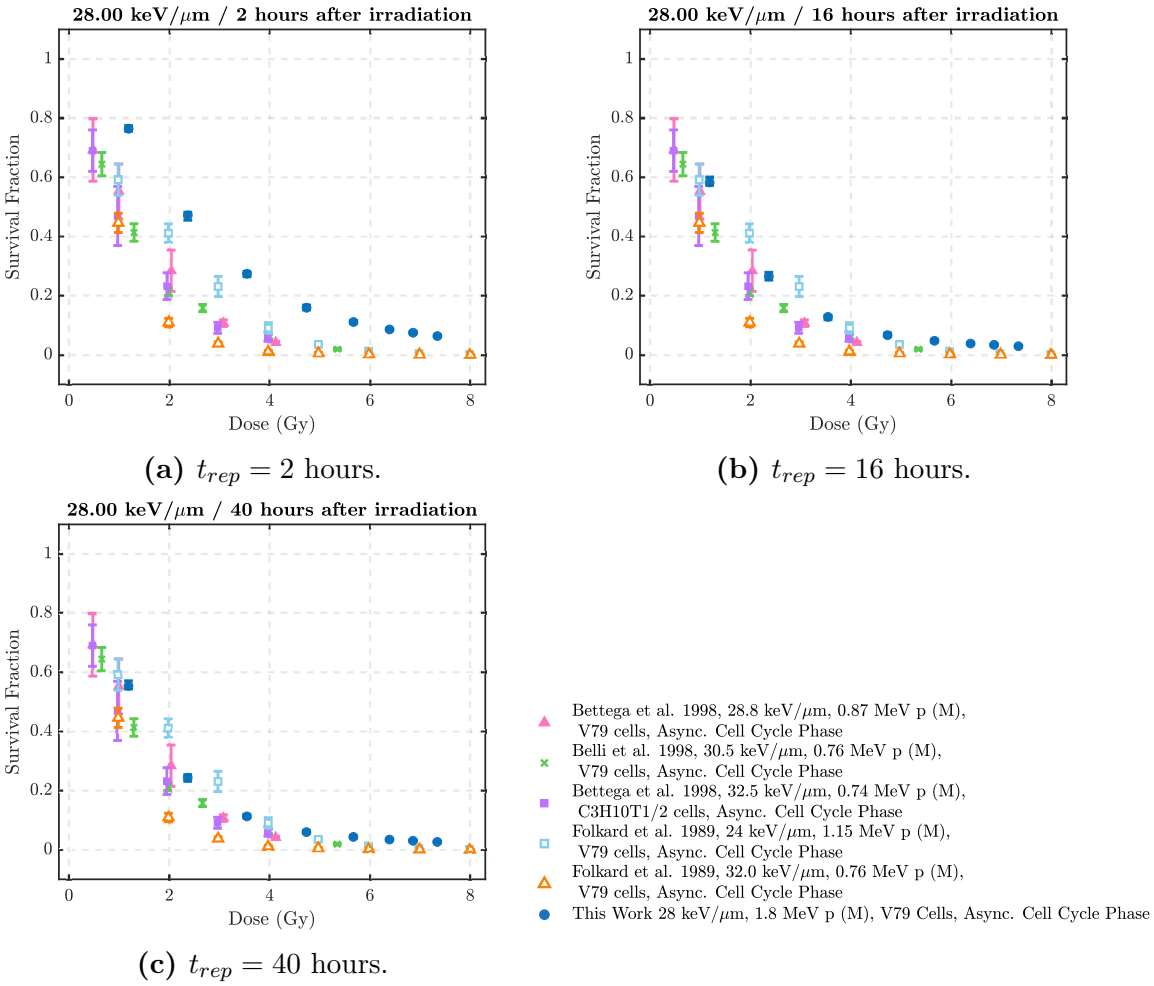


Figure J.1: Our simulated cell survival measurements for the V79 cell line at various post-irradiation times (t_{rep}) after exposure to a 1.7MeV (32.0keV/μm) proton beam. The experimental results presented utilise a monoenergetic proton beam (denoted by M) incident on cells with asynchronous cell cycle phases [26, 61, 96].

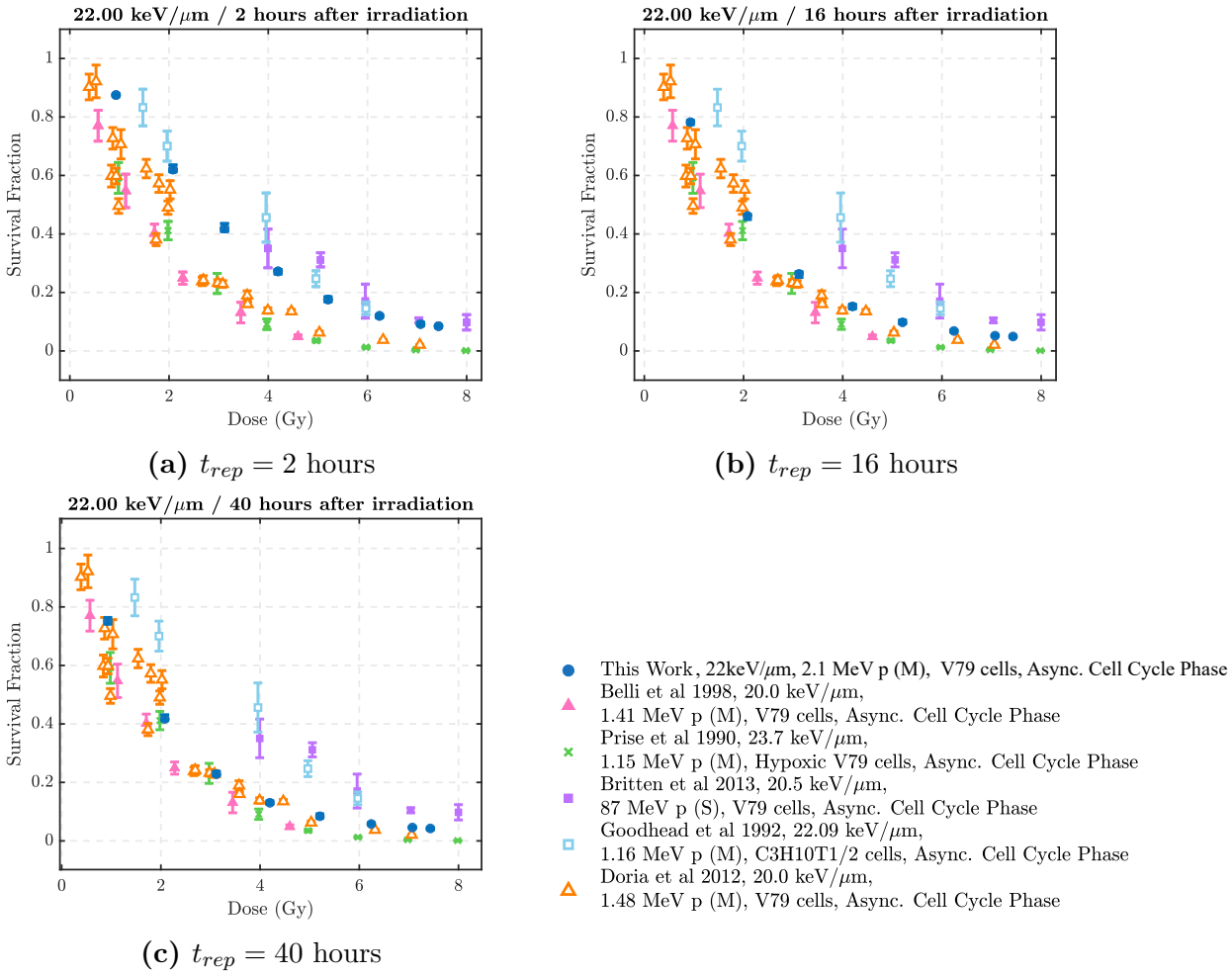


(a) $t_{rep} = 2$ hours.

(b) $t_{rep} = 16$ hours.

(c) $t_{rep} = 40$ hours.

Figure J.2: Our simulated cell survival measurements for the V79 cell line at various post-irradiation times (t_{rep}) after exposure to a 1.8MeV (28.0keV/μm) proton beam. The experimental results presented utilise a monoenergetic and modulated proton beam (denoted by M and S respectively) incident on cells with asynchronous cell cycle phases [26, 61, 96, 168].

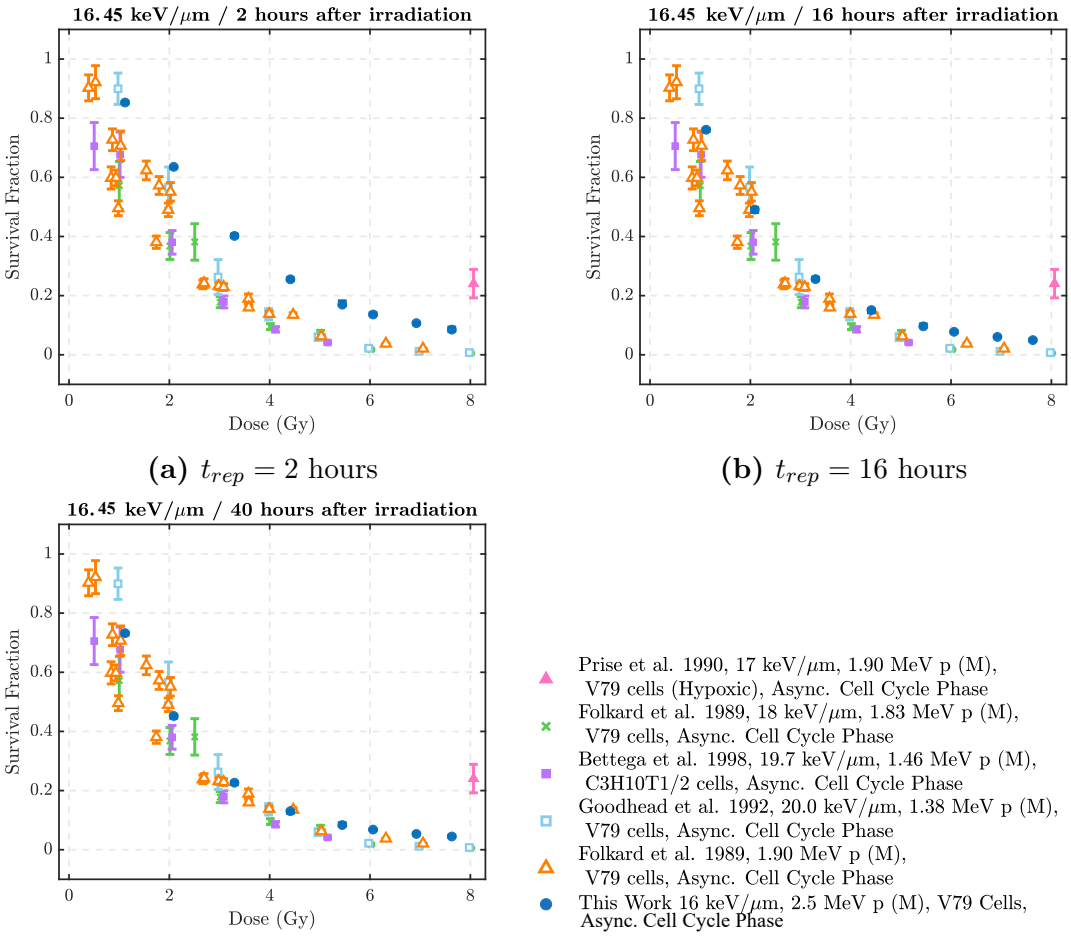


(a) $t_{rep} = 2$ hours

(b) $t_{rep} = 16$ hours

(c) $t_{rep} = 40$ hours

Figure J.3: Our simulated cell survival measurements for the V79 cell line at various post-irradiation times (t_{rep}) after exposure to a 2.1MeV (21.58keV/μm) proton beam. The experimental results presented utilise a monoenergetic and modulated proton beam (denoted by M and S respectively) incident on cells with asynchronous cell cycle phases [26, 29, 153, 168].

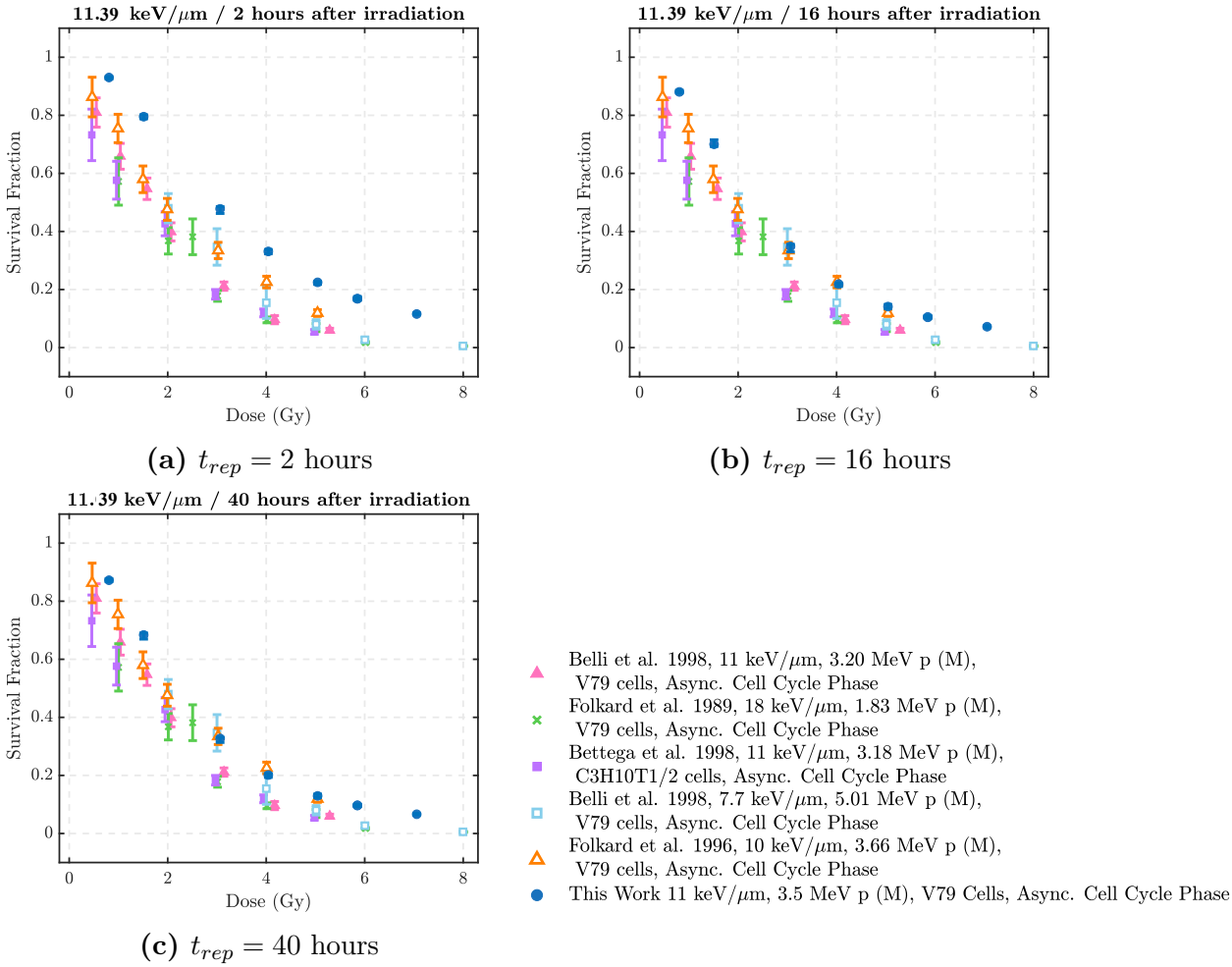


(a) $t_{rep} = 2$ hours

(b) $t_{rep} = 16$ hours

(c) $t_{rep} = 40$ hours

Figure J.4: Our simulated cell survival measurements for the V79 cell line at various post-irradiation times (t_{rep}) after exposure to a 2.5MeV (16.45keV/μm) proton beam. The experimental results presented utilise a monoenergetic and modulated proton beam (denoted by M and S respectively) incident on cells with asynchronous cell cycle phases [26, 29, 153, 168].

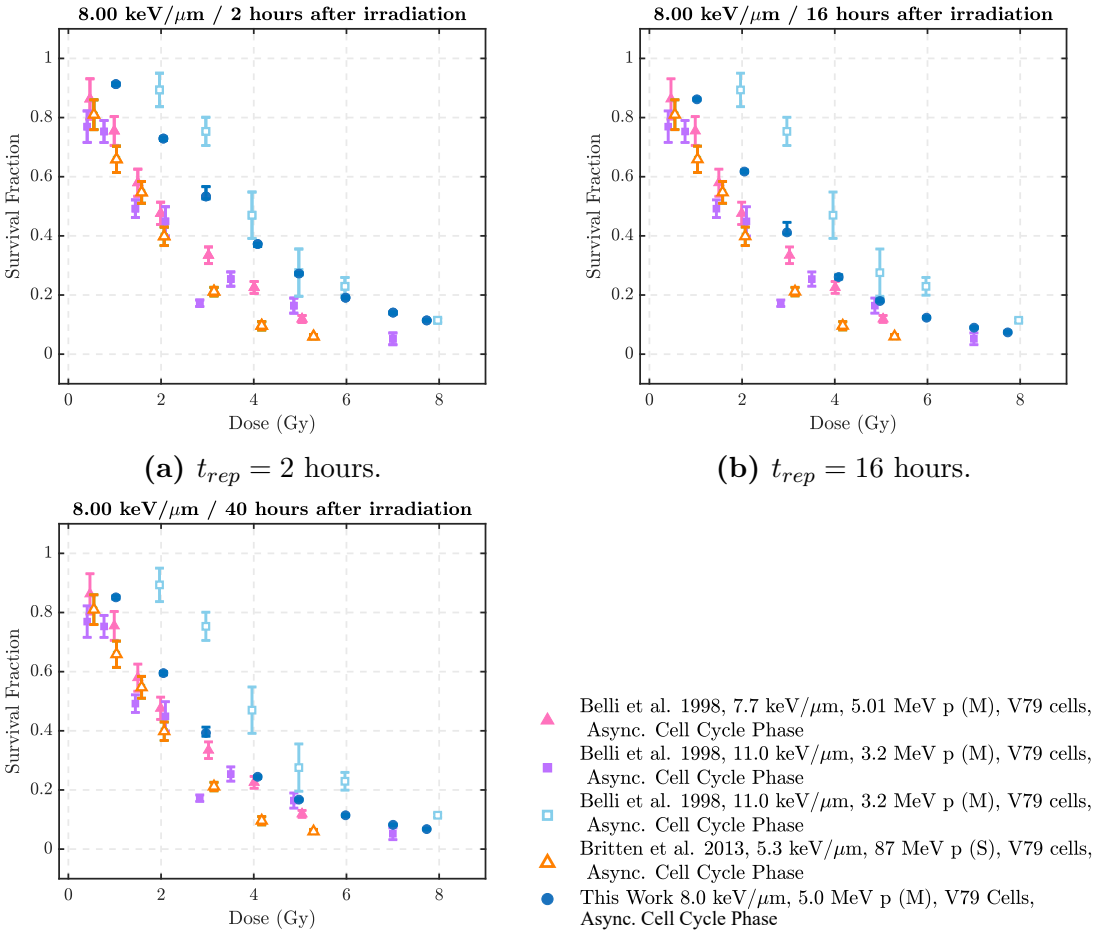


(a) $t_{rep} = 2$ hours

(b) $t_{rep} = 16$ hours

(c) $t_{rep} = 40$ hours

Figure J.5: Our simulated cell survival measurements for the V79 cell line at various post-irradiation times (t_{rep}) after exposure to a 3.5MeV (11.39keV/μm) proton beam. The experimental results presented utilise a monoenergetic proton beam (denoted by M) incident on cells with asynchronous cell cycle phases [26, 29, 153, 168].

