The long term impact of ionising radiation on living systems

Nathanael Lampe

N° d'Ordre : D.U. 2809

UNIVERSITE CLERMONT AUVERGNE

ECOLE DOCTORALE DES SCIENCES FONDAMENTALES
N° 909

THESE

présentée pour obtenir le grade de

DOCTEUR D UNIVERSITE

Spécialité : Physique

Par **LAMPE Nathanael**

B. Sci (Hons) B. Aero Eng (Hons)

The Long Term Impact of Ionising Radiation on Living Systems

De l impact à longue terme des radiations ionisantes sur les systèmes vivants

Soutenue publiquement le 5 mai 2017 devant la commission d examen.

Président : Dr Fabrice PIQUEMAL

Examinateurs : Dr Marie-Claude BORDAGE, Dr Marie DAVIDKOVA, Dr Marie DUTREIX, Dr Aleksandra RISTIC-FIRA

Directeur de thèse : Vincent BRETON

Co-Directeurs de thèse : Dr David SARRAMIA, Dr Sébastien INCERTI

Like many others have said, I couldn’t have seen this far if not for the giants upon who’s shoulders we all stand.

Comme on l’a dit, c’est seulement grâce aux géants qu’on regarde si loin.

前略、この文章は、日本語が読めるかどうかのテストです。

Abstract

All life on earth has adapted to an environment where there is a small, persistent, radiation background interacting with cells. Unlike evaluating the clearly harmful effects of high radiation doses, understanding the effects of this low persistent radiation dose on living systems is incredibly difficult. We have attempted to study whether background radiation is an important factor in evolution by conducting identical evolution experiments with *Escherichia coli* in the Clermont-Ferrand Particle Physics Laboratory and the Modane Underground Laboratory. Despite a 7.3 fold difference in the rate of interactions between the radiation background and cells between the two environments, no significant difference was found in the competitive fitness of the cell populations grown at each location. Using simulations, we showed that the rate at which ionising radiation interacts with cells is one hundred times less frequent than *E. coli*'s mutation rate in our experimental conditions, supporting the contention that natural radiation has no strong evolutionary effect. To further support this conclusion, we developed a mechanistic simulation for DNA damage as part of the Geant4-DNA project. Using this application, we irradiated a model of an *E. coli* genome, showing that for electron irradiation > 10 keV, the double strand break yield can be reasonably estimated to be between $0.006 - 0.010$ DSB Gy^{-1} Mbp^{-1}, depending upon the modelling of radical scavenging. This is in agreement with experimental data, further highlighting the small role natural ionising radation plays as a cause of mutations.

Résumé

La vie sur Terre s'est adaptée à un environnement où il y a un faible et persistent bruit de fond radiatif qui interagit avec les cellules. Loin des effets clairement nocifs des radiations à haute dose, il est difficile d'évaluer et de comprendre les impacts des faibles doses de la radioactivité naturelle sur les systèmes vivants. Nous avons tenté d'étudier si le bruit de fond radiatif est un facteur important dans l'évolution, en menant des expériences évolutives identiques avec *Escherichia coli* au Laboratoire de Physique Corpusculaire de Clermont-Ferrand, et au Laboratoire Souterrain de Modane. Malgré une différence d'un facteur 7,3 entre les taux d'interaction des rayonnements ionisants avec les cellules dans les deux laboratoires, aucune différence significative n'a pu être trouvée dans le fitness compétitif des populations cellulaires évoluées dans chaque laboratoire. Par simulation, nous avons montré que le taux d'interaction entre le bruit de fond radiatif et *E. coli* est cent fois plus faible que le taux de mutations d'origine endémique, ce qui renforce l'hypothèse que les radiations naturelles ont peu d'effet sur l'évolution. Dans le cadre du projet Geant4-DNA, nous avons développé une application complète de simulation mécanistique des dommages radio-induits à l'ADN, afin d'explorer davantage cette hypothèse. Avec cette application, on a irradié un modèle du génome d'*E. coli*, montrant que pour l'irradiation par des électrons d'énergies > 10 keV, le rendement des cassures double brin est de $0,006 - 0,010$ CDB Gy^{-1} Mbp^{-1}, selon le modèle de piégeage des radicaux chimiques. Ce résultat est en accord avec des données expérimentales, et souligne plus encore que les radiations ionisantes d'origine naturelle n'ont qu'une contribution mineure aux mutations responsables de l'évolution.