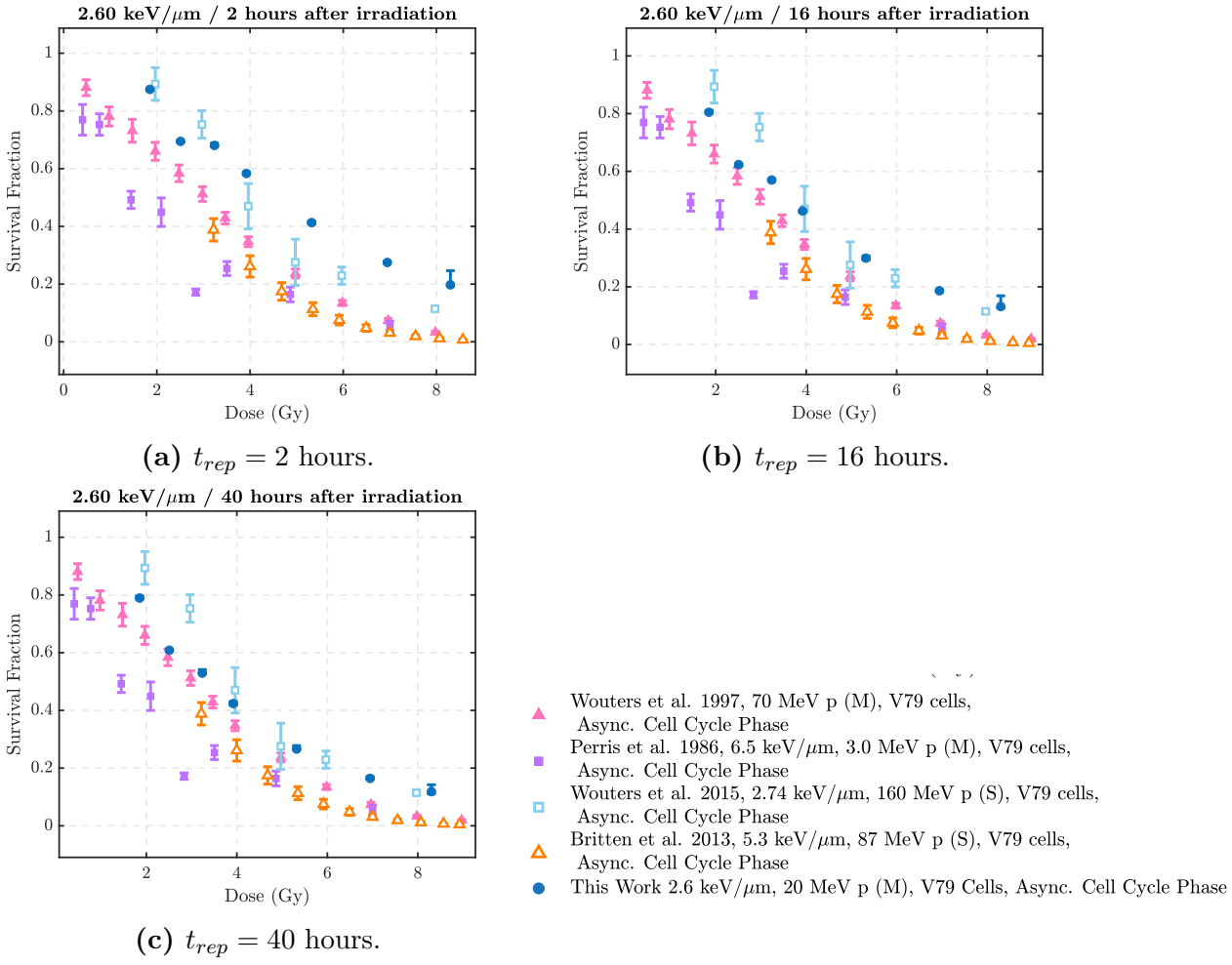


(a) $t_{rep} = 2$ hours.

(b) $t_{rep} = 16$ hours.

(c) $t_{rep} = 40$ hours.

Figure J.6: Our simulated cell survival measurements for the V79 cell line at various post-irradiation times (t_{rep}) after exposure to a 5MeV (8.0keV/μm) proton beam. The experimental results presented utilise a monoenergetic proton beam (denoted by M) incident on cells with asynchronous cell cycle phases [26, 61, 96].



(a) $t_{rep} = 2$ hours.

(b) $t_{rep} = 16$ hours.

(c) $t_{rep} = 40$ hours.

Figure J.7: Our simulated cell survival measurements for the V79 cell line at various post-irradiation times (t_{rep}) after exposure to a 20MeV (2.6keV/μm) proton beam. The experimental results presented utilise a monoenergetic proton beam (denoted by M) incident on cells with asynchronous cell cycle phases [26, 61, 96].

Appendix K

FPp Fits on Experimental Dose-Response Data

This Appendix summarises the testing of the fractional survival probability (Mittag-Leffler) function on experimental world data outlined in Appendix D. For each set of conditions, the fPp model parameters are presented with their GOF metrics and that of the Poisson process (LQ model) are presented in Appendix F for comparison. The specified conditions are the reported proton energy (MeV), LET (keV/μm) and cell line. The errors on the fit parameters are the 95% confidence interval. Additionally, we report the RMSE, R^2, χ^2 and RBE (at a 10% and 37% survival fraction).

- Kolmogorov-Smirnov (KS) Test: ✓= Residual errors distributed normally, ✗= Residual errors are not distributed normally.

- Anderson-Darling (AD) Test: ✓= Residual errors distributed normally, ✗= Residual errors are not distributed normally.

- Breusch-Pagan (BP) Test:
  ✓= residual errors are homoscedastic (have constant variance),
  ✗= residual errors are heteroscedastic (do not have constant variance).

- White Test: ✓= residual errors are homoscedastic (have constant variance), ✗= residual errors are heteroscedastic (do not have constant variance).

K.1 Regression Fit Results

Table K.1: The fit parameters and goodness-of-fit metrics for the fPp (Mittag-Leffler) model ($E_\gamma[-(aD+bD^2)^\gamma]$) fits to selected experimental data using the non-linear least squares (NLLS) and search algorithm methods (Sch.). Cell lines of human origin are denoted by a dagger (†). The *et al.* is omitted from the table to save page space, however all publications in this table have multiple authors. **Note:** The corresponding Pp fit values are presented in Appendix F.

Study	LET (keV/μm)	Energy (MeV)	Cell Line	γ_{LSQ}	γ_{Sch}	γ 95% Conf. Interval	a_{LSQ}	a_{Sch}	a 95% Conf. Interval	b_{LSQ}	b_{Sch}	β 95% Conf. Interval	χ^2	R^2	RMSE	RBE_{10}	RBE_{37}
Belli [60]	9.11	5.04	M10†	0.909	0.909	(0.516, 1.000)	0.435	0.435	(0.000, 0.982)	0.052	0.052	(0.000, 0.392)	0.350	0.997	0.592	0.935	1.050
Belli [60]	33	0.88	M10†	0.909	0.909	(0.516, 1.000)	0.435	0.435	(0.000, 0.982)	0.052	0.052	(0.000, 0.393)	0.350	0.997	0.592	0.935	1.050
Belli [60]	7.7	5.04	SCC25†	0.901	0.901	(0.696, 1.000)	0.177	0.177	(0.000, 0.864)	0.251	0.251	(0.000, 0.706)	0.448	0.996	0.669	1.048	0.950
Belli [60]	19.7	1.49	SCC25†	0.896	0.896	(0.871, 0.922)	0.642	0.642	(0.560, 0.725)	0.227	0.227	(0.146, 0.308)	0.002	0.999	0.042	1.303	1.420
Belli [60]	29.5	0.88	SCC25†	0.931	0.931	(0.476, 1.000)	0.788	0.788	(0.000, 1.968)	0.082	0.082	(0.000, 1.088)	0.345	0.995	0.587	1.207	1.410
Belli [60]	7.7	5.04	SQ20B†	0.832	0.832	(0.000, 1.000)	0.078	0.078	(0.000, 0.528)	0.030	0.030	(0.000, 0.148)	0.186	0.994	0.431	0.817	1.000
Belli [60]	19.8	1.49	SQ20B†	0.698	0.698	(0.401, 0.994)	0.015	0.015	(0.000, 0.276)	0.072	0.072	(0.000, 0.173)	0.677	0.987	0.987	0.835	1.230
Belli [60]	30	0.88	SQ20B†	0.593	0.593	(0.358, 0.828)	0.061	0.061	(0.000, 0.792)	0.313	0.313	(0.000, 0.828)	0.993	0.986	0.997	1.236	2.463
Belli [60]	7.7	5.04	HF19†	0.970	0.997	(0.547, 1.000)	0.51	0.525	(0.000, 1.191)	0.012	0.01	(0.000, 0.299)	0.728	0.974	2.384	0.883	0.890
Belli [60]	19.5	1.49	HF19†	0.726	0.724	(0.652, 0.801)	0.000	0.000	(0.000, 0.000)	0.335	0.335	(0.231, 0.439)	0.816	0.990	0.782	0.837	0.809
Belli [60]	29	0.88	HF19†	0.926	0.953	(0.845, 1.000)	0.649	0.671	(0.453, 0.845)	0.000	0.000	(0.000, 0.000)	2.567	0.962	1.398	0.843	1.050
Belli [26]	7.7	5.01	V79	0.958	0.980	(0.729, 1.188)	0.271	0.291	(0.082, 0.461)	0.035	0.029	(0.000, 0.116)	0.484	0.994	0.765	1.089	1.353
Belli [26]	11	3.2	V79	0.997	0.990	(0.990, 0.999)	0.361	0.313	(0.351, 0.370)	0.041	0.049	(0.039, 0.042)	0.295	0.997	0.670	1.360	1.660
Belli [26]	20	1.41	V79	0.909	0.909	(0.779, 1.039)	0.316	0.316	(0.000, 0.670)	0.145	0.145	(0.000, 0.358)	0.315	0.998	0.561	1.661	2.067
Belli [26]	30.5	0.76	V79	0.972	0.992	(0.691, 1.253)	0.691	0.665	(0.000, 1.404)	0.031	0.021	(0.000, 0.462)	0.078	0.990	1.569	1.930	2.400
Belli [26]	36.4	0.64	V79	0.755	0.755	(0.669, 0.841)	0.112	0.112	(0.000, 0.493)	0.474	0.474	(0.138, 0.810)	0.305	0.998	0.552	1.800	2.590
Belli [26]	37.8	0.57	V79	0.751	0.751	(0.714, 0.787)	0.000	0.000	(0.000, 0.000)	0.451	0.451	(0.372, 0.530)	0.678	0.996	0.821	1.766	2.229
Bettega [61]	11	3.18	C3H10T12†	0.969	0.969	(0.221, 1.000)	0.474	0.474	(0.000, 1.499)	0.034	0.034	(0.000, 0.585)	0.086	0.979	1.258	1.451	2.100
Bettega [61]	19.7	1.46	C3H10T12†	0.967	0.967	(0.776, 1.000)	0.375	0.375	(0.000, 0.905)	0.066	0.066	(0.000, 0.285)	0.758	0.991	0.871	1.900	1.600
Bettega [61]	28.8	0.87	C3H10T12†	0.974	0.974	(0.791, 1.000)	0.521	0.521	(0.000, 1.109)	0.082	0.082	(0.000, 0.382)	0.288	0.995	0.536	2.380	2.060
Bettega [61]	31.6	0.77	C3H10T12†	0.955	0.955	(0.000, 1.000)	0.709	0.709	(0.000, 3.579)	0.004	0.004	(0.000, 2.236)	0.765	0.988	0.875	2.106	2.250
Bettega [61]	32.5	0.74	C3H10T12†	0.837	0.837	(0.350, 1.000)	0.488	0.487	(0.000, 1.851)	0.266	0.266	(0.000, 1.562)	0.287	0.994	0.536	2.519	2.490
Bettega [61]	33.2	0.72	C3H10T12†	0.805	0.805	(0.684, 0.926)	0.360	0.360	(0.000, 0.876)	0.740	0.740	(0.047, 1.434)	0.040	0.999	0.199	3.416	3.227
Guan [27]	10.8	79.7	H460†	0.956	0.956	(0.896, 1.016)	0.179	0.179	(0.000, 0.460)	0.243	0.243	(0.084, 0.403)	1.193	0.999	1.092	1.325	1.420
Guan [27]	15.2	79.7	H460†	0.963	0.963	(0.746, 1.180)	0.272	0.272	(0.000, 1.490)	0.498	0.498	(0.000, 1.547)	1.309	0.999	1.144	1.927	2.050
Guan [27]	17.7	79.7	H460†	0.965	0.995	(0.889, 1.040)	0.343	0.575	(0.000, 0.950)	0.972	0.644	(0.230, 1.713)	0.035	0.998	1.685	2.549	2.760
Guan [27]	19	79.7	H460†	0.983	0.983	(0.955, 1.011)	0.759	0.758	(0.000, 1.586)	1.179	1.180	(0.324, 2.034)	4.798	0.993	2.190	3.370	3.540

Study	LET (keV/μm)	Energy (MeV)	Cell Line	γ_{LSQ}	γ_{Sch}	γ 95% Conf. Interval	a_{LSQ}	a_{Sch}	a 95% Conf. Interval	b_{LSQ}	b_{Sch}	b 95% Conf. Interval	χ^2	R^2	RMSE	RBE_{10}	RBE_{37}
Howard [123]	1.78	71	A549†	0.995	0.995	(0.978, 1.013)	0.287	0.287	(0.244, 0.331)	0.034	0.034	(0.030, 0.037)	0.019	0.999	0.139	1.391	1.471
Howard [123]	3.36	71	A549†	0.987	0.987	(0.968, 1.006)	0.211	0.211	(0.000, 0.441)	0.061	0.061	(0.005, 0.116)	0.337	0.995	0.581	1.334	1.410
Howard [123]	4.35	71	A549†	0.969	0.988	(0.739, 1.199)	0.351	0.420	(0.000, 0.922)	0.071	0.046	(0.000, 0.326)	0.492	0.991	0.795	1.624	2.040
Howard [123]	7.29	71	A549†	0.967	0.967	(0.776, 1.159)	0.395	0.395	(0.000, 0.991)	0.071	0.071	(0.000, 0.333)	0.653	0.985	0.808	1.677	2.070
Howard [123]	0.99	160	A549†	0.980	0.980	(0.947, 1.000)	0.214	0.214	(0.154, 0.275)	0.036	0.036	(0.019, 0.054)	0.019	0.999	0.005	1.107	1.280
Howard [123]	2.26	160	A549†	0.984	0.984	(0.982, 0.987)	0.161	0.161	(0.147, 0.174)	0.037	0.037	(0.035, 0.040)	< 0.001	0.999	< 0.001	1.034	1.170
Howard [123]	4.19	160	A549†	0.965	0.965	(0.929, 1.001)	0.074	0.074	(0.000, 0.180)	0.067	0.067	(0.038, 0.096)	0.029	0.998	0.013	1.129	1.160
Howard [123]	7.29	160	A549†	0.978	0.978	(0.915, 1.040)	0.233	0.233	(0.133, 0.333)	0.055	0.055	(0.015, 0.095)	0.037	0.999	0.010	1.298	1.480
Howard [123]	1.78	71	CHO	0.992	0.992	(0.986, 0.998)	0.187	0.187	(0.137, 0.237)	0.043	0.043	(0.03, 0.055)	0.152	0.999	0.389	1.079	1.190
Howard [123]	3.36	71	CHO	0.988	0.988	(0.968, 1.008)	0.180	0.180	(0.089, 0.272)	0.070	0.070	(0.035, 0.105)	0.082	0.999	0.286	1.312	1.370
Howard [123]	4.35	71	CHO	0.983	0.995	(0.859, 1.108)	0.189	0.208	(0.000, 0.546)	0.096	0.082	(0.000, 0.231)	0.174	0.995	0.57	1.604	1.450
Howard [123]	7.34	71	CHO	0.991	0.991	(0.979, 1.002)	0.519	0.519	(0.420, 0.618)	0.042	0.042	(0.005, 0.079)	0.084	0.999	0.29	1.708	2.050
Howard [123]	0.99	160	CHO	0.988	0.989	(0.974, 1.003)	0.164	0.169	(0.071, 0.256)	0.050	0.049	(0.027, 0.072)	0.743	0.998	0.415	1.116	1.190
Howard [123]	2.26	160	CHO	0.98	0.995	(0.918, 1.042)	0.167	0.107	(0.000, 0.405)	0.054	0.050	(0.000, 0.111)	1.816	0.992	0.897	1.123	1.001
Howard [123]	4.19	160	CHO	0.996	1.000	(0.993, 0.999)	0.241	0.261	(0.205, 0.277)	0.060	0.052	(0.053, 0.068)	0.319	0.998	0.355	1.306	1.280
Howard [123]	7.29	160	CHO	0.941	0.941	(0.811, 1.071)	0.121	0.121	(0.000, 0.547)	0.148	0.148	(0.000, 0.346)	0.270	0.995	0.519	1.550	1.661
Wouters [38]	1.10	160	V79	0.995	1.000	(0.993, 0.996)	0.115	0.123	(0.102, 0.128)	0.052	0.047	(0.050, 0.055)	< 0.001	0.981	0.675	1.004	1.250
Wouters [38]	2.06	160	V79	0.993	1.000	(0.990, 0.996)	0.134	0.112	(0.106, 0.161)	0.065	0.054	(0.058, 0.072)	0.001	0.959	0.911	1.042	1.280
Wouters [38]	2.41	160	V79	0.993	1.000	(0.991, 0.996)	0.041	0.123	(0.017, 0.064)	0.073	0.053	(0.068, 0.078)	< 0.001	0.977	0.772	1.055	1.310
Wouters [38]	3.20	160	V79	0.993	1.000	(0.989, 0.996)	0.114	0.130	(0.078, 0.150)	0.077	0.061	(0.069, 0.086)	0.036	0.966	0.886	1.124	1.390
Wouters [38]	4.74	160	V79	0.993	1.000	(0.992, 0.995)	0.083	0.141	(0.070, 0.096)	0.069	0.054	(0.065, 0.072)	< 0.001	0.982	0.608	1.086	1.360
Wouters [38]	1.03	230	V79	0.995	1.000	(0.993, 0.997)	0.099	0.103	(0.087, 0.112)	0.055	0.05	(0.052, 0.057)	< 0.001	0.981	0.911	0.998	1.080
Wouters [38]	1.95	230	V79	0.993	1.000	(0.990, 0.996)	0.046	0.101	(0.004, 0.087)	0.072	0.054	(0.062, 0.082)	< 0.001	0.968	0.836	1.025	1.100
Wouters [38]	2.28	230	V79	0.992	1.000	(0.988, 0.996)	0.031	0.103	(0.000, 0.078)	0.078	0.057	(0.067, 0.089)	< 0.001	0.967	0.927	1.051	1.130
Wouters [38]	2.95	230	V79	0.996	1.000	(0.995, 0.997)	0.121	0.123	(0.116, 0.126)	0.067	0.062	(0.065, 0.068)	< 0.001	0.977	0.737	1.121	1.220
Wouters [38]	4.02	230	V79	0.995	1.000	(0.993, 0.997)	0.171	0.122	(0.155, 0.186)	0.057	0.057	(0.054, 0.060)	< 0.001	0.963	0.936	1.078	1.170

K.2 Hypothesis Testing

Table K.3: Hypothesis and "regression assumption" test results for the fPp model ($E_\gamma[-(aD + bD^2)^\gamma]$) on selected experimental data. **Note:** The BP test cannot be applied to data with non-Normally distributed errors (i.e. they fail the KS or AD test) and is therefore denoted with a "-".

Study	LET (keV/μm)	Energy (MeV)	Cell Line	KS Test	AD Test	BP Test	White Test
Belli et al. [60]	9.11	5.04	M10†	✓	✓	✓	✓
Belli et al. [60]	33	0.88	M10†	✓	✓	✓	✓
Belli et al. [60]	7.7	5.04	SCC25†	✓	✓	✓	✓
Belli et al. [60]	19.7	1.49	SCC25†	✓	✓	✓	✓
Belli et al. [60]	29.5	1.35	SCC25†	✗	✗	-	✓
Belli et al. [60]	7.7	5.04	SQ20B†	✓	✓	✓	✓
Belli et al. [60]	19.8	1.49	SQ20B†	✓	✓	✓	✓
Belli et al. [60]	30	0.88	SQ20B†	✗	✗	-	✓
Belli et al. [26]	7.7	5.01	V79	✓	✓	✓	✓
Belli et al. [26]	11	3.2	V79	✓	✓	✓	✓
Belli et al. [26]	20	1.41	V79	✓	✓	✓	✓
Belli et al. [26]	30.5	0.76	V79	✓	✓	✓	✓
Belli et al. [26]	34.6	0.64	V79	✓	✓	✓	✓
Belli et al. [26]	37.8	0.57	V79	✓	✓	✓	✓
Bettega et al. [61]	11	3.18	C3H10T12†	✓	✗	-	✓
Bettega et al. [61]	19.7	1.46	C3H10T12†	✗	✓	-	✓
Bettega et al. [61]	28.8	0.87	C3H10T12†	✗	✓	-	✓
Bettega et al. [61]	31.6	0.77	C3H10T12†	✓	✓	✓	✓
Bettega et al. [61]	32.5	0.74	C3H10T12†	✓	✓	✓	✓
Bettega et al. [61]	33.2	0.72	C3H10T12†	✓	✓	✓	✓
Guan et al. [27]	10.8	79.9	H460†	✓	✓	✓	✓
Guan et al. [27]	15.2	79.9	H460†	✓	✓	✓	✓
Guan et al. [27]	17.7	79.9	H460†	✓	✓	✓	✓
Guan et al. [27]	19	79.9	H460†	✓	✓	✓	✓
Howard et al. [123]	1.78	71	A549†	✓	✓	✓	✓
Howard et al. [123]	3.36	71	A549†	✓	✓	✗	✓
Howard et al. [123]	4.35	71	A549†	✓	✓	✓	✓
Howard et al. [123]	7.29	71	A549†	✓	✓	✗	✓
Howard et al. [123]	0.99	160	A549†	✓	✓	✓	✓
Howard et al. [123]	2.26	160	A549†	✗	✗	-	✓
Howard et al. [123]	4.19	160	A549†	✓	✓	✓	✓
Howard et al. [123]	7.29	160	A549†	✗	✗	-	✓
Howard et al. [123]	0.99	160	CHO†	✓	✓	✗	✓
Howard et al. [123]	2.26	160	CHO†	✗	✗	-	✓
Howard et al. [123]	4.19	160	CHO†	✓	✓	✓	✓
Howard et al. [123]	7.29	160	CHO†	✓	✓	✗	✓
Wouters et al. [38]	1.1	230	V79	✓	✓	✓	✓
Wouters et al. [38]	2.06	230	V79	✗	✗	-	✗
Wouters et al. [38]	2.41	230	V79	✗	✗	-	✗
Wouters et al. [38]	3.2	230	V79	✗	✗	-	✗
Wouters et al. [38]	4.74	230	V79	✗	✗	-	✗
Wouters et al. [38]	1.03	160	V79	✗	✗	-	✗
Wouters et al. [38]	1.95	160	V79	✗	✗	-	✗
Wouters et al. [38]	2.28	160	V79	✓	✗	-	✗
Wouters et al. [38]	2.95	160	V79	✗	✗	-	✗
Wouters et al. [38]	4.01	160	V79	✗	✗	-	✗

Appendix L

FPp Fits on Simulated Dose-Response Data

This Appendix summarises the testing of both the fractional and standard poisson survival functions on simulated data generated using the simulation framework discussed in Chapters 4 and 7. The specified conditions are the mean Gaussian energy of the beam (MeV), LET (keV/μm) and repair time (hours). The tables presented summarise the fit results using :

- FPp fits using Method 3
- Pp fits using Method 3
- FPp using Method 1
- Hypothesis testing results on the fPp and Pp fits using Method 1 after repair times of 2 and 40 hours

in order.

The hypothesis testing results are defined as follows :

- Kolmogorov-Smirnov (KS) Test: ✓= Residual errors distributed normally, ✗= Residual errors are not distributed normally.
- Anderson-Darling (AD) Test: ✓= Residual errors distributed normally, ✗= Residual errors are not distributed normally.
- Breusch-Pagan (BP) Test:
  ✓= residual errors are homoscedastic (have constant variance),

✗= residual errors are heteroscedastic (do not have constant variance).

- White Test: ✓= residual errors are homoscedastic (have constant variance), ✗= residual errors are heteroscedastic (do not have constant variance).

The final section contains additional plots to aid discussion in Chapter 8.

L.1 Energy-varied LET

Table L.1: The fit parameters and goodness-of-fit metrics for the fPp (Mittag-Leffler) model fits ($E_{\gamma,1}[-(aD+bD^2)^\gamma]$) to our simulated data using Method 3 in Section (8.3.2). **Note:** Where $a = 0$, the true parameter is not exactly zero but is < 0.001.

Energy (MeV)	LET (keV/μm)	Repair Time (hrs)	γ	γ 95% Conf. Interval	a	a 95% Conf. Interval	b	b 95% Conf. Interval	χ^2	SSE	RMSE	R^2	RBE_{10}	RBE_{37}
50	1.2	2	0.851	(0.830, 0.872)	0.000	(0.000, 0.009)	0.027	(0.025, 0.029)	0.087	0.003	0.009	0.998	0.921	1.018
20	2.6	2	0.806	(0.776, 0.836)	0.000	(0.000, 0.000)	0.034	(0.033, 0.036)	0.781	0.027	0.028	0.985	0.948	1.128
5	8	2	0.832	(0.825, 0.839)	0.000	(0.000, 0.000)	0.061	(0.060, 0.061)	0.232	0.033	0.015	0.997	1.333	1.504
3.5	11	2	0.829	(0.825, 0.834)	0.000	(0.000, 0.000)	0.073	(0.072, 0.074)	0.133	0.012	0.009	0.999	1.450	1.657
2.5	16	2	0.813	(0.808, 0.818)	0.000	(0.000, 0.000)	0.092	(0.091, 0.093)	0.247	0.016	0.011	0.998	1.572	1.831
2.1	22	2	0.804	(0.797, 0.812)	0.000	(0.000, 0.008)	0.099	(0.096, 0.102)	0.122	0.008	0.007	0.999	1.602	1.915
2	24	2	0.804	(0.801, 0.807)	0.000	(0.000, 0.000)	0.103	(0.102, 0.104)	0.142	0.018	0.009	0.999	1.637	1.960
1.8	29	2	0.794	(0.784, 0.804)	0.000	(0.000, 0.000)	0.109	(0.105, 0.114)	0.255	0.020	0.010	0.998	1.652	2.004
1.7	32	2	0.796	(0.787, 0.805)	0.004	(0.000, 0.014)	0.117	(0.112, 0.122)	0.173	0.017	0.009	0.990	1.712	2.071
1.6	35	2	0.795	(0.787, 0.802)	0.003	(0.000, 0.013)	0.127	(0.122, 0.132)	0.159	0.014	0.008	0.999	1.778	2.146
50	1.2	16	0.842	(0.831, 0.853)	0.000	(0.000, 0.000)	0.041	(0.040, 0.042)	0.092	0.003	0.009	0.999	1.119	1.242
20	2.6	16	0.799	(0.776, 0.822)	0.000	(0.000, 0.000)	0.051	(0.049, 0.053)	0.595	0.017	0.022	0.991	1.143	1.374
5	8	16	0.815	(0.808, 0.822)	0.000	(0.000, 0.000)	0.097	(0.095, 0.098)	0.299	0.033	0.015	0.997	1.623	1.881
3.5	11	16	0.815	(0.811, 0.819)	0.000	(0.000, 0.000)	0.119	(0.118, 0.120)	0.163	0.009	0.008	0.999	1.793	2.097
2.5	16	16	0.803	(0.791, 0.815)	0.000	(0.000, 0.019)	0.161	(0.151, 0.171)	0.263	0.013	0.010	0.998	2.028	2.401
2.1	22	16	0.799	(0.791, 0.807)	0.031	(0.018, 0.043)	0.172	(0.164, 0.179)	0.137	0.006	0.007	0.999	2.118	2.617
2	24	16	0.794	(0.792, 0.797)	0.000	(0.000, 0.000)	0.197	(0.196, 0.199)	0.117	0.012	0.007	0.999	2.223	2.686
1.8	29	16	0.800	(0.792, 0.808)	0.054	(0.038, 0.069)	0.210	(0.200, 0.220)	0.191	0.010	0.007	0.999	2.358	2.895
1.7	32	16	0.813	(0.803, 0.823)	0.086	(0.066, 0.106)	0.215	(0.201, 0.229)	0.123	0.009	0.007	0.999	2.500	3.027
1.6	35	16	0.812	(0.800, 0.825)	0.103	(0.071, 0.135)	0.241	(0.219, 0.263)	0.181	0.012	0.007	0.998	2.683	3.222
50	1.2	40	0.841	(0.831, 0.851)	0.000	(0.000, 0.000)	0.044	(0.043, 0.045)	0.094	0.002	0.008	0.999	1.154	1.297
20	2.6	40	0.801	(0.739, 0.863)	0.000	(0.000, 0.049)	0.055	(0.040, 0.070)	0.541	0.014	0.020	0.992	1.189	1.417
5	8	40	0.812	(0.805, 0.819)	0.000	(0.000, 0.000)	0.106	(0.104, 0.107)	0.292	0.034	0.015	0.997	1.687	1.989
3.5	11	40	0.813	(0.805, 0.821)	0.000	(0.000, 0.009)	0.130	(0.125, 0.135)	0.135	0.007	0.007	0.999	1.876	2.178
2.5	16	40	0.799	(0.786, 0.812)	0.000	(0.000, 0.022)	0.178	(0.166, 0.191)	0.269	0.014	0.010	0.998	2.119	2.564
2.1	22	40	0.793	(0.785, 0.801)	0.030	(0.016, 0.043)	0.193	(0.184, 0.201)	0.119	0.006	0.007	0.999	2.224	2.731
2	24	40	0.794	(0.787, 0.801)	0.000	(0.000, 0.013)	0.219	(0.211, 0.227)	0.137	0.013	0.008	0.999	2.318	2.802
1.8	29	40	0.810	(0.800, 0.820)	0.086	(0.067, 0.106)	0.213	(0.200, 0.226)	0.174	0.011	0.007	0.999	2.473	3.013
1.7	32	40	0.815	(0.804, 0.825)	0.109	(0.087, 0.132)	0.230	(0.214, 0.247)	0.123	0.008	0.007	0.999	2.651	3.184
1.6	35	40	0.812	(0.800, 0.824)	0.121	(0.088, 0.154)	0.265	(0.242, 0.288)	0.157	0.011	0.007	0.998	2.798	3.423

Table L.2: The fit parameters and goodness-of-fit metrics for the Pp (LQ) model fits ($\exp(-\alpha D - \beta D^2)$) to our simulated data using Method 3 from Section (8.3.2).

Energy (MeV)	LET (keV/μm)	Repair Time (hrs)	α	α 95% Conf. Interval	β	β 95% Conf. Interval	χ^2	SSE	RMSE	R^2	RBE_{10}	RBE_{37}
50	1.2	2	0.040	(0.032, 0.048)	0.020	(0.018, 0.022)	0.251	0.007	0.015	0.996	1.091	1.021
20	2.6	2	0.085	(0.070, 0.100)	0.016	(0.013, 0.019)	1.167	0.034	0.031	0.981	1.094	1.091
5	8	2	0.108	(0.099, 0.118)	0.029	(0.026, 0.031)	2.043	0.132	0.029	0.988	1.461	1.446
3.5	11	2	0.115	(0.105, 0.125)	0.035	(0.032, 0.037)	1.864	0.106	0.026	0.992	1.596	1.582
2.5	16	2	0.168	(0.156, 0.180)	0.031	(0.028, 0.034)	2.975	0.106	0.028	0.988	1.678	1.736
2.1	22	2	0.184	(0.172, 0.195)	0.030	(0.027, 0.033)	2.784	0.096	0.026	0.990	1.711	1.799
2	24	2	0.186	(0.175, 0.196)	0.032	(0.029, 0.034)	5.344	0.221	0.031	0.987	1.740	1.831
1.8	29	2	0.207	(0.196, 0.218)	0.029	(0.026, 0.032)	4.645	0.157	0.028	0.988	1.756	1.892
1.7	32	2	0.227	(0.217, 0.237)	0.026	(0.024, 0.029)	1.835	0.126	0.026	0.990	1.771	1.954
1.6	35	2	0.241	(0.232, 0.251)	0.027	(0.025, 0.030)	2.437	0.133	0.024	0.989	1.835	2.030
50	1.2	16	0.070	(0.057, 0.084)	0.025	(0.022, 0.028)	0.72	0.014	0.021	0.993	1.272	1.224
20	2.6	16	0.124	(0.107, 0.142)	0.018	(0.014, 0.022)	1.658	0.031	0.030	0.983	1.273	1.319
5	8	16	0.170	(0.157, 0.184)	0.033	(0.030, 0.037)	4.480	0.155	0.032	0.986	1.726	1.787
3.5	11	16	0.181	(0.167, 0.195)	0.042	(0.038, 0.046)	4.418	0.131	0.029	0.989	1.917	1.971
2.5	16	16	0.266	(0.249, 0.283)	0.036	(0.031, 0.041)	7.617	0.109	0.028	0.986	2.065	2.265
2.1	22	16	0.308	(0.291, 0.324)	0.034	(0.028, 0.039)	7.280	0.092	0.026	0.989	2.174	2.448
2	24	16	0.299	(0.283, 0.314)	0.042	(0.037, 0.048)	18.207	0.210	0.030	0.984	2.278	2.519
1.8	29	16	0.372	(0.357, 0.387)	0.033	(0.028, 0.039)	13.279	0.112	0.024	0.988	2.383	2.777
1.7	32	16	0.417	(0.401, 0.432)	0.028	(0.022, 0.033)	6.325	0.102	0.023	0.988	2.455	2.936
1.6	35	16	0.470	(0.456, 0.485)	0.021	(0.017, 0.026)	6.062	0.096	0.021	0.987	2.569	3.152
50	1.2	40	0.076	(0.062, 0.090)	0.026	(0.022, 0.029)	0.775	0.014	0.021	0.993	1.309	1.265
20	2.6	40	0.131	(0.113, 0.148)	0.019	(0.015, 0.023)	1.724	0.028	0.028	0.984	1.313	1.361
5	8	40	0.185	(0.170, 0.199)	0.034	(0.030, 0.038)	5.012	0.161	0.032	0.985	1.781	1.865
3.5	11	40	0.195	(0.180, 0.209)	0.044	(0.040, 0.048)	5.266	0.134	0.03	0.989	1.974	2.047
2.5	16	40	0.290	(0.273, 0.308)	0.035	(0.030, 0.041)	8.476	0.109	0.028	0.985	2.143	2.382
2.1	22	40	0.333	(0.316, 0.351)	0.033	(0.027, 0.039)	8.307	0.093	0.026	0.988	2.256	2.591
2	24	40	0.319	(0.303, 0.336)	0.045	(0.039, 0.051)	23.697	0.210	0.030	0.983	2.374	2.656
1.8	29	40	0.402	(0.387, 0.417)	0.033	(0.028, 0.038)	13.889	0.102	0.023	0.989	2.483	2.917
1.7	32	40	0.453	(0.437, 0.469)	0.026	(0.020, 0.032)	7.279	0.094	0.022	0.988	2.574	3.107
1.6	35	40	0.513	(0.499, 0.528)	0.018	(0.013, 0.023)	6.003	0.084	0.019	0.987	2.695	3.365

Table L.3: The fit parameters and goodness-of-fit metrics for the fPp and Pp model fits ($E_{\gamma,1}[-(aD+bD^2)]$ and $\exp(-\alpha D-\beta D^2)$, respectively) to our simulated data using Method 1 of Section (8.3.2) . **Note:** Where $a = 0$, the true parameter is not exactly zero but is < 0.001.

Energy (MeV)	LET (keV/μm)	Repair Time (hrs)	γ	γ 95% Conf. Interval	a	a 95% Conf. Interval	b	b 95% Conf. Interval	χ^2	α	α 95% Conf. Interval	β	β 95% Conf. Interval	χ^2
50	1.2	2	0.855	(0.829, 0.881)	0.000	(0.000, 0.000)	0.027	(0.026, 0.028)	7.637	0.036	(0.021, 0.059)	0.021	(-0.003, 0.024)	20.169
20	2.6	2	0.866	(0.657, 1.000)	0.070	(0.000, 0.192)	0.021	(0.000, 0.052)	83.534	0.118	(0.039, 0.128)	0.011	(-0.014, 0.025)	67.150
5	8	2	0.833	(0.783, 0.880)	0.001	(0.000, 0.036)	0.060	(0.047, 0.072)	4.162	0.112	(0.053, 0.162)	0.026	(-0.016, 0.042)	56.588
3.5	11	2	0.834	(0.815, 0.854)	0.000	(0.000, 0.000)	0.071	(0.068, 0.075)	1.754	0.099	(0.055, 0.174)	0.036	(-0.014, 0.049)	30.519
2.5	16	2	0.814	(0.803, 0.825)	0.000	(0.000, 0.000)	0.091	(0.089, 0.093)	0.751	0.161	(0.102, 0.234)	0.030	(-0.017, 0.047)	26.229
2.1	22	2	0.803	(0.799, 0.807)	0.000	(0.000, 0.000)	0.099	(0.098, 0.100)	0.088	0.196	(0.12, 0.253)	0.024	(-0.022, 0.046)	19.381
2	24	2	0.804	(0.785, 0.824)	0.000	(0.000, 0.000)	0.101	(0.096, 0.107)	5.125	0.197	(0.106, 0.266)	0.025	(-0.028, 0.053)	102.257
1.8	29	2	0.792	(0.788, 0.796)	0.000	(0.000, 0.000)	0.109	(0.108, 0.111)	0.058	0.225	(0.134, 0.28)	0.020	(-0.028, 0.049)	18.057
1.7	32	2	0.814	(0.796, 0.831)	0.021	(0.000, 0.044)	0.107	(0.095, 0.118)	0.283	0.237	(0.161, 0.293)	0.021	(-0.022, 0.044)	22.507
1.6	35	2	0.798	(0.789, 0.808)	0.006	(0.000, 0.022)	0.125	(0.117, 0.133)	0.040	0.269	(0.172, 0.312)	0.017	(-0.028, 0.045)	11.564
50	1.2	16	0.838	(0.818, 0.857)	0.000	(0.000, 0.000)	0.041	(0.039, 0.042)	5.545	0.072	(0.036, 0.104)	0.023	(-0.009, 0.032)	34.898
20	2.6	16	0.863	(0.522, 1.000)	0.078	(0.000, 0.299)	0.032	(0.000, 0.093)	79.102	0.149	(0.067, 0.176)	0.014	(-0.017, 0.032)	62.514
5	8	16	0.818	(0.805, 0.831)	0.000	(0.000, 0.000)	0.095	(0.091, 0.098)	1.857	0.165	(0.093, 0.245)	0.030	(-0.024, 0.054)	52.020
3.5	11	16	0.811	(0.799, 0.823)	0.000	(0.000, 0.000)	0.119	(0.114, 0.124)	0.767	0.194	(0.098, 0.263)	0.034	(-0.032, 0.065)	36.648
2.5	16	16	0.800	(0.793, 0.807)	0.000	(0.000, 0.000)	0.160	(0.157, 0.163)	0.354	0.276	(0.172, 0.359)	0.030	(-0.034, 0.064)	45.706
2.1	22	16	0.793	(0.784, 0.802)	0.024	(0.004, 0.043)	0.176	(0.165, 0.187)	0.072	0.370	(0.215, 0.407)	0.010	(-0.052, 0.062)	19.758
2	24	16	0.792	(0.780, 0.804)	0.000	(0.000, 0.000)	0.196	(0.187, 0.204)	2.251	0.366	(0.181, 0.416)	0.015	(-0.068, 0.083)	115.696
1.8	29	16	0.793	(0.781, 0.803)	0.043	(0.013, 0.069)	0.218	(0.201, 0.236)	0.080	0.459	(0.27, 0.473)	0.002	(-0.066, 0.068)	20.658
1.7	32	16	0.823	(0.811, 0.834)	0.105	(0.073, 0.138)	0.200	(0.179, 0.221)	0.073	0.510	(0.311, 0.522)	-0.001	(-0.066, 0.065)	18.555
1.6	35	16	0.805	(0.794, 0.816)	0.085	(0.046, 0.123)	0.254	(0.228, 0.279)	0.034	0.563	(0.368, 0.571)	-0.008	(-0.065, 0.056)	9.335
50	1.2	40	0.851	(0.830, 0.872)	0.000	(0.000, 0.000)	0.043	(0.041, 0.045)	3.669	0.051	(0.041, 0.113)	0.030	(-0.004, 0.034)	32.451
20	2.6	40	0.780	(0.695, 0.866)	0.000	(0.000, 0.000)	0.058	(0.051, 0.066)	105.078	0.140	(0.075, 0.181)	0.018	(-0.014, 0.032)	117.449
5	8	40	0.813	(0.799, 0.827)	0.000	(0.000, 0.000)	0.104	(0.100, 0.108)	2.119	0.202	(0.101, 0.264)	0.026	(-0.031, 0.057)	73.400
3.5	11	40	0.809	(0.798, 0.820)	0.000	(0.000, 0.000)	0.129	(0.124, 0.134)	0.938	0.204	(0.107, 0.282)	0.034	(-0.035, 0.069)	46.065
2.5	16	40	0.795	(0.788, 0.802)	0.000	(0.000, 0.000)	0.179	(0.175, 0.182)	0.373	0.351	(0.192, 0.389)	0.017	(-0.050, 0.066)	47.502
2.1	22	40	0.788	(0.778, 0.799)	0.021	(0.000, 0.046)	0.198	(0.183, 0.213)	0.079	0.415	(0.235, 0.439)	0.007	(-0.058, 0.065)	19.302
2	24	40	0.792	(0.780, 0.804)	0.000	(0.000, 0.000)	0.216	(0.206, 0.225)	2.273	0.374	(0.192, 0.446)	0.017	(-0.073, 0.091)	105.222
1.8	29	40	0.797	(0.784, 0.810)	0.060	(0.021, 0.100)	0.231	(0.206, 0.257)	0.113	0.530	(0.299, 0.504)	-0.008	(-0.078, 0.069)	19.200
1.7	32	40	0.813	(0.798, 0.829)	0.104	(0.056, 0.151)	0.234	(0.201, 0.268)	0.129	0.552	(0.343, 0.561)	-0.005	(-0.071, 0.066)	23.365
1.6	35	40	0.807	(0.793, 0.821)	0.105	(0.048, 0.161)	0.277	(0.240, 0.314)	0.057	0.626	(0.411, 0.616)	-0.015	(-0.069, 0.054)	8.245

L.2 Depth-varied LET

Table L.4: The fit parameters and goodness-of-fit metrics for the fPp (Mittag-Leffler) model fits ($aD + bD^2$) of the fPp model fits to our simulated data using Method 3 of Section (8.3.2) for a range of depth in liquid water. **Note:** Where $a = 0$, the true parameter is not exactly zero but is < 0.001.