Table of contents

List of figures **xiii**

List of tables **xvii**

Nomenclature **xix**

Summary **1**

Synthèse **5**

1 Introduction **9**

1.1 A Brief Introduction to Low-Dose Radiation Biology 11

1.2 Biology in Underground Laboratories 14

1.2.1 The Laboratoire Souterrain de Modane 18

1.3 Evolution and Ionising Radiation 19

1.3.1 Some Basic Evolutionary Dynamics 21

1.3.2 a fitness landscape for bacteria 25

1.4 The Long Term Evolution Experiment 28

1.5 Simulating the Impact of Radiation on DNA 31

1.5.1 The Geant4-DNA Project 33

1.5.2 DNA Damage and the Radiolysis of Water 34

1.6 Putting it all together 36

2 Biological Experiments in Low Radiation Environments **39**

2.1 Radiation Environments in Biology 41

2.1.1 Radon in biological experiments 42

2.1.2 Potassium in biological systems 43

2.1.3 Activity measurements for biological media 44

2.2 Evaluating Dose in Different Environments 45

2.2.1 Simple dosimetry . . . 45
2.2.2 Dosimetry simulations . . . 47
2.2.3 Results . . . 51
2.2.4 Discussion . . . 52
2.2.5 Conclusions and Perspectives from Dosimetry . . . 55
2.3 The impact of radiation environments on cells . . . 55
2.3.1 Biological Conditions . . . 57
2.3.2 The Macroscopic Level . . . 57
2.3.3 The Microscopic Level . . . 58
2.3.4 Background sources and simulation parameters . . . 62
2.3.5 Results . . . 66
2.3.6 Discussion I: Physics . . . 70
2.3.7 Discussion II: Interpretation for an LTEE . . . 71
2.3.8 Discussion III: Implications for Low Background Experiments . . . 73
2.3.9 Perspectives . . . 74
2.4 Conclusions from Dosimetry and Microdosimetry . . . 75

3 Evaluating radiation damage with Geant4-DNA in bacteria 77
3.1 Requirements for the simulation . . . 79
3.2 Generating Geometries . . . 82
3.2.1 Generating Fractal Geometries . . . 84
3.2.2 Generating DNA Volumes . . . 88
3.3 Implementation of the molecular level DNA simulation . . . 90
3.3.1 Chromosome definition and structure . . . 90
3.3.2 Reading in geometries . . . 91
3.3.3 Implementation of Chemistry . . . 98
3.3.4 DNA Damage Model . . . 99
3.4 Model Parameter Studies . . . 104
3.4.1 Geometry for Parameter Studies . . . 105
3.4.2 Results . . . 105
3.4.3 Discussion of Parameter Sweep . . . 118
3.4.4 Reproducing Nikjoo et al., 1997 . . . 121
3.5 Simulation of an *E. coli* bacterium. . . . 127
3.5.1 Results . . . 129
3.5.2 Discussion . . . 133
3.6 Conclusions and Perspectives . . . 137

4 Long Term Evolution Experiments in Different Radiation Environments 139
4.1 Fitness Trajectories and Mutation Rates . . . 142
4.1.1 Mathematical Predictions . . . 143
4.1.2 Experimental Predictions . . . 144
4.1.3 Summary of hypotheses . . . 146
4.2 Method and Experimental Protocols . . . 147
4.2.1 Experimental Principle . . . 148
4.2.2 Preparation of solutions . . . 150
4.2.3 Starting the experiment from bacterial beads . . . 151
4.2.4 Using lines from a frozen microplate . . . 151
4.2.5 Daily Growth . . . 152
4.2.6 Freezing bacterial microplates for storage . . . 152
4.2.7 Fitness Assays . . . 153
4.3 Colony Counting With OpenCFU . . . 153
4.4 Results . . . 155
4.5 Discussion . . . 163
4.6 Perspectives for Future Experiments . . . 164

Conclusions and Perspectives 167

References 171

Appendix A DNA level simulation commands 185
A.1 Command Line Options . . . 185
A.2 Geometry Related Commands . . . 185
A.3 Damage Model Related Commands . . . 186
A.4 Analysis Related Commands . . . 187
A.5 Testing related commands . . . 188

Appendix B DNA level simulation class summary 189

Appendix C Model Parameter Sweep Tables 193

Appendix D Rejection of Data Points In Fitness Analysis 205

Appendix E P-values from Bootstrapped KS tests 209

List of figures

1.1 Radiation hormesis 12
1.2 the bystander effect 14
1.3 the LSM within the Fréyjus tunnel 19
1.4 Evolution viewed through fitness landscapes 22
1.5 Fitness landscapes and selection 22
1.6 Typical bacterial growth curve 25
1.7 Bacterial Fitness landscapes 27
1.8 Mutation accumulation across an LTEE 29
1.9 Discrete changes in fitness curves 30
1.10 Punctuations in the development of cell size 31
1.11 Schematic of the radiolysis of water 35