Depth (mm)	LET (keV/μm)	Repair Time (hrs)	γ	γ 95% Conf. Interval	a	a 95% Conf. Interval	b	b 95% Conf. Interval	χ^2	SSE	RMSE	R^2	RBE_{10}	RBE_{37}
1.2	3	2	0.843	(0.834, 0.853)	0.000	(0.000, 0.000)	0.035	(0.034, 0.035)	0.149	0.002	0.008	0.999	1.030	1.152
2.6	4	2	0.841	(0.833, 0.849)	0.000	(0.000, 0.000)	0.050	(0.049, 0.050)	0.075	0.002	0.007	0.999	1.226	1.374
3.9	8	2	0.826	(0.806, 0.846)	0.000	(0.000, 0.016)	0.080	(0.073, 0.086)	0.191	0.002	0.008	0.999	1.506	1.719
4.08	14	2	0.833	(0.808, 0.858)	0.000	(0.000, 0.026)	0.096	(0.084, 0.107)	0.597	0.003	0.010	0.999	1.676	1.879
4.2	20	2	0.799	(0.782, 0.817)	0.000	(0.000, 0.018)	0.110	(0.102, 0.118)	0.198	0.001	0.007	0.999	1.674	2.011
4.32	25	2	0.801	(0.784, 0.817)	0.000	(0.000, 0.000)	0.108	(0.104, 0.112)	0.452	0.006	0.014	0.998	1.668	1.999
4.4	28	2	0.816	(0.764, 0.868)	0.020	(0.000, 0.071)	0.099	(0.077, 0.121)	1.489	0.011	0.019	0.994	1.673	1.989
1.2	3	16	0.822	(0.804, 0.840)	0.000	(0.000, 0.013)	0.052	(0.048, 0.057)	0.223	0.002	0.008	0.999	1.214	1.398
2.6	4	16	0.824	(0.809, 0.839)	0.000	(0.000, 0.012)	0.078	(0.073, 0.083)	0.143	0.002	0.007	0.999	1.489	1.706
3.9	8	16	0.807	(0.789, 0.824)	0.000	(0.000, 0.022)	0.137	(0.125, 0.148)	0.254	0.001	0.007	0.999	1.902	2.220
4.08	14	16	0.811	(0.796, 0.827)	0.002	(0.000, 0.026)	0.179	(0.166, 0.192)	0.290	0.001	0.006	0.999	2.194	2.598
4.2	20	16	0.789	(0.771, 0.808)	0.026	(0.000, 0.058)	0.219	(0.197, 0.241)	0.070	0.001	0.004	0.999	2.321	2.858
4.32	25	16	0.790	(0.751, 0.829)	0.020	(0.000, 0.076)	0.240	(0.196, 0.284)	0.411	0.005	0.012	0.999	2.421	2.955
4.4	28	16	0.820	(0.740, 0.900)	0.122	(0.000, 0.252)	0.186	(0.096, 0.276)	0.877	0.008	0.016	0.994	2.414	2.975
1.2	3	40	0.817	(0.807, 0.828)	0.000	(0.000, 0.000)	0.057	(0.056, 0.059)	0.242	0.003	0.010	0.999	1.256	1.455
2.6	4	40	0.819	(0.809, 0.829)	0.000	(0.000, 0.000)	0.087	(0.085, 0.089)	0.194	0.002	0.009	0.999	1.545	1.782
3.9	8	40	0.806	(0.788, 0.824)	0.000	(0.000, 0.024)	0.152	(0.139, 0.165)	0.219	0.001	0.007	0.999	1.987	2.334
4.08	14	40	0.820	(0.800, 0.841)	0.026	(0.000, 0.058)	0.185	(0.165, 0.205)	0.368	0.001	0.006	0.999	2.290	2.708
4.2	20	40	0.782	(0.770, 0.795)	0.020	(0.000, 0.045)	0.253	(0.235, 0.271)	0.048	<0.001	0.003	0.999	2.441	3.013
4.32	25	40	0.801	(0.764, 0.839)	0.049	(0.000, 0.108)	0.250	(0.204, 0.296)	0.404	0.004	0.011	0.999	2.584	3.099
4.4	28	40	0.828	(0.745, 0.910)	0.157	(0.011, 0.302)	0.194	(0.089, 0.299)	0.831	0.008	0.016	0.994	2.567	3.137

Table L.5: The fit parameters and goodness-of-fit metrics for the Pp (LQ) model fits ($\exp(-\alpha D - \beta D^2)$) to our simulated data using Method 3 of Section (8.3.2) for a range of depth in liquid water.

Depth (mm)	LET (keV/μm)	Repair Time (hrs)	α	α 95% Conf. Interval	β	β 95% Conf. Interval	χ^2	SSE	RMSE	R^2	RBE_{10}	RBE_{37}
1.2	3	2	0.074	(0.057, 0.092)	0.018	(0.015, 0.021)	2.951	0.023	0.028	0.992	1.119	1.095
2.6	4	2	0.092	(0.073, 0.110)	0.025	(0.021, 0.029)	1.891	0.020	0.026	0.993	1.336	1.313
3.9	8	2	0.130	(0.103, 0.156)	0.034	(0.027, 0.041)	6.405	0.029	0.031	0.989	1.625	1.627
4.08	14	2	0.145	(0.111, 0.179)	0.041	(0.032, 0.051)	27.039	0.028	0.031	0.988	1.792	1.799
4.2	20	2	0.172	(0.136, 0.207)	0.041	(0.030, 0.052)	246.741	0.027	0.030	0.989	1.865	1.910
4.32	25	2	0.158	(0.123, 0.192)	0.039	(0.028, 0.049)	50.799	0.022	0.030	0.993	1.803	1.839
4.4	28	2	0.183	(0.156, 0.210)	0.037	(0.029, 0.045)	2.086	0.019	0.025	0.990	1.833	1.906
1.2	3	16	0.114	(0.090, 0.138)	0.020	(0.015, 0.024)	5.716	0.030	0.032	0.990	1.281	1.306
2.6	4	16	0.145	(0.115, 0.176)	0.030	(0.022, 0.037)	5.552	0.032	0.033	0.989	1.582	1.621
3.9	8	16	0.208	(0.170, 0.246)	0.043	(0.031, 0.056)	20.029	0.032	0.033	0.987	2.001	2.094
4.08	14	16	0.258	(0.214, 0.302)	0.051	(0.036, 0.067)	83.064	0.022	0.027	0.988	2.282	2.425
4.2	20	16	0.327	(0.285, 0.367)	0.049	(0.033, 0.065)	1282.917	0.017	0.023	0.991	2.455	2.729
4.32	25	16	0.288	(0.241, 0.335)	0.068	(0.045, 0.092)	1167.958	0.016	0.026	0.994	2.735	2.822
4.4	28	16	0.387	(0.344, 0.429)	0.038	(0.023, 0.053)	3.624	0.018	0.024	0.988	2.504	2.894
1.2	3	40	0.124	(0.098, 0.150)	0.020	(0.015, 0.025)	6.605	0.033	0.033	0.989	1.324	1.356
2.6	4	40	0.159	(0.124, 0.193)	0.031	(0.022, 0.040)	7.248	0.037	0.035	0.987	1.652	1.697
3.9	8	40	0.223	(0.183, 0.263)	0.046	(0.032, 0.060)	24.824	0.032	0.033	0.986	2.095	2.199
4.08	14	40	0.281	(0.235, 0.326)	0.053	(0.037, 0.070)	102.569	0.020	0.026	0.988	2.372	2.572
4.2	20	40	0.359	(0.316, 0.402)	0.049	(0.032, 0.067)	1628.576	0.016	0.023	0.991	2.576	2.877
4.32	25	40	0.313	(0.267, 0.358)	0.074	(0.051, 0.099)	2414.682	0.013	0.023	0.995	2.841	2.970
4.4	28	40	0.422	(0.379, 0.466)	0.037	(0.021, 0.053)	3.713	0.016	0.023	0.988	2.645	3.066

Table L.6: The fit parameters and goodness-of-fit metrics for the fPp (Mittag-Leffler) and Pp (LQ) model fits ($E_{\gamma,1}[-(aD+bD^2)]$ and $\exp(-\alpha D-\beta D^2)$, respectively) to our simulated data using Method 1 of Section (8.3.2) for a range of depth in liquid water.

Depth (mm)	LET (keV/μm)	Repair Time (hrs)	γ	γ 95% Conf. Interval	a	a 95% Conf. Interval	b	b 95% Conf. Interval	χ^2	α	α 95% Conf. Interval	β	β 95% Conf. Interval	χ^2
1.2	3	2	0.839	(0.801, 0.877)	0.000	(0.000, 0.000)	0.035	(0.032, 0.038)	125.540	0.087	(0.013, 0.136)	0.016	(-0.012, 0.028)	125.540
2.6	4	2	0.849	(0.814, 0.883)	0.000	(0.000, 0.000)	0.049	(0.044, 0.053)	149.396	0.065	(0.026, 0.157)	0.029	(-0.010, 0.039)	149.396
3.9	8	2	0.816	(0.789, 0.843)	0.000	(0.000, 0.000)	0.081	(0.076, 0.086)	124.323	0.175	(0.034, 0.225)	0.023	(-0.036, 0.060)	124.323
4.08	14	2	0.834	(0.788, 0.880)	0.000	(0.000, 0.000)	0.097	(0.084, 0.110)	113.424	0.155	(0.023, 0.253)	0.037	(-0.039, 0.076)	113.424
4.2	20	2	0.791	(0.770, 0.812)	0.000	(0.000, 0.000)	0.110	(0.103, 0.118)	345.856	0.297	(0.046, 0.297)	0.001	(-0.079, 0.080)	345.856
4.32	25	2	0.810	(0.796, 0.824)	0.004	(0.000, 0.013)	0.103	(0.097, 0.108)	125.326	0.122	(0.051, 0.265)	0.046	(-0.026, 0.072)	125.326
4.4	28	2	0.786	(0.774, 0.799)	0.000	(0.000, 0.000)	0.107	(0.105, 0.110)	6.939	0.232	(0.115, 0.250)	0.024	(-0.033, 0.057)	6.939
1.2	3	16	0.820	(0.788, 0.853)	0.000	(0.000, 0.000)	0.052	(0.048, 0.057)	204.272	0.120	(0.025, 0.215)	0.018	(-0.018, 0.036)	204.272
2.6	4	16	0.829	(0.795, 0.863)	0.000	(0.000, 0.000)	0.078	(0.073, 0.083)	353.093	0.109	(0.037, 0.254)	0.040	(-0.016, 0.057)	353.093
3.9	8	16	0.808	(0.781, 0.834)	0.000	(0.000, 0.000)	0.137	(0.129, 0.144)	220.929	0.182	(0.074, 0.343)	0.056	(-0.032, 0.087)	220.929
4.08	14	16	0.805	(0.784, 0.827)	0.000	(0.000, 0.000)	0.182	(0.168, 0.195)	249.344	0.273	(0.100, 0.395)	0.037	(-0.071, 0.108)	249.344
4.2	20	16	0.781	(0.757, 0.804)	0.008	(0.000, 0.081)	0.228	(0.189, 0.267)	87.192	0.467	(0.181, 0.471)	0.000	(-0.106, 0.106)	170.100
4.32	25	16	0.799	(0.788, 0.810)	0.047	(0.027, 0.066)	0.213	(0.199, 0.228)	292.502	0.369	(0.142, 0.435)	0.025	(-0.116, 0.141)	292.502
4.4	28	16	0.781	(0.768, 0.795)	0.000	(0.000, 0.000)	0.250	(0.241, 0.259)	8.737	0.504	(0.272, 0.502)	0.003	(-0.075, 0.078)	8.737
1.2	3	40	0.826	(0.785, 0.867)	0.000	(0.000, 0.000)	0.056	(0.050, 0.062)	128.705	0.096	(0.032, 0.216)	0.023	(-0.015, 0.038)	128.705
2.6	4	40	0.835	(0.793, 0.877)	0.000	(0.000, 0.000)	0.083	(0.073, 0.092)	219.575	0.095	(0.035, 0.281)	0.043	(-0.020, 0.063)	219.575
3.9	8	40	0.799	(0.776, 0.821)	0.000	(0.000, 0.000)	0.152	(0.144, 0.160)	111.316	0.270	(0.079, 0.366)	0.025	(-0.071, 0.096)	111.316
4.08	14	40	0.804	(0.728, 0.880)	0.001	(0.000, 0.143)	0.202	(0.108, 0.295)	65.606	0.314	(0.120, 0.421)	0.035	(-0.078, 0.114)	65.606
4.2	20	40	0.783	(0.771, 0.796)	0.022	(0.000, 0.054)	0.251	(0.228, 0.274)	188.718	0.450	(0.206, 0.511)	0.000	(-0.111, 0.111)	209.709
4.32	25	40	0.812	(0.798, 0.827)	0.081	(0.053, 0.109)	0.219	(0.194, 0.244)	126.007	0.363	(0.173, 0.454)	0.019	(-0.129, 0.149)	126.007
4.4	28	40	0.793	(0.750, 0.837)	0.046	(0.000, 0.213)	0.256	(0.170, 0.342)	5.355	0.559	(0.309, 0.535)	-0.002	(-0.081, 0.079)	5.355

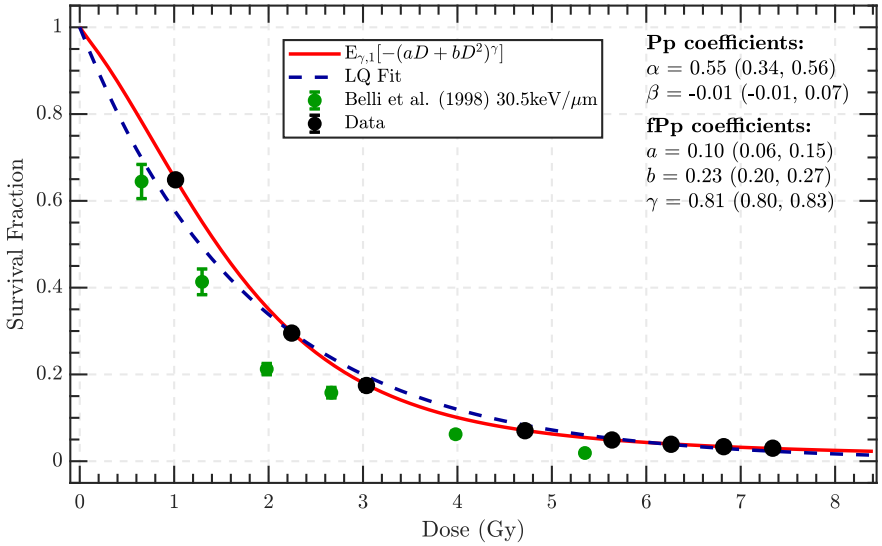
L.3 Hypothesis Testing Results

Table L.7: Hypothesis testing results of the fPp (Mittag-Leffler) and Pp (LQ) model fits to our simulated data after 2 and 40 hours of repair time for both zero and a range of depths in liquid water.

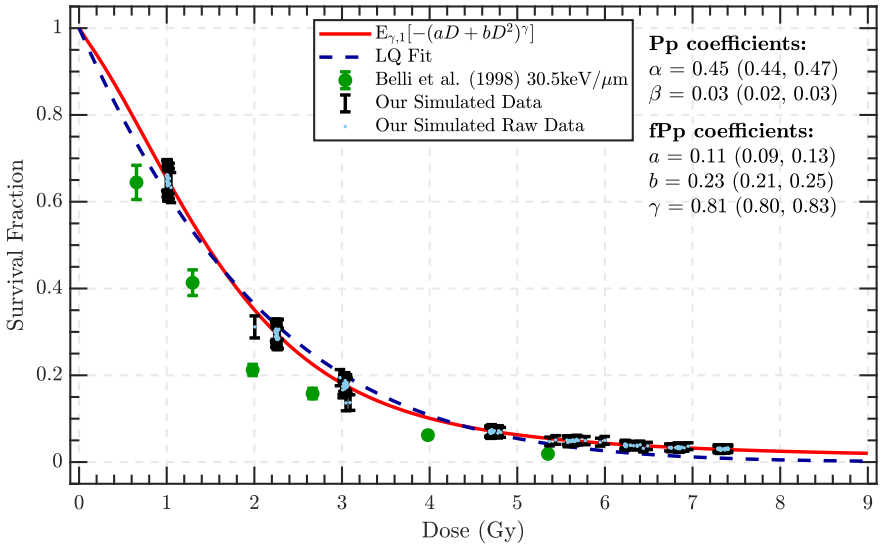
Energy	LET	Repair	KS Test		AD Test		BP Test		White Test	
(MeV)	(keV/μm)	Time (hrs)	FPp	Pp	FPp	Pp	FPp	Pp	FPp	Pp
50	1	2	✓	✓	✓	✓	✓	✓	✓	✓
20	3	2	✓	✓	✓	✓	✓	✓	✓	✓
5	8	2	✓	✓	✓	✓	✓	✓	✓	✓
3.5	11	2	✓	✓	✓	✓	✓	✓	✓	✓
2.5	16	2	✓	✓	✓	✓	✓	✓	✓	✓
2.1	22	2	✓	✓	✓	✓	✓	✓	✓	✓
2	24	2	✓	✓	✓	✓	✗	✓	✗	✓
1.8	29	2	✓	✓	✓	✓	✓	✓	✓	✓
1.7	32	2	✗	✓	✗	✓	✓	✗	✓	✓
1.6	35	2	✓	✓	✓	✓	✓	✗	✓	✓
50	1	40	✓	✓	✓	✓	✓	✓	✓	✓
20	3	40	✓	✓	✓	✓	✗	✓	✓	✓
5	8	40	✓	✓	✓	✓	✓	✓	✓	✓
3.5	11	40	✓	✓	✓	✓	✓	✓	✓	✓
2.5	16	40	✓	✓	✓	✓	✓	✓	✓	✓
2.1	22	40	✓	✗	✓	✓	✓	✓	✓	✓
2	24	40	✓	✓	✓	✓	✓	✗	✓	✓
1.8	29	40	✓	✓	✗	✗	✓	✓	✓	✓
1.7	32	40	✓	✗	✓	✗	✓	✓	✓	✗
1.6	35	40	✓	✗	✓	✗	✓	✓	✓	✗

Depth	LET	Repair	KS Test		AD Test		BP Test		White Test	
(mm)	(keV/μm)	Time (hrs)	fPp	cPp	fPp	cPp	fPp	cPp	fPp	cPp
1.2	3	2	✓	✓	✓	✓	✓	✓	✓	✓
2.6	4	2	✓	✓	✓	✓	✓	✓	✓	✓
3.9	8	2	✓	✓	✓	✓	✓	✓	✓	✓
4.1	14	2	✓	✓	✓	✓	✓	✓	✓	✓
4.2	20	2	✗	✓	✓	✓	✓	✓	✓	✓
4.3	25	2	✓	✓	✓	✓	✓	✓	✓	✓
4.4	28	2	✓	✓	✓	✗	✓	✓	✓	✓
1.2	3	40	✓	✓	✓	✓	✓	✓	✓	✓
2.6	4	40	✓	✓	✓	✓	✓	✓	✓	✓
3.9	8	40	✓	✓	✓	✓	✓	✓	✓	✓
4.1	14	40	✓	✓	✓	✓	✓	✓	✓	✓
4.2	20	40	✓	✓	✓	✓	✓	✗	✓	✓
4.3	25	40	✓	✓	✓	✗	✓	✓	✓	✓
4.4	28	40	✗	✗	✗	✗	✓	✓	✓	✓

L.4 Additional Fit Plots

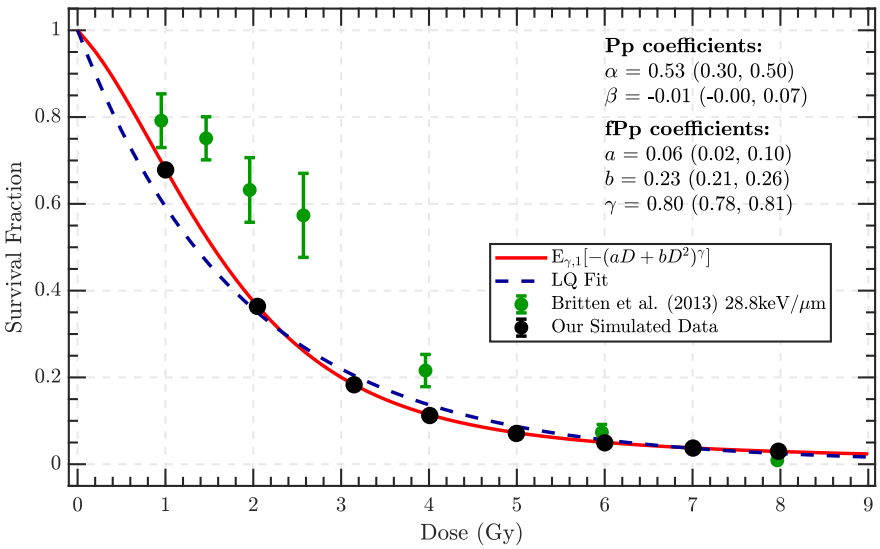


(a) Method 1

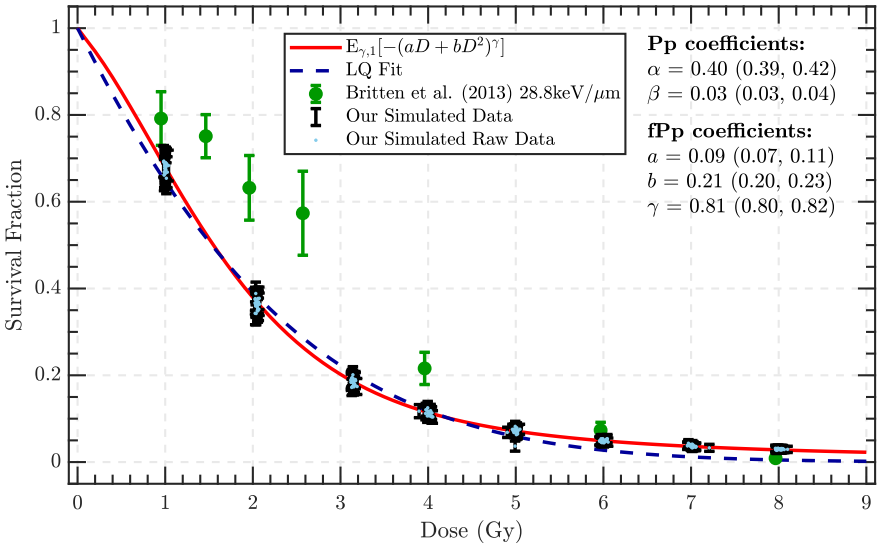


(b) Method 3

Figure L.1: Fits of the fPp (Mittag-Leffler) and Pp (LQ) models to our simulated survival data of V79 cells exposed to 1.7MeV (32keV/μm) proton radiation after 40 hours of repair time.

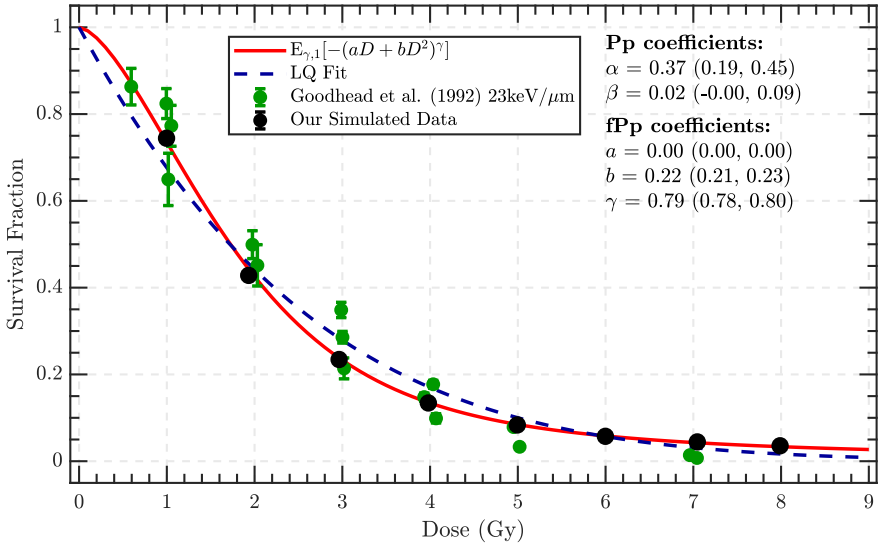


(a) Method 1

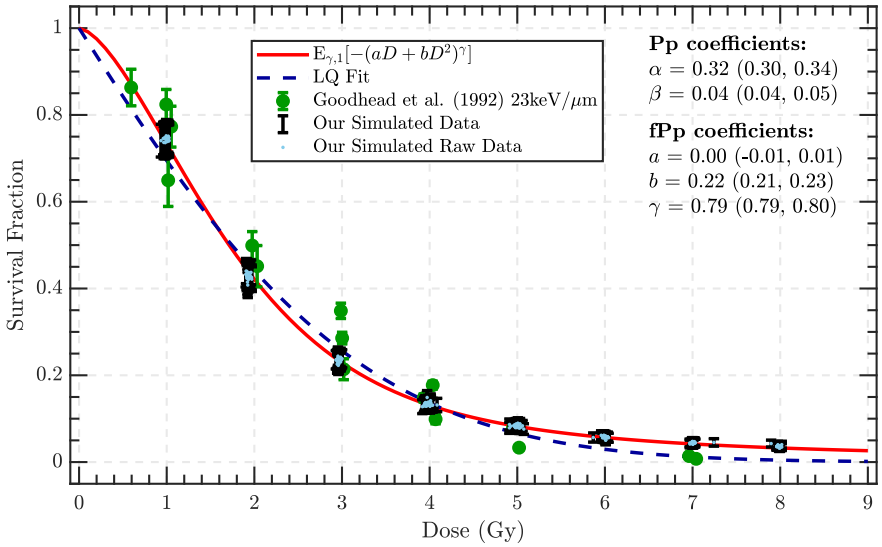


(b) Method 3

Figure L.2: Fits of the fPp (Mittag-Leffler) and Pp (LQ) models to our simulated survival data of V79 cells exposed to 1.8MeV (29keV/μm) proton radiation after 40 hours of repair time.

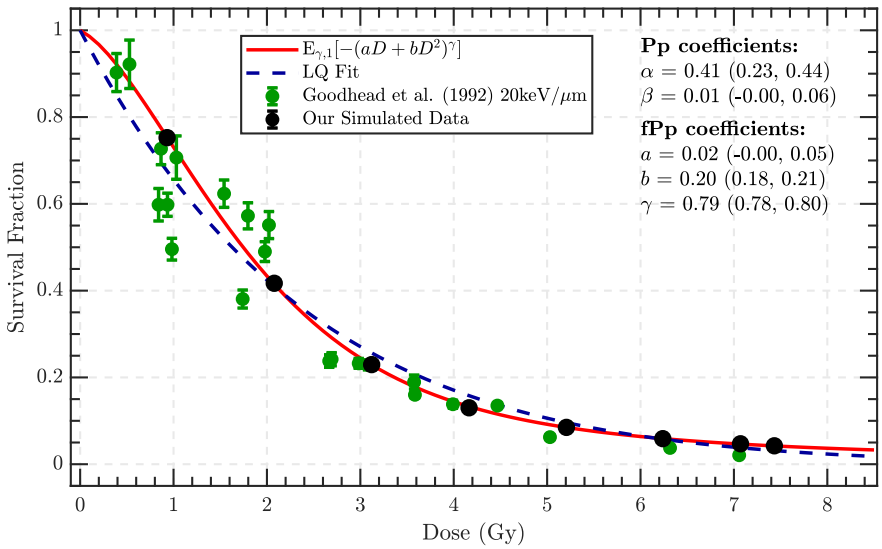


(a) Method 1

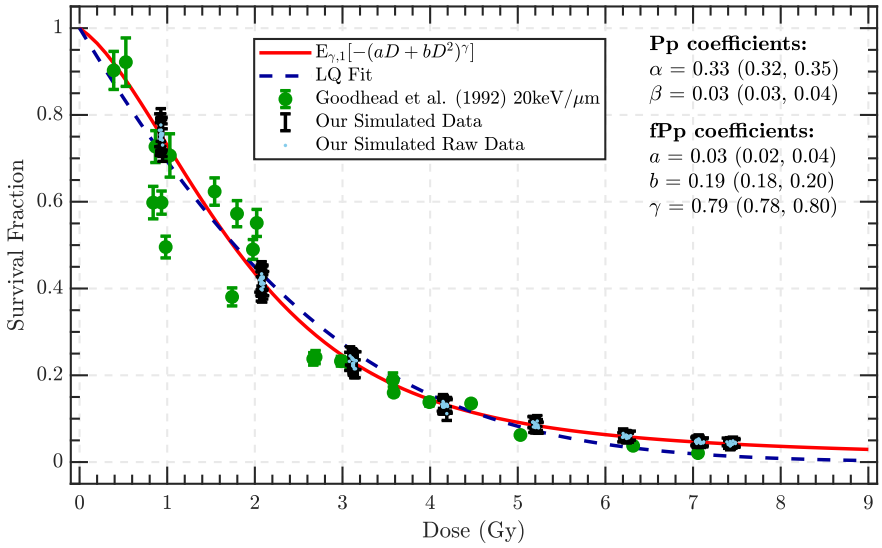


(b) Method 3

Figure L.3: Fits of the fPp (Mittag-Leffler) and Pp (LQ) models to our simulated survival data of V79 cells exposed to 2MeV (24keV/μm) proton radiation after 40 hours of repair time.

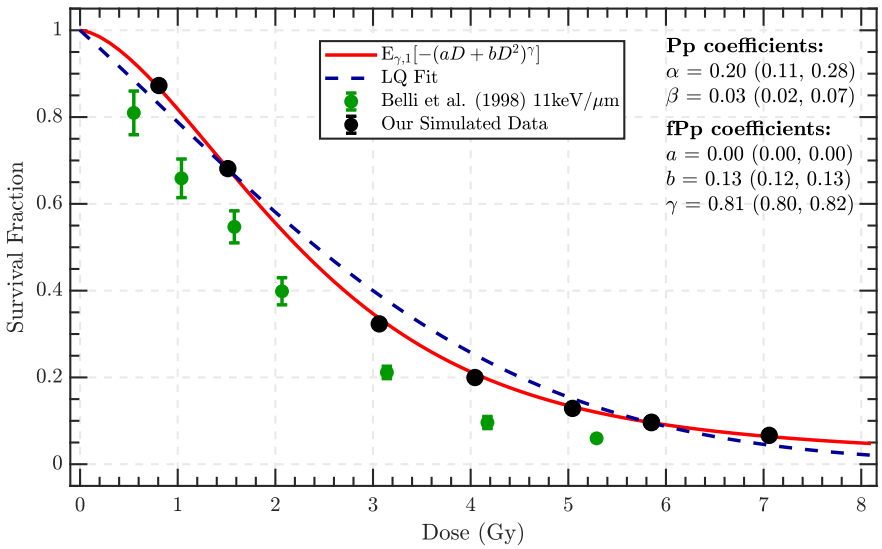


(a) Method 1

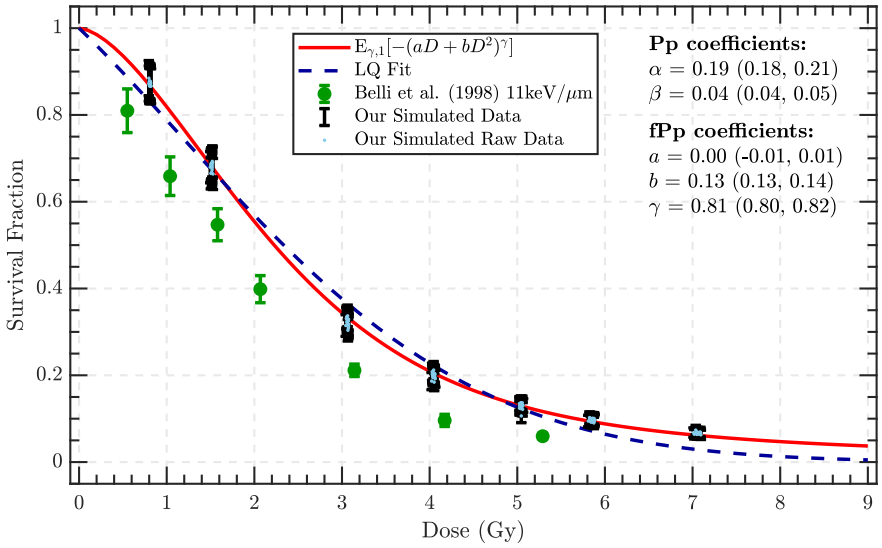


(b) Method 3

Figure L.4: Fits of the fPp (Mittag-Leffler) and Pp (LQ) models to our simulated survival data of V79 cells exposed to 2.1MeV (22keV/μm) proton radiation after 40 hours of repair time.

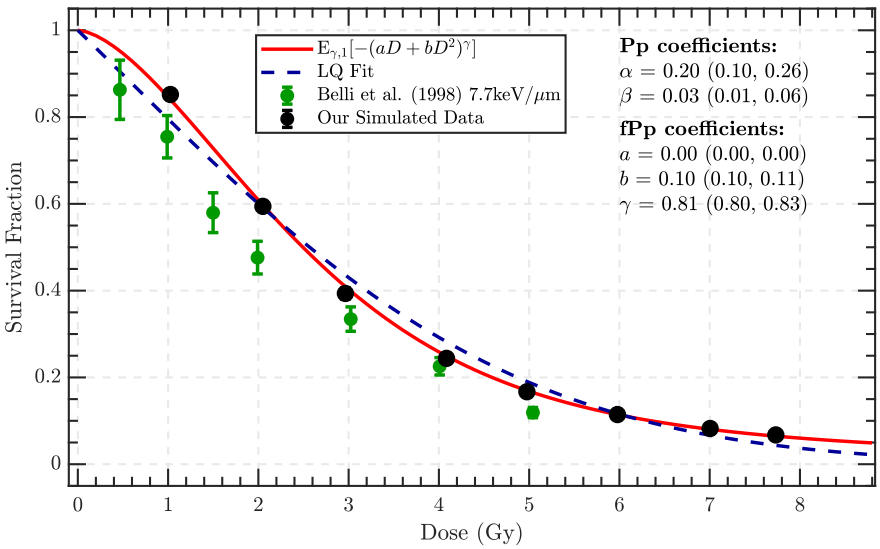


(a) Method 1

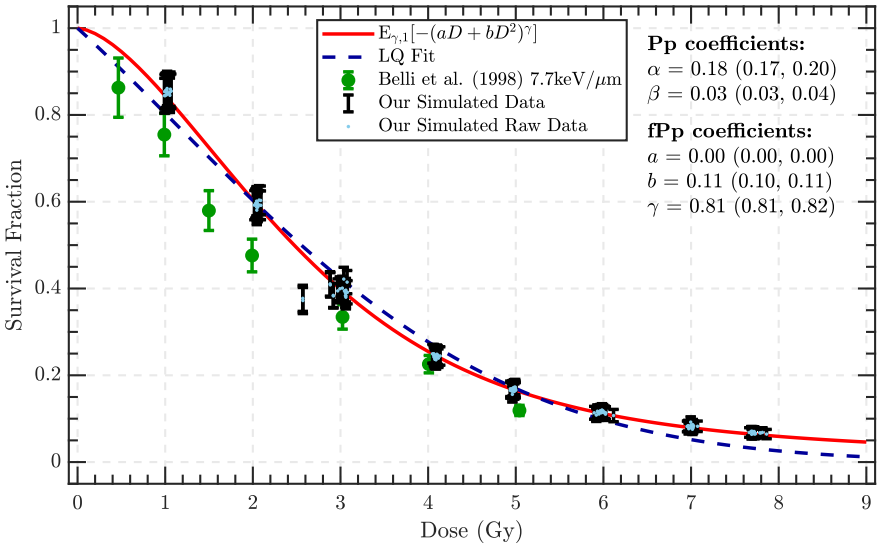


(b) Method 3

Figure L.5: Fits of the fPp (Mittag-Leffler) and Pp (LQ) models to our simulated survival data of V79 cells exposed to 3.5MeV (11keV/μm) proton radiation after 40 hours of repair time.

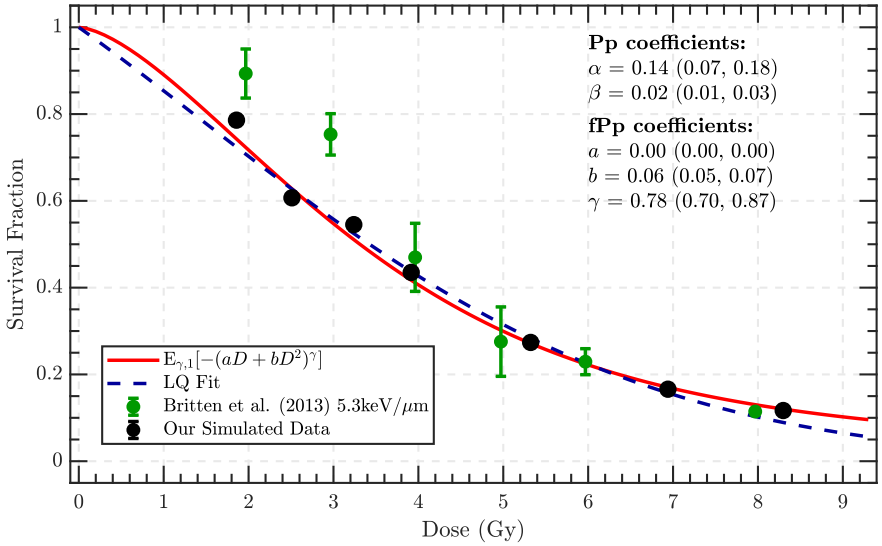


(a) Method 1

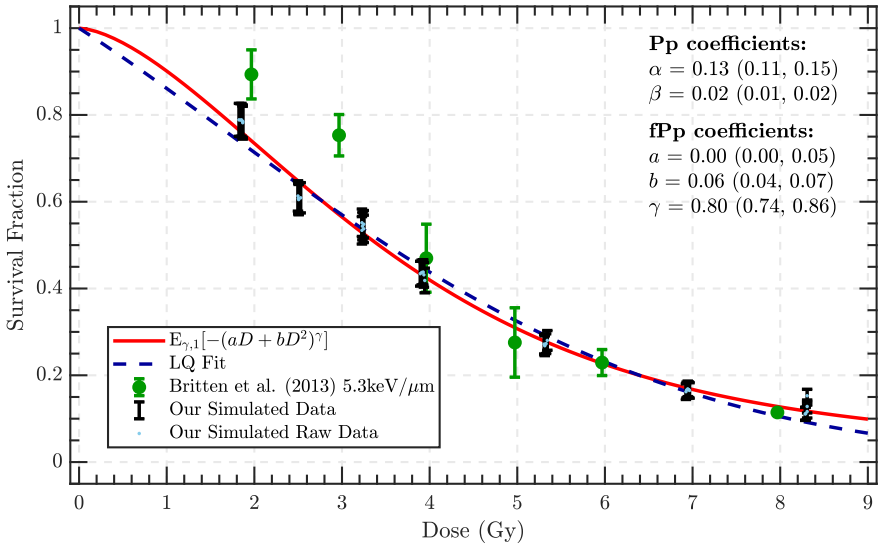


(b) Method 3

Figure L.6: Fits of the fPp (Mittag-Leffler) and Pp (LQ) models to our simulated survival data of V79 cells exposed to 5MeV (8keV/μm) proton radiation after 40 hours of repair time.

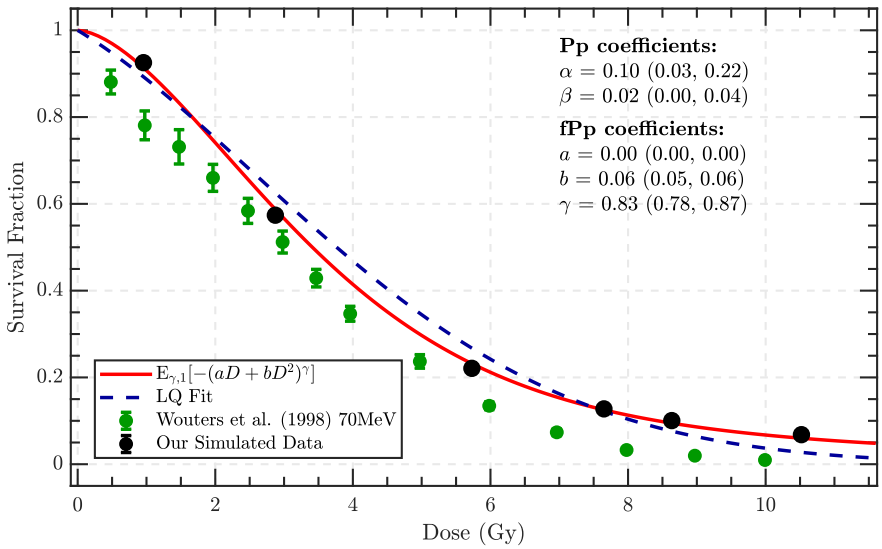


(a) Method 1

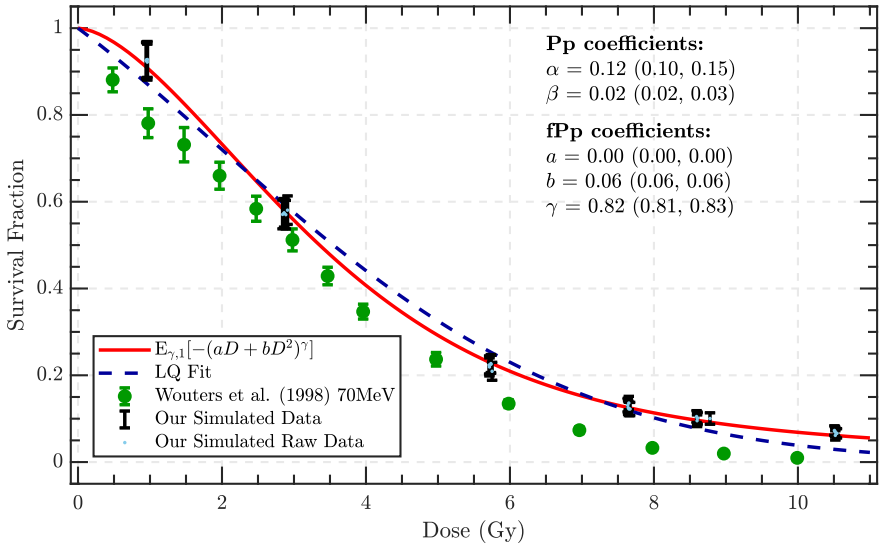


(b) Method 3

Figure L.7: Fits of the fPp (Mittag-Leffler) and Pp (LQ) models to our simulated survival data of V79 cells exposed to 20MeV (2.6keV/μm) proton radiation after 40 hours of repair time.

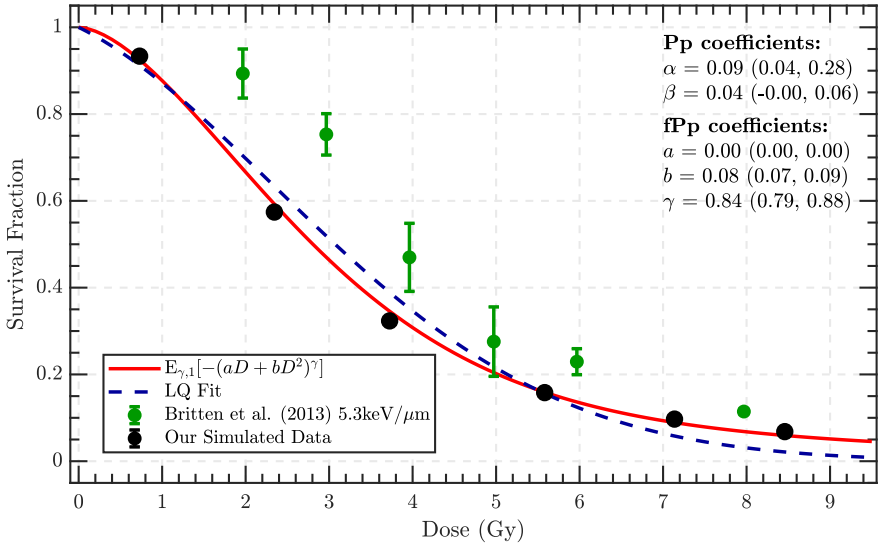


(a) Method 1

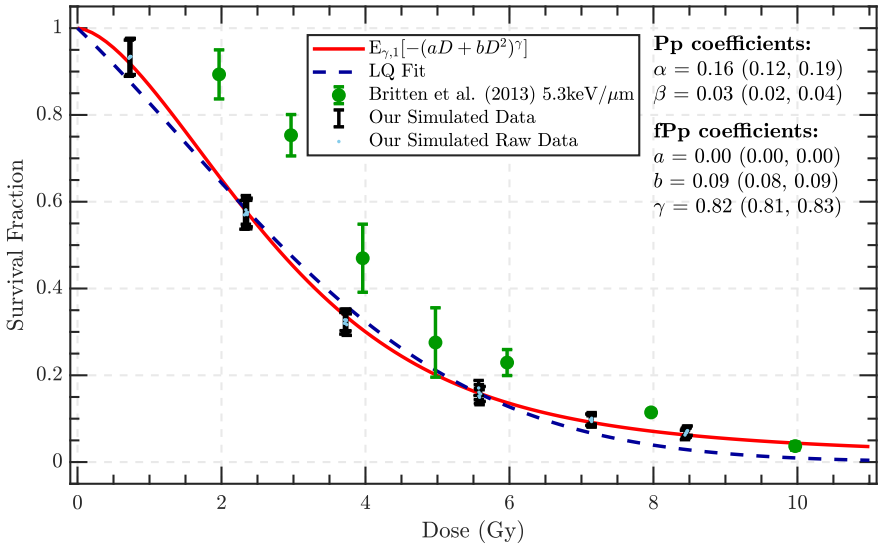


(b) Method 3

Figure L.8: Fits of the fPp (Mittag-Leffler) and Pp (LQ) models to our simulated survival data of V79 cells exposed to 20MeV (3keV/μm) proton radiation at 1.2mm depth in liquid water after 40 hours of repair time.

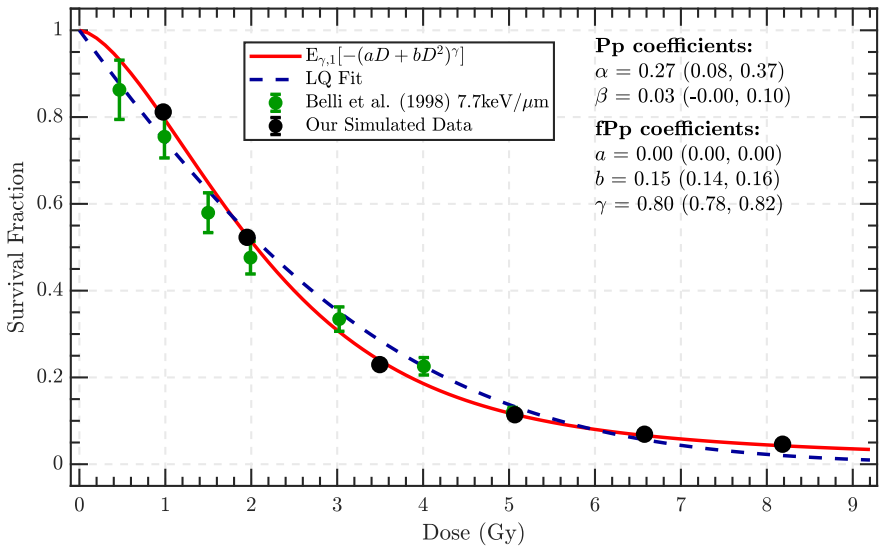


(a) Method 1

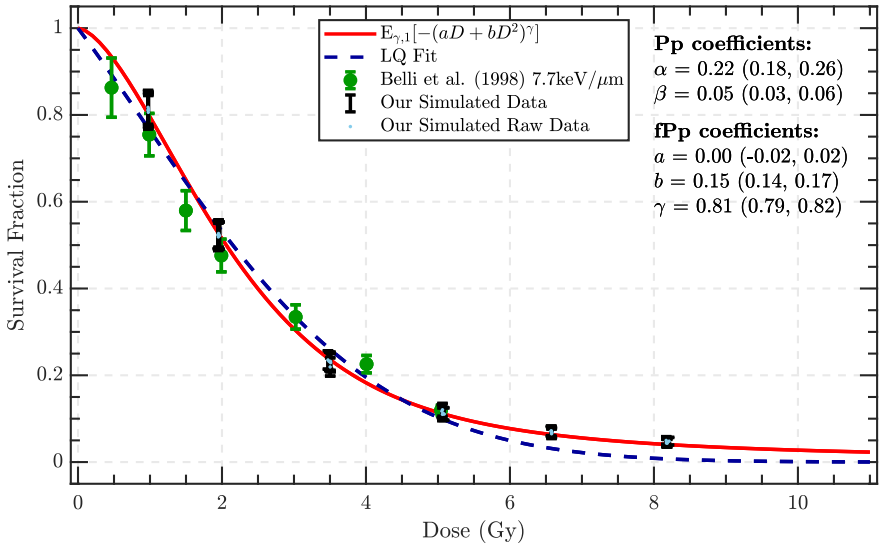


(b) Method 3

Figure L.9: Fits of the fPp (Mittag-Leffler) and Pp (LQ) models to our simulated survival data of V79 cells exposed to 20MeV (4keV/μm) proton radiation at 2.6mm depth in liquid water after 40 hours of repair time.

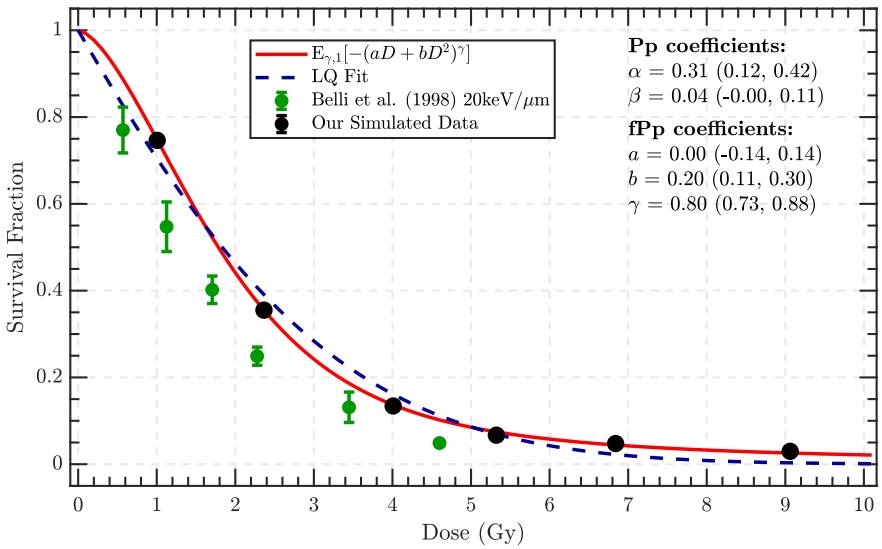


(a) Method 1

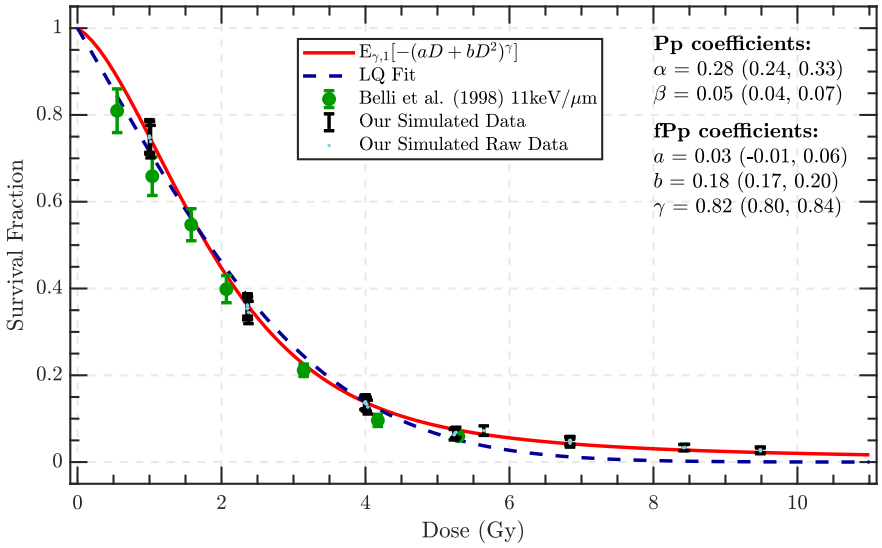


(b) Method 3

Figure L.10: Fits of the fPp (Mittag-Leffler) and Pp (LQ) models to our simulated survival data of V79 cells exposed to 20MeV (8keV/μm) proton radiation at 3.9mm depth in liquid water after 40 hours of repair time.

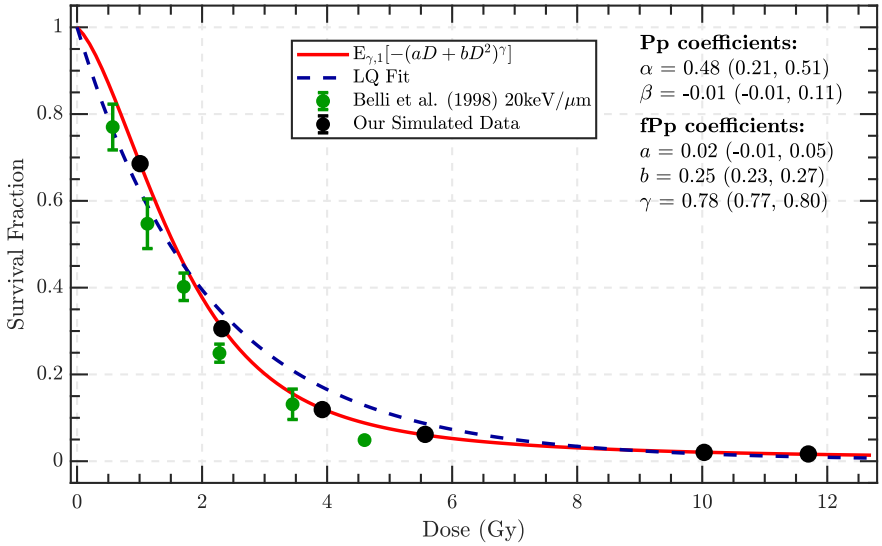


(a) Method 1

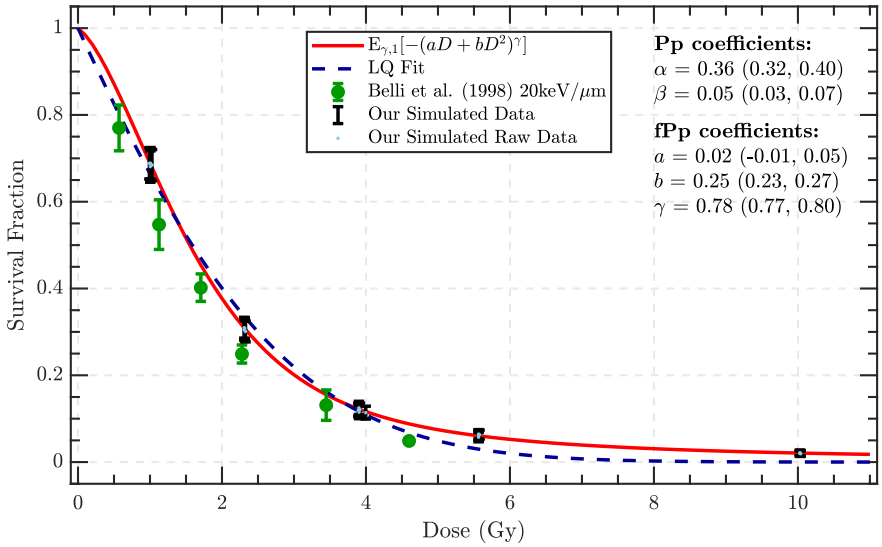


(b) Method 3

Figure L.11: Fits of the fPp (Mittag-Leffler) and Pp (LQ) models to our simulated survival data of V79 cells exposed to 20MeV (14keV/μm) proton radiation at 4.08mm depth in liquid water after 40 hours of repair time.

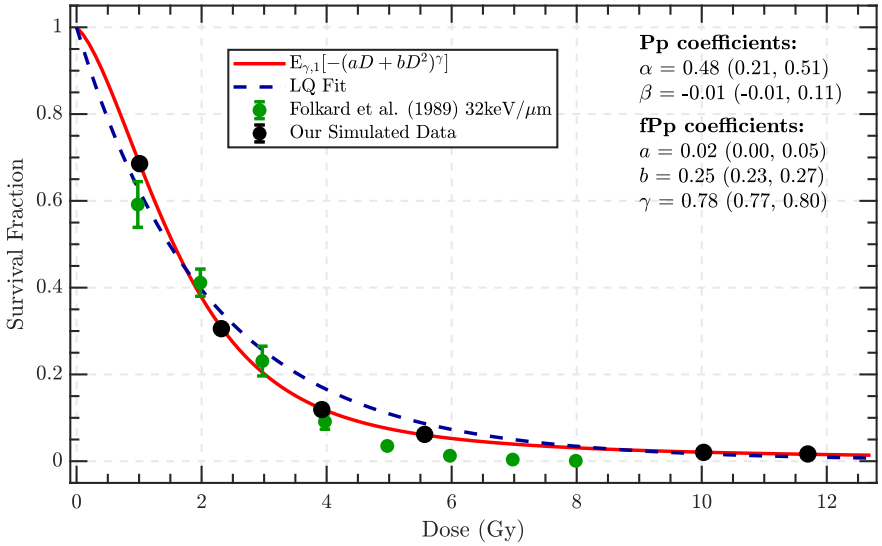


(a) Method 1

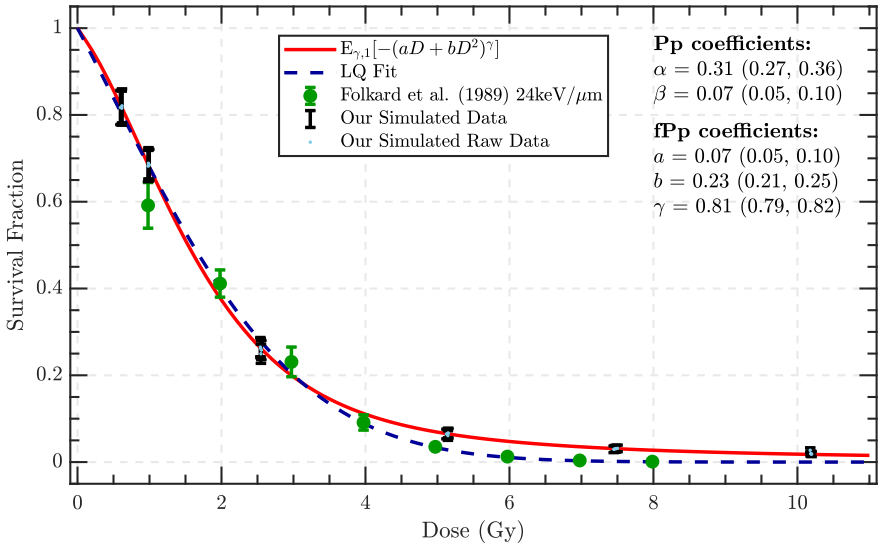


(b) Method 3

Figure L.12: Fits of the fPp (Mittag-Leffler) and Pp (LQ) models to our simulated survival data of V79 cells exposed to 20MeV (20keV/μm) proton radiation at 4.2mm depth in liquid water after 40 hours of repair time.

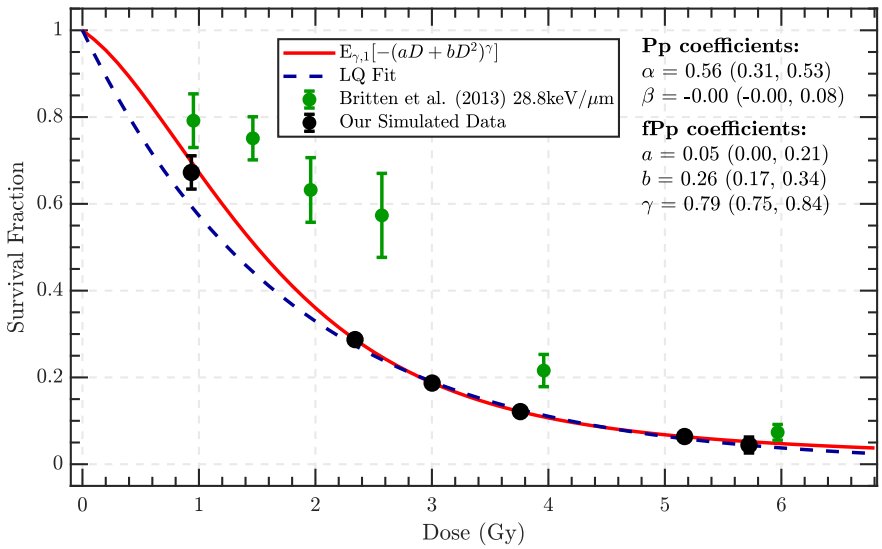


(a) Method 1

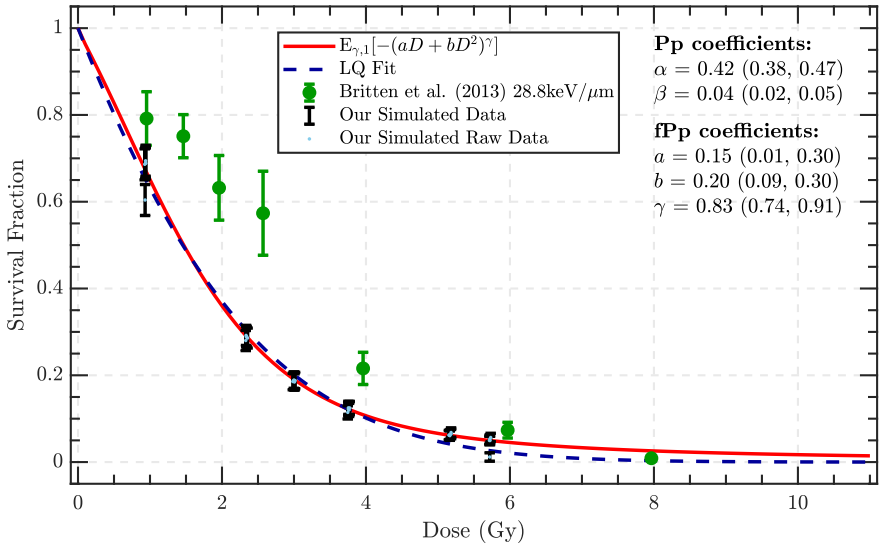


(b) Method 3

Figure L.13: Fits of the fPp (Mittag-Leffler) and Pp (LQ) models to our simulated survival data of V79 cells exposed to 20MeV (25keV/μm) proton radiation at 4.32mm depth in liquid water after 40 hours of repair time.



(a) Method 1



(b) Method 3

Figure L.14: Fits of the fPp (Mittag-Leffler) and Pp (LQ) models to our simulated survival data of V79 cells exposed to 20MeV (28keV/μm) proton radiation at 4.4mm depth in liquid water after 40 hours of repair time.

References

[1] "Australian Institute of Health and Welfare", Cancer in Australia, (2019).

[2] H. Paganetti, "Proton Beam Physics". CRC Press, (2012).

[3] H. Paganetti, "Proton Beam Therapy". IOP Publishing, (2017).

[4] Cancer Council Australia. "Treatment Methods: Chemotherapy".

[5] A. Nahum, "The Radiobiology of Hypofractionation," *Clinical Oncology*, **vol. 27**, no. 5, (2015).

[6] A. Pompos, M. Durante, and H. Choy, "Heavy Ions in Cancer Therapy," *JAMA Oncology*, **vol. 2**, no. 12, 1539–1540, (2016).

[7] C. Parker, V. Lewington, N. Shore *et al.*, "Targeted Alpha Therapy, an Emerging Class of Cancer Agents: A Review," *JAMA Oncology*, **vol. 4**, no. 12, 1765–1772, (2018).

[8] H. Blattmann, "Treatment Planning and Optimization for Pion Therapy", Boron Neutron Capture Therapy, (1992), 129–141.

[9] R. R. Wilson, "Radiological use of Fast Protons," *Radiology*, **vol. 47**, 487–491, (1946).

[10] J. Van der Veen, A. Laenen, and S. Nuyts, "Modern Radiotherapy Techniques versus Three-Dimensional Conformal Radiotherapy for Head and Neck Cancer," *Cochrane Database of Systematic Reviews*, no. 12, (2017).

[11] L. Marcu, E. Bezak, and B. Allen, "Biomedical Physics in Radiotherapy for Cancer". Springer, (2012).

[12] G. Stein and K. Luebbers, "Cancer : Prevention, Early Detection, Treatment and Recovery", 2nd ed. Hoboken, NJ: Wiley-Blackwell, (2020).

[13] K. Haume, S. Rosa, S. Grellet *et al.*, "Gold Nanoparticles for Cancer Radiotherapy: A Review," *Cancer Nanotechnology*, **vol. 7**, no. 8, (2016), doi: /10.1186/s12645-016-0021-x.

[14] R. Baskar, K. A. Lee, R. Yeo *et al.*, "Cancer and Radiation Therapy: Current Advances and Future Directions," *International Journal of Medical Sciences*, **vol. 9**, no. 3, 193–199, (2012).

[15] D. Schulz-Ertner and H. Tsujii, "Particle Radiation Therapy Using Proton and Heavier Ion Beams," *Journal of Clinical Oncology*, **vol. 25**, no. 8, 953–964, (2007).

[16] A. Chi, H. Chen, S. Wen *et al.*, "Comparison of Particle Beam Therapy and Stereotactic Body Radiotherapy for early stage Non-Small Cell Lung Cancer: A systematic review and hypothesis-generating meta-analysis," *Radiotherapy and Oncology*, **vol. 123**, no. 3, 346–354, (2017).

[17] C. Keppel, P. Ambrozewicz, and V. Nazaryan, "Beyond the Linear-Quadratic Model: Proton RBE Regression Analysis Studies, (2019)."

[18] R. Abolfath, C. R. Peeler, M. Newpower *et al.*, "A Model for Relative Biological Effectiveness of Therapeutic Proton Beams based on a global fit of Cell Survival Data," *Scientific Reports*, **vol. 7**, no. 1, (2017), doi: 10.3109/0284186X.2012.705892.

[19] A. Miyakawa, Y. Shibamoto, S. Otsuka *et al.*, "Applicability of the Linear–Quadratic Model to Single and Fractionated Radiotherapy schedules: an experimental study," *Journal of Radiation Research*, **vol. 55**, no. 3, 451–454, (2014).

[20] J. P. Kirkpatrick, D. J. Brenner, and C. G. Orton, "The Linear-Quadratic model is inappropriate to model High Dose per Fraction effects in Radiosurgery," *Medical Physics (Lancaster)*, **vol. 36**, no. 8, 3381–3384, (2009).

[21] H. Liu and J. Y. Chang, "Proton Therapy in Clinical Practice," *Chinese Journal of Cancer*, **vol. 30**, no. 5, 315–326, (2011).

[22] W. D. Newhauser and R. Zhang, "The Physics of Proton Therapy," *Physics in Medicine and Biology*, **vol. 60**, no. 8, 155–209, (2015).

[23] A. M. Kellerer and H. H. Rossi, "RBE and the Primary Mechanism of Radiation Action," *Radiation Research*, **vol. 47**, no. 1, 15–34, (1971).

[24] E. J. Hall, "Radiobiology for the Radiologist", 5th ed. Wolters Kluwer Health/Lippincott Williams & Wilkins, (2000).

[25] H. Paganetti, "Relative Biological Effectiveness (RBE) values for Proton Beam Therapy. Variations as a function of Biological Endpoint, Dose and Linear Energy Transfer," *Physics in Medicine and Biology*, **vol. 59**, no. 22, 419–472, (2014).

[26] F. Belli, R. Cera, M. Cherubini *et al.*, "RBE-LET relationships for cell inactivation and mutation induced by low energy protons in V79 cells: further results at the LNL facility," *International Journal of Radiation Biology*, **vol. 74**, no. 4, 501–509, (1998).

[27] F. Guan, L. Bronk, U. Titt *et al.*, "Spatial mapping of the Biologic Effectiveness of scanned particle beams: Towards biologically optimized particle therapy," *Scientific Reports*, **vol. 5**, (2015), doi: /10.1038/srep09850.

[28] H. Paganetti, "Proton Relative Biological Effectiveness – Uncertainties and Opportunities," *International Journal of Particle Therapy*, **vol. 5**, no. 1, 2–14, (2018).

[29] K. Prise, M. Folkard, S. Davies *et al.*, "The Irradiation of V79 Mammalian Cells by Protons with Energies below 2 MeV. Part II. Measurement of Oxygen Enhancement Ratios and DNA Damage," *International Journal of Radiation Biology*, **vol. 58**, no. 2, 261–277, (1990).

[30] S. J. McMahon, "The Linear-Quadratic Model: Usage, Interpretation and Challenges," *Physics in Medicine and Biology*, **vol. 64**, no. 1, (2018), doi: /10.1088/1361-6560/aaf26a.

[31] C. A. Tobias, "The Repair-Misrepair Model in Radiobiology: Comparison to Other Models," *Radiation research*, **vol. 104**, no. 2, 77–95, (1985).

[32] S. B. Curtis, "Lethal and Potentially Lethal Lesions Induced by Radiation - A Unified Repair Model," *Radiation Research*, **vol. 106**, no. 2, 252–270, (1986).

[33] R. Stewart, "Two-Lesion Kinetic Model of Double-Strand Break Rejoining and Cell Killing," *Radiation Research*, **vol. 156**, no. 4, 365–378, (2001).

[34] D. E. Lea, "Actions of Radiations on Living Cells". University Press, Cambridge, (1946).

[35] K. H. Chadwick, "A Molecular Theory of Cell Survival," *Physics in Medicine and Biology*, **vol. 18**, no. 1, 78–87, (1973).

[36] A. M. Kellerer and H. H. Rossi, "A Generalized Formulation of Dual Radiation Action," *Radiation Research*, **vol. 75**, no. 3, 471–488, (1978).

[37] F. Ballarini, "From DNA Radiation Damage to Cell Death: Theoretical Approaches," *Journal of Nucleic Acids*, **vol. 2010**, (2010), doi: 10.4061/2010/350608.

[38] B. G. Wouters, L. D. Skarsgard, L. E. Gerweck *et al.*, "Radiobiological Intercomparison of the 160 MeV and 230 MeV Proton Therapy Beams at the Harvard Cyclotron Laboratory and at Massachusetts General Hospital," *Radiation Research*, **vol. 183**, no. 2, 174–187, (2015).

[39] B. Emami, G. Woloschak, and S. Jr, "Beyond the Linear-Quadratic Model: Intraoperative Radiotherapy and Normal Tissue Tolerance," *Translational Cancer Research*, **vol. 4**, no. 2, 140–147, (2015).

[40] B. Fertil, I. Reydellet, and P. Deschavanne, "A Benchmark of Cell Survival Models Using Survival Curves for Human Cells after Completion of Repair of Potentially Lethal Damage," *Radiation Research*, **vol. 138**, no. 1, 61–69, (1994).

[41] D. J. Brenner, L. R. Hlatky, P. J. Hahnfeldt *et al.*, "A Convenient Extension of the Linear-Quadratic Model to Include Redistribution and Reoxygenation," *International Journal of Radiation Oncology, Biology, Physics*, **vol. 32**, no. 2, 379 – 390, (1995).

[42] J. Z. Wang, Z. Huang, S. S. Lo *et al.*, "A Generalized Linear-Quadratic Model for Radiosurgery, Stereotactic Body Radiation Therapy, and High–Dose Rate Brachytherapy," *Science Translational Medicine*, **vol. 2**, no. 39, 39–48, (2010).

[43] B. Douglas and J. Fowler, "The Effect of Multiple Small Doses of X Rays on Skin Reactions in the Mouse and a Basic Interpretation," *Radiation Research*, **vol. 178**, 25–38, (1976).

[44] M. Scholz, A. M. Kellerer, W. Kraft-Weyrather *et al.*, "Computation of Cell Survival in Heavy Ion Beams for therapy: The model and its approximation," *Radiation and Environmental Biophysics*, **vol. 36**, no. 1, 59–66, (1997).

[45] R. B. Hawkins, "A statistical theory of cell killing by radiation of varying linear energy transfer," *Radiation Research*, **vol. 140**, no. 3, 366–374, (1994).

[46] T. Friedrich, U. Scholz, T. Elsaäser *et al.*, "Systematic Analysis of RBE and Related Quantities using a Database of Cell Survival Experiments with Ion beam irradiation," *Journal of Radiation Research*, **vol. 54**, no. 3, 494–514, (2012).

[47] J. Perl, J. Shin, J. Schümann *et al.*, "TOPAS: An innovative proton Monte Carlo platform for research and clinical applications," *Medical physics (Lancaster)*, **vol. 39**, no. 11, 6818–6837, (2012).

[48] M. Douglass, E. Bezak, and S. Penfold, "Development of a Radiation Track Structure Clustering Algorithm for the Prediction of DNA DSB Yields and Radiation Induced Cell Death in Eukaryotic cells," *Physics in Medicine and Biology*, **vol. 60**, no. 8, 3217–3236, (2015).

[49] J. Forster, M. Douglass, W. Phillips *et al.*, "Stochastic Multicellular Modeling of X-ray Irradiation, DNA Damage Induction, DNA free-end Misrejoining and Cell Death," **vol. 9**, no. 1, 18888, (2019).

[50] S. Le Caër, "Water Radiolysis: Influence of Oxide Surfaces on H2 Production under Ionizing Radiation," *Water (Basel)*, **vol. 3**, no. 1, 235–253, (2011).

[51] S. P. Jackson, "Sensing and Repairing DNA Double-Strand Breaks," *Carcinogenesis*, **vol. 23**, no. 5, 687–696, (2002).

[52] Y. Tokuyama, Y. Furusawa, H. Ide *et al.*, "Role of Isolated and Clustered DNA Damage and the Post-Irradiating Repair Process in the Effects of Heavy Ion Beam Irradiation," *Journal of Radiation Research*, **vol. 56**, no. 3, 446–455, (2015).

[53] E. J. Hall, A. M. Kellerer, H. H. Rossi *et al.*, "The Relative Biological Effectiveness of 160 MeV protons—II Biological data and their interpretation in terms of microdosimetry," *International Journal of Radiation Oncology, Biology, Physics*, **vol. 4**, no. 11, 1009–1013, (1978).

[54] N. Chatterjee and G. C. Walker, "Mechanisms of DNA Damage, Repair, and Mutagenesis," *Environmental and molecular mutagenesis*, **vol. 58**, no. 5, 235–263, (2017).

[55] G.-M. Li, "Mechanisms and Functions of DNA Mismatch Repair," *Cell Research*, **vol. 18**, no. 1, 85–98, (2008).

[56] C. A. Waters, N. T. Strande, D. W. Wyatt *et al.*, "Nonhomologous End Joining: A good solution for bad ends," *DNA repair*, **vol. 17**, 39–51, (2014).

[57] J. C. Forster, M. J. J. Douglass, W. M. Phillips *et al.*, "Monte Carlo Simulation of the Oxygen Effect in DNA Damage Induction by Ionizing Radiation," *Radiation research*, **vol. 190**, no. 3, 248–261, (2018).

[58] K. Otani, Y. Naito, Y. Sakaguchi *et al.*, "Cell-cycle-controlled radiation therapy was effective for treating a murine malignant melanoma cell line in vitro and in vivo," *Scientific reports*, **vol. 6**, no. 1, 30689, (2016), doi: 10.1038/srep30689.

[59] G. Cheng, D. Kong, X. Hou *et al.*, "The Tumor Suppressor, p53, Contributes to Radiosensitivity of Lung Cancer Cells by Regulating Autophagy and Apoptosis," *Cancer Biotherapy and Radiopharmaceuticals*, **vol. 28**, no. 2, 153–159, (2013).

[60] D. Belli, P. Bettega, F. Calzolari *et al.*, "Inactivation of Human Normal and Tumour Cells Irradiated with Low Energy Protons," *International Journal of Radiation Biology*, **vol. 76**, no. 6, 831–839, (2000).

[61] D. Bettega, P. Calzolari, R. Marchesini *et al.*, "Inactivation of C3H10T1/2 cells by low energy Protons and Deuterons," *International Journal of Radiation Biology*, **vol. 73**, no. 3, 303–309, (1998).

[62] L. Bodgi, A. Canet, L. Pujo-Menjouet *et al.*, "Mathematical Models of Radiation Action on Living Cells: From the Target Theory to the Modern Approaches. A historical and critical review." *Journal of Theoretical Biology*, **vol. 58**, no. 394, 93–101, (2016).

[63] J. Crowther, "Some Considerations Relative to the Action of X-Rays on Tissue Cells," *Proceedings of The Royal Society of London. Series B*, **vol. 96**, 207–211, (1924).

[64] T. S. P. Strangeways and H. E. H. Oakley, "The immediate changes observed in tissue cells after exposure to soft x-rays while growing in vitro," *Proceedings of the Royal Society of London. Series B, Containing Papers of a Biological Character*, **vol. 95**, no. 669, 373–381, (1923).

[65] F. Holweck, "Production de rayons X monochromatiques de grande longueur d'onde. Action quantique sur les microbes," *Comptes Rendus Academy of Science*, **vol. 188**, no. 1, 197–199, (1929).

[66] P. Curie and M. Curie, "Sur l'étude des courbes de probabilité relatives à l'action des rayons X sur les bacilles," *Comptes rendus l'Académie des Sci*, **vol. 188**, no. 2, 202–204, (1929).

[67] M. M. Elkind and G. F. Whitmore., "In the Radiobiology of Cultured Mammalian Cells". Gordon & Breach Science Publishers, (1967).

[68] R. P. Virsik and D. Harder, "Statistical Interpretation of the Overdispersed Distribution of Radiation-Induced Dicentric Chromosome Aberrations at High LET," *Radiation Research*, **vol. 85**, no. 1, 13–23, (1981).

[69] I. Shuryak, B. D. Loucas, and M. N. Cornforth, "Straightening Beta: Overdispersion of Lethal Chromosome Aberrations following Radiotherapeutic Doses Leads to Terminal Linearity in the Alpha–Beta Model," *Frontiers in Oncology*, **vol. 7**, 318, (2017), doi: 10.3389/fonc.2017.00318.

[70] R. B. Hawkins, "A Statistical Theory of Cell Killing by Radiation of Varying Linear Energy Transfer," *Radiation Research*, **vol. 140**, no. 3, 366–374, (1994).

[71] T. Sato and Y. Furusawa, "Cell Survival Fraction Estimation Based on the Probability Densities of Domain and Cell Nucleus Specific Energies Using Improved Microdosimetric Kinetic Models," *Radiation Research*, **vol. 178**, no. 4, 341 – 356, (2012).

[72] M. Guerrero, R. D. Stewart, J. Z. Wang *et al.*, "Equivalence of the Linear–Quadratic and Two-Lesion Kinetic Models," *Physics in Medicine and Biology*, **vol. 47**, no. 17, 3197–3209, (2002).

[73] P. Andreo, "Monte Carlo simulations in Radiotherapy Dosimetry," *Radiation oncology (London, England)*, **vol. 13**, no. 1, 121–121, (2018).

[74] M. Asai and D. Wright, "Basics of Monte Carlo Simulation, (2019)."

[75] S. Agostinelli, J. Allison, K. Amako *et al.*, "Geant4—A Simulation Toolkit," *Nuclear instruments and Methods in Physics Research*, **vol. 506**, no. 3, 250–303, (2003).

[76] J.-F. Carrier, L. Archambault, L. Beaulieu *et al.*, "Validation of GEANT4, an object-oriented Monte Carlo toolkit, for simulations in medical physics," *Medical physics (Lancaster)*, **vol. 31**, no. 3, 484–492, (2004).

[77] D. Bolst, S. Guatelli, L. T. Tran *et al.*, "Validation of Geant4 for silicon microdosimetry in Heavy Ion Therapy," *Physics in Medicine and Biology*, **vol. 65**, no. 4, 045014, (2019), doi: 10.1088/1361-6560/ab586a.

[78] D. Peukert, S. Incerti, I. Kempson *et al.*, "Validation and investigation of reactive species yields of Geant4-DNA chemistry models," *Medical Physics (Lancaster)*, **vol. 46**, no. 2, 983–998, (2019).

[79] G. Battistoni, J. Bauer, T. Boehlen *et al.*, "The FLUKA code: An accurate simulation tool for particle therapy," *Frontiers in Oncology*, **vol. 6**, (2016), doi: /10.3389/fonc.2016.00116.

[80] V. Semenenko and R. Stewart, "A Fast Monte Carlo Algorithm to Simulate the Spectrum of DNA Damages Formed by Ionizing Radiation," *Radiation Research*, **vol. 161**, no. 4, 451–457, (2004).

[81] F. Salvat, J. Fernández-Varea, and J. Sempau, "Penelope. A code system for Monte Carlo simulation of Electron and Photon Transport," *NEA Data Bank, Workshop Proceeding, Barcelona*," (2001).

[82] D. Bolst, "The Geant4 Kernel," Geant4 School, Wollongong, Australia 2019," (2019).

[83] B. Faddegon, J. Ramos-Méndez, J. Schuemann *et al.*, "The TOPAS tool for particle simulation, a Monte Carlo simulation tool for Physics, Biology and Clinical Research," *Physica Medica*, **vol. 72**, 114–121, (2020).

[84] J. Schuemann, A. McNamara, J. Ramos *et al.*, "TOPAS-nBio: An Extension to the TOPAS Simulation Toolkit for Cellular and Sub-cellular Radiobiology," *Radiation Research*, **vol. 191**, no. 2, 125–138, (2019).

[85] "Low Energy Electromagnetic Physics - physics lists," https://geant4.web.cern.ch/node/1617, accessed: 2020-12-12.

[86] "Geant4-DNA: Extending the Geant4 Monte Carlo Simulation Toolkit for Radiobiology - Physics processes & models," http://geant4-dna.in2p3.fr/styled-3/styled-8/index.html, accessed: 2020-12-12.

[87] M. Douglass, "Development of an Integrated Stochastic Radiobiological Model for Electromagnetic Particle Interactions in a 4D Cellular Geometry [Ph.D Thesis]," *Adelaide (SA): University of Adelaide*," (2013).

[88] P. de Vera, I. Abril, and R. Garcia-Molina, "Excitation and Ionisation Cross-Sections in Condensed-Phase Biomaterials by electrons down to very low energy: Application to liquid water and genetic building blocks," *Physical Chemistry Chemical Physics*," (2020), doi: 10.1039/D0CP04951D.

[89] M. Douglass, E. Bezak, and S. Penfold, "Development of a Randomized 3D Cell Model for Monte Carlo Microdosimetry Simulations," *Medical Physics (Lancaster)*, **vol. 39**, no. 6, 3509–3519, (2012).

[90] H. L. Byrne, A. L. McNamara, W. Domanova *et al.*, "Radiation Damage on Sub-Cellular Scales: Beyond DNA," *Physics in Medicine and Biology*, **vol. 58**, no. 5, 1251–1267, (2013).

[91] D. W. Ross and H. C. Mel, "Growth Dynamics of Mitochondria in Synchronized Chinese Hamster Cells," *Biophysical Journal*, **vol. 12**, no. 11, 1562–1572, (1972).

[92] B. Alberts, D. Bray, J. Lewis *et al.*, "Molecular Biology of the Cell", 4th ed. Garland, (2002).

[93] R. E. Durand, "Oxygen Enhancement Ratio in V79 Spheroids," *Radiation Research*, **vol. 96**, no. 2, 322–334, (1983).

[94] S. Incerti, H. Seznec, M. Simon *et al.*, "Monte Carlo Microdosimetry for Targeted Irradiation of Individual Cells using a Microbeam Facility," *Radiation Protection Dosimetry*, **vol. 133**, 2–11, (2009).

[95] M. Šefl, S. Incerti, G. Papamichael *et al.*, "Calculation of Cellular S-values using Geant4-DNA: The effect of cell geometry," *Applied Radiation and Isotopes*, **vol. 104**, 113–123, (2015).

[96] M. Folkard, K. Prise, B. Vojnovic *et al.*, "The Irradiation of V79 Mammalian Cells by Protons with Energies below 2 MeV: Part I: Experimental Arrangement and Measurements of Cell Survival," *International Journal of Radiation Biology*, **vol. 56**, no. 3, 221–237, (1989).

[97] M. Dingfelder, M. Inokuti, and H. Paretzke, "Inelastic-Collision Cross-Sections of liquid water for interactions of energetic protons," *Radiation Physics and Chemistry*, **vol. 59**, 255–275, (2000).

[98] C. Champion, S. Incerti, H. Aouchiche *et al.*, "A free-parameter theoretical model for describing the electron elastic scattering in water in the Geant4 toolkit," *Radiation Physics and Chemistry*, **vol. 78**, no. 9, 745–750, (2009).

[99] E. Rutherford, "The Scattering of Alpha and Beta Particles by Matter and the Structure of the Atom," *Philosophical Magazine*, **vol. 21**, no. 6, 669–688, (1911).

[100] J. M. Heller, R. N. Hamm, R. D. Birkhoff *et al.*, "Collective oscillation in liquid water," *The Journal of Chemical Physics*, **vol. 60**, no. 9, 3483–3486, (1974).

[101] D. Emfietzoglou and H. Nikjoo, "The Effect of Model Approximations on Single-Collision Distributions of Low-Energy Electrons in Liquid Water," *Radiation Research*, **vol. 163**, 98–111, (2005).

[102] M. Dingfelder, D. Hantke, M. Inokuti *et al.*, "Electron inelastic-scattering cross sections in liquid water," *Radiation Physics and Chemistry*, **vol. 53**, no. 1, 1–18, (1999).

[103] M. E. Rudd, Y. K. Kim, D. H. Madison *et al.*, "Electron production in proton collisions: total cross sections," *Reviews of Modern Physics*, **vol. 57**, no. 4, 965–994, (1985).

[104] M. Karamitros, A. Mantero, S. Incerti *et al.*, "Modeling Radiation Chemistry in the Geant4 Toolkit," *Progress in Nuclear Science and Technology*, **vol. 2**, 503–508, (2011).

[105] R. Mueller, "Genome Biology and the Evolution of Cell-Size Diversity," *Cold Spring Harbor perspectives in Biology*, **vol. 7**, no. 11, (2015), doi: 10.1101/cshperspect.a019125.

[106] K. Bachmann, "Genome size in mammals," *Chromosoma*, **vol. 37**, no. 1, 85–93, (1972).

[107] A. Verkhovtsev, E. Surdutovich, and A. V. Solov'yov, "Multiscale approach predictions for biological outcomes in Ion-Beam Cancer Therapy," *Scientific Reports*, **vol. 6**, no. 1, (2016).

[108] N. Lewis, X. Liu, Y. Li *et al.*, "Genomic landscapes of Chinese hamster ovary cell lines as revealed by the Cricetulus griseus draft genome," *Nature Biotechnology*, **vol. 31**, no. 8, 759–765, (2013).

[109] S. Walenta and W. Mueller-Klieser, "Differential Superiority of Heavy Charged-Particle Irradiation to X-Rays: Studies on Biological Effectiveness and Side Effect Mechanisms in Multicellular Tumor and Normal Tissue Models," *Frontiers in oncology*, **vol. 6**, 30–30, (2016).

[110] B. L. Tracy, D. L. Stevens, D. T. Goodhead *et al.*, "Variation in RBE for Survival of V79-4 Cells as a Function of Alpha-Particle (Helium Ion) Energy," *Radiation Research*, **vol. 184**, no. 1, 33–45, (2015).

[111] W. Pohlit and I. R. Heyder, "The Shape of Dose-Survival Curves for Mammalian Cells and Repair of Potentially Lethal Damage Analyzed by Hypertonic Treatment," *Radiation Research*, **vol. 87**, no. 3, 613–634, (1981).

[112] S. Terashima, Y. Hosokawa, E. Tsuruga *et al.*, "Impact of Time Interval and Dose Rate on Cell Survival following Low-Dose Fractionated Exposures," *Journal of Radiation Research*, **vol. 58**, 1–9, (2017).

[113] B. G. Wouters, G. K. Y. Lam, U. Oelfke *et al.*, "Measurements of Relative Biological Effectiveness of the 70 MeV Proton Beam at TRIUMF Using Chinese Hamster V79 Cells and the High-Precision Cell Sorter Assay," *Radiation Research*, **vol. 146**, no. 2, 159–170, (1996).

[114] J. O. Rawlings, "Applied Regression Analysis: A Research Tool", 2nd ed., ser. Springer Texts in Statistics. New York, NY: Springer, (1998).

[115] T. S. Breusch and A. R. Pagan, "A Simple Test for Heteroscedasticity and Random Coefficient Variation," *Econometrica*, **vol. 47**, no. 5, 1287–1294, (1979).

[116] R. Koenker, "A note on Studentizing a Test for Heteroscedasticity," *Journal of Econometrics*, **vol. 17**, no. 1, 107–112, (1981).

[117] J. Durbin and G. S. Watson, "Testing for Series Correlation in Least Squares Regression. I," *Biometrika*, **vol. 37**, no. 3–4, 409–428, (1950).

[118] A. N. Kolmogorov, "Sulla Determinazione Empirica di una Legge di Distribuzione," *Giornale dell'Istituto Italiano degli Attuari*, **vol. 4**, 83–91, (1933).

[119] N. Smirnov, "Table for Estimating the Goodness of Fit of Empirical Distributions," *The Annals of Mathematical Statistics*, **vol. 19**, no. 2, 279–281, (1948).

[120] H. W. Lilliefors, "On the Kolmogorov-Smirnov Test for Normality with Mean and Variance Unknown," *Journal of the American Statistical Association*, **vol. 62**, no. 318, 399–402, (1967).

[121] T. W. Anderson and D. A. Darling, "A Test of Goodness of Fit," *Journal of the American Statistical Association*, **vol. 49**, no. 268, 765–769, (1954).

[122] M. Stephens, "Anderson-Darling Test for Goodness of Fit." *Encyclopedia of Statistical Sciences*, **vol. 1**, 81–85, (1982).

[123] M. Howard, C. Beltran, S. Anderson *et al.*, "Investigating Dependencies of Relative Biological Effectiveness for Proton Therapy in Cancer Cells," *International Journal of Particle Therapy*, **vol. 4**, 12–22, (2017).

[124] H.-J. Baek, T.-H. Kim, D. Shin *et al.*, "Radiobiological Characterization of Proton Beam at the National Cancer Center in Korea," *Journal of Radiation Research*, **vol. 49**, no. 5, 509–515, (2008).

[125] J. Gueulette, V. Grégoire, M. Octave-Prignot *et al.*, "Measurements of Radiobiological Effectiveness in the 85 MeV Proton Beam Produced at the Cyclotron CYCLONE of Louvain-la-Neuve, Belgium," *Radiation Research*, **vol. 145**, no. 1, 70–74, (1996).

[126] M. Folkard, K. Prise, B. Vojnovic *et al.*, "Inactivation of V79 cells by low-energy Protons, Deuterons and Helium-3 Ions," *International Journal of Radiation Oncology, Biology, Physics*, **vol. 69**, no. 6, 729–738, (1996).

[127] M. Belli, R. Cherubini, S. Finotto *et al.*, "RBE-LET Relationship for the Survival of V79 Cells Irradiated with Low Energy Protons," *International Journal of Radiation Biology*, **vol. 55**, no. 1, 93–104, (1989).

[128] G. Ljung and G. Box, "On a Measure of Lack of Fit in Time Series Models," *Biometrika*, **vol. 65**, no. 2, 297–303, (1978).

[129] K. J. White, "The Durbin-Watson Test for Autocorrelation in Nonlinear Models," *The Review of Economics and Statistics*, **vol. 74**, no. 2, 370–373, (1992).

[130] T. Tian, T. Inoue, H. Yamazaki *et al.*, "Comparison of Radiobiological Effective Depths in 65 MeV modulated Proton Beams," *British Journal of Cancer*, **vol. 76**, 220–225, (1997).

[131] Y. Furusawa, K. Fukutsu, M. Aoki *et al.*, "Inactivation of Aerobic and Hypoxic Cells from Three Different Cell Lines by Accelerated 3He-, 12C- and 20Ne-Ion Beams," *Radiation Research*, **vol. 154**, no. 5, 485 – 496, (2000).

[132] J.-H. Choi and J.-O. Kang, "Basics of Particle Therapy II: Relative Biological Effectiveness," *Radiation Oncology Journal*, **vol. 30**, no. 1, 1–13, (2012).

[133] K. Kagawa, M. Murakami, Y. Hishikawa *et al.*, "Preclinical biological assessment of proton and carbon ion beams at Hyogo Ion Beam Medical Center," *International Journal of Radiation Oncology, Biology, Physics*, **vol. 54**, 928–938, (2002).

[134] G. Last and M. Penrose, "Lectures on the Poisson Process", ser. Institute of Mathematical Statistics Textbooks. Cambridge University Press, (2017).

[135] W. Feller, "An Introduction to Probability Theory and Its Applications", 2nd ed. Wiley, (1968), **vol. 1**.

[136] D. R. Cox, "Regression Models and Life-Tables," *Journal of the Royal Statistical Society. Series B (Methodological)*, **vol. 34**, no. 2, 187–220, (1972).

[137] J. Hilbe, "Negative Binomial Regression". Cambridge: Cambridge University Press, (2007).

[138] X. Harrison, "Using observation-level random effects to model overdispersion in count data in Ecology and Evolution," *PeerJ*, **vol. 2**, 616, (2014), doi: 10.7717/peerj.616.

[139] R. Winkelmann, "Duration Dependence and Dispersion in Count-Data Models," *Journal of Business and Economic Statistics*, **vol. 13**, no. 4, 467–474, (1995).

[140] F. Mainardi, A. Mura, and G. Pagnini, "The M-Wright Function in Time-Fractional Diffusion Processes: A Tutorial Survey," *International Journal of Differential Equations*, **vol. 2010**, no. 1, 1–29, (2010), doi: 10.1155/2010/104505.

[141] R. N. Pillai, "On Mittag-Leffler functions and related distributions," *Annals of the Institute of Statistical Mathematics*, **vol. 42**, no. 1, 157–161, (1990).

[142] I. Podlubny, "Fractional differential equations: An Introduction to Fractional Derivatives, Fractional Differential Equations, to methods of their solution and some of their applications", ser. Mathematics in Science and Engineering. London: Academic Press, (1999).

[143] N. Laskin, "Fractional Poisson process," *Communications in Nonlinear Science and Numerical Simulation*, **vol. 8**, no. 3–4, 201–213, (2003).

[144] E. Orsingher and L. Beghin, "Fractional Diffusion Equations and Processes with Randomly Varying Time," *The Annals of Probability*, **vol. 37**, no. 1, 206–249, (2009).

[145] M. Kreer, A. Kızılersü, and A. Thomas, "Fractional Poisson processes and their representation by infinite systems of ordinary differential equations," *Statistics Probability Letters*, **vol. 84**, 27–32, (2014).

[146] R. Gorenflo and F. Mainardi, "Laplace-Laplace analysis of the fractional Poisson process," *S. Rogosin (Ed), Analytical Methods of Analysis and Differential Equations*, 43–58, (2012), [Kilbas Memorial Volume, AMADE 2011].

[147] F. Mainardi, "The Time Fractional Diffusion-Wave Equation," *Radiophysics and Quantum Electronics*, **vol. 38**, no. 1-2, 13–24, (1995).

[148] M. Meerschaert, E. Nane, and P. Vellaisamy, "The Fractional Poisson Process and the Inverse Stable Subordinator," *Electronic Journal of Probability*, **vol. 16**, 1600–1620, (2011).

[149] R. Garrappa, "Trapezoidal methods for fractional differential equations: Theoretical and computational aspects," *Mathematics and Computers in Simulation*, **vol. 110**, no. 1, 96–112, (2015).

[150] "Prescribing, Recording, and Reporting Proton-Beam Therapy", ser. ICRU Report. Bethesda, MD: ICRU, (2007), **vol. 7**.

[151] H. Nikjoo, P. O'Neill, M. Terrissol *et al.*, "Quantitative modelling of DNA damage using Monte Carlo track structure method," *Radiation and Environmental Biophysics*, **vol. 38**, no. 1, 31–38, (1999).

[152] H. Nikjoo, P. O'Neill, W. E. Wilson *et al.*, "Computational Approach for Determining the Spectrum of DNA Damage Induced by Ionizing Radiation," *Radiation Research*, **vol. 156**, no. 5, 577 – 583, (2001).

[153] D. T. Goodhead, "Radiation effects in living cells," *Canadian Journal of Physics*, **vol. 68**, no. 9, 872–886, (1990).

[154] M. Dingfelder, "Track Structure: time evolution from Physics to Chemistry," *Radiation Protection Dosimetry*, **vol. 122**, no. 1–4, 16–21, (2006).

[155] S. Incerti, G. Baldacchino, M. Bernal *et al.*, "The Geant4-DNA Project," *International Journal of Modeling Simulation and Scientific Computing*, **vol. 1**, 1–27, (2010).

[156] M. Rao, I. Iga, and S. K. Srivastava, "Ionization cross-sections for the production of positive ions from H_2O by electron impact," *Journal of Geophysical Research: Planets*, **vol. 100**, (1995), doi: 10.1029/95JE02314.

[157] M. A. Bolorizadeh and M. E. Rudd, "Angular and energy dependence of cross sections for ejection of electrons from water vapor. I. 50–2000-eV electron impact," *Physical Review A*, **vol. 33**, 882—887, (1986).

[158] N. Djurić, I. Čadež, and M. Kurepa, "H2O and D2O total ionization cross-sections by electron impact," *International Journal of Mass Spectrometry and Ion Processes*, **vol. 83**, no. 3, 7–10, (1988).

[159] M. Dingfelder, R. Ritchie, J. Turner *et al.*, "Comparisons of calculations with PARTRAC and NOREC: Transport of electrons in liquid water," *Radiation Research*, **vol. 169**, 584–594, (2008).

[160] A. Munoz, F. Blanco, G. Garcia *et al.*, "Single electron tracks in water vapour for energies below 100 eV," *International Journal of Mass Spectrometry*, **vol. 277**, no. 1-3, 175–179, (2008).

[161] L. H. Toburen, "Ionization and Charge-Transfer: Basic data for track structure calculations," *Radiation and Environmental Biophysics*, **vol. 37**, no. 4, 221–233, (1998).

[162] R. Dagnac, D. Blanc, and D. Molina, "A study on the collision of Hydrogen Ions H1+, H2+ and H3+ with a water-vapour target," *Journal of Physics B: Atomic and Molecular Physics*, **vol. 3**, no. 9, 1239, (2001).

[163] M. Dapor, I. Abril, P. de Vera *et al.*, "Simulation of the Secondary Electrons energy deposition produced by Proton Beams in PMMA: influence of the target Electronic Excitation description," *The European Physical Journal D*, **vol. 69**, no. 6, 165, (2015).

[164] W. Friedland, P. Jacob, P. Bernhardt *et al.*, "Simulation of DNA Damage after Proton Irradiation," *Radiation Research*, **vol. 159**, 401–410, (2003).

[165] D. Frankenberg, H. J. Brede, U. J. Schrewe *et al.*, "Induction of DNA Double-Strand Breaks by 1H and 4He Ions in Primary Human Skin Fibroblasts in the LET Range of 8 to 124 keV/μm," *Radiation Research*, **vol. 151**, no. 5, 540–, (1999).

[166] S. Meylan, S. Incerti, M. Karamitros *et al.*, "Simulation of early DNA damage after the irradiation of a fibroblast cell nucleus using Geant4-DNA," *Scientific Reports*, **vol. 7**, no. 1, 11 851–11 854, (2017).

[167] W. Chaung, L. Mi, and R. Boorstein, "The p53 status of Chinese hamster V79 cells frequently used for studies on DNA damage and DNA repair." *Nucleic Acids Research*, **vol. 25**, no. 5, 992–994, (1997).

[168] R. A. Britten, V. Nazaryan, L. K. Davis *et al.*, "Variations in the RBE for Cell Killing Along the Depth-Dose Profile of a Modulated Proton Therapy Beam," *Radiation Research*, **vol. 179**, no. 1, 21–28, (2013).

[169] D. Goodhead, M. Belli, A. Mill *et al.*, "Direct Comparison between Protons and Alpha-particles of the Same LET: I. Irradiation Methods and Inactivation of Asynchronous V79, HeLa and C3H 10T1/2 Cells," *International Journal of Radiation Biology*, **vol. 61**, 611–624, (2009).

[170] C. P. Karger and P. Peschke, "RBE and related modeling in Carbon-Ion therapy," *Physics in Medicine and Biology*, **vol. 63**, no. 1, (2017).

[171] L. Hanin and M. Zaider, "Cell-survival probability at large doses: An alternative to the Linear-Quadratic Model," *Physics in Medicine and Biology*, **vol. 55**, 4687–4702, (2010).

[172] H. Paganetti, "Biological Uncertainties in Proton (Ion) Therapy, (2019)," presentation.

[173] G. Neary, "Chromosome aberrations and the theory of rbe," *International Journal of Radiation Biology and Related Studies in Physics, Chemistry and Medicine*, **vol. 9**, no. 5, 477–502, (1965).

[174] K. Ando, Y. Furusawa, M. Suzuki *et al.*, "Relative Biological Effectiveness of the 235 MeV Proton Beams at the National Cancer Center Hospital East," *Journal of Radiation Research*, **vol. 42**, no. 1, 79–89, (2001).

[175] A. Antoccia, A. Sgura, F. Berardinelli *et al.*, "Cell cycle perturbations and genotoxic effects in human primary fibroblasts induced by low-energy Protons and X/Gamma-rays," *Journal of Radiation Research*, **vol. 50**, 457–468, (2009).

[176] D. Bettega, P. Calzolari, P. Chauvel *et al.*, "Radiobiological studies on the 65 MeV therapeutic Proton Beam at Nice using human tumour cells," *International Journal of Radiation Biology*, **vol. 76**, no. 10, 1297–1303, (2000).

[177] V. Calugaru, C. Nauraye, G. Noël *et al.*, "Radiobiological Characterization of Two Therapeutic Proton Beams With Different Initial Energy Spectra Used at the Institut Curie Proton Therapy Center in Orsay," *International Journal of Radiation Oncology, Biology, Physics*, **vol. 81**, 1136–1143, (2010).

[178] Y. Matsumoto, T. Matsuura, M. Wada *et al.*, "Enhanced Radiobiological Effects at the distal end of a clinical Proton Beam: In vitro study," *Journal of Radiation Research*, **vol. 55**, no. 4, 816–822, (2014).

[179] M. Miller and J. van den Brand, "Measurement of quasielastic ^{3}He(p, pN) scattering from polarized ^{3}He and the three-body ground state spin structure," *Physical Review Letters*, **vol. 74**, (1995).

[180] M. Raju, "A heavy particle comparative study Part III : OER and RBE," *The British Journal of Radiology*, **vol. 609**, no. 51, 712–719, (1979).

[181] J. Robertson, J. Eaddy, J. Archambeau *et al.*, "Relative Biological Effectiveness and Microdosimetry of a mixed energy field of protons up to 200 MeV," *Advances in Space Research*, **vol. 14**, no. 10, 271–275, (1994).

[182] R. L. Plackett, "Karl Pearson and the Chi-Squared Test," *International Statistical Review / Revue Internationale de Statistique*, **vol. 51**, no. 1, 59–72, (1983).