2.1 Cells traversed by a radiation track 40
2.2 Measured gamma spectra at the LSM and LPC 46
2.3 Simulation geometry for dosimetry 47
2.4 LSM γ spectrum 48
2.5 Binned LSM γ spectrum 48
2.6 ^{40}K β spectrum 50
2.7 Neutron energy spectrum 51
2.8 Electron interaction cross sections in Davis medium 59
2.9 Microdomains within the periodic domain 61
2.10 Implementation of the repeating boundary condition 63
2.11 Cosmic muon spectrum 66
2.12 Normalised cell energy depostions 69
2.13 Unnormalised cell energy depositions 69

3.1 Contribution of indirect damage on DNA strand breaks 78
3.2 Fractal structure of DNA 83

3.3 The *E. coli* macrostructure 84
3.4 Koch flake fractal built from an L-system 85
3.5 Sierpinski gasket built from an L-system 85
3.6 Co-ordinate system for 3D turtle graphics 86
3.7 Path of a 3D Hilbert curve 87
3.8 A 'voxelised' 2-D Hilbert curve 87
3.9 Molecule representation of DNA 89
3.10 Turned DNA placement elements 89
3.11 Schematic of chromosomes/ROIs 91
3.12 Cutting molecules to stop overlaps 92
3.13 DNA placed in Geant4 93
3.14 Base pair index counting in placement volumes 95
3.15 DNA volume sensitive to physical damage 100
3.16 Break likelihood and energy 101
3.17 Classifying Damage 103
3.18 Parameter Sweep Geometry 105
3.19 Impact of physical damage parameters I 106
3.20 Impact of physical damage parameters II 107
3.21 Comparison of break yield to Nikjoo et al. (1997) 108
3.22 Impact of physical damage parameters III 109
3.23 Impact of physical damage parameters IV 110
3.24 Killing distant radical tracks (breakdown by complexity) 112
3.25 Killing distant radical tracks (breakdown by source) 113
3.26 Diffusion and chemical reaction abundances 115
3.27 Killing distant radical tracks with limited diffusion 115
3.28 Impact of limiting diffusion on break yield 116
3.29 Impact of SSB efficiency 117
3.30 Saturation of indirect breaks with SSB efficiency 117
3.31 Comparison of Geant4-DNA break yields Nikjoo et al. (1997) 123
3.32 Strand break yields in the DNA hydration shell. 124
3.33 Changes in break types in the hydration shield 126
3.34 *E. coli* geometry used 128
3.35 Indirect and direct breaks 129
3.36 Strand break yields in *E. coli* from electrons 131
3.37 Strand break yields in *E. coli* from electrons 131
3.38 Distribution of strand break types in *E. coli* 134

4.1 Impact of population size on selection . . . 143
4.2 Fitness changes with mutation supply . . . 145
4.3 Microplate organisational scheme . . . 147
4.4 Bacterial growth curve during an LTEE . . . 148
4.5 Long term evolution experiments . . . 149
4.6 Fitness experiments . . . 150
4.7 The RGB and Lab colour schemes . . . 154
4.8 Colour recognition in OpenCFU . . . 155
4.9 LPC Fitness at 500 Generations . . . 156
4.10 Fitness measurements . . . 157
4.11 Mean fitness measurements . . . 158
4.12 Comparison of LPC and LSM fitness . . . 160
4.13 Fitness histograms between locations . . . 160
4.14 Fitness histograms between generations . . . 161

D.1 Fitness histograms at different times (without data rejection) . . . 206
D.2 Fitness histograms at different locations (without data rejection) . . . 206

E.1 Distribution of p-values at different locations . . . 210
E.2 p-value distributions at different generations . . . 211

List of tables

1.1 Recent results from underground laboratories 16

2.1 Underground and surface radiation sources 41
2.2 Activity of potassium in biological media 45
2.3 Gamma flux measurements in the LSM 49
2.4 Simulated Dose Rates in the LSM 52
2.5 Dose summary in the LSM 53
2.6 Elemental composition of Davis Medium 58
2.7 Impact of physics models on cell dosimetry 67
2.8 Interactions between background radiation and cells 68
2.9 Dose normalised cell/radiation interactions 70

3.1 Chemical reactions for DNA 98
3.2 Frequency of different energy deposits in DNA 114
3.3 Parameters required to reproduce Nikjoo et al. (1997) 122
3.4 SSB and DSB yields compared to Nikjoo et al. (1997) 122
3.5 Impact of a hydration shell 132

4.1 Mean Fitnesses 159

C.1 Direct damage parameter sweep (3 Å damage radius) 194
C.2 Direct damage parameter sweep (4 Å damage radius) 195
C.3 Direct damage parameter sweep (5 Å damage radius) 196
C.4 Direct damage parameter sweep (6 Å damage radius) 197
C.5 Direct damage parameter sweep (7 Å damage radius) 198
C.6 Direct damage parameter sweep (8 Å damage radius) 199
C.7 Direct damage parameter sweep (9 Å damage radius) 200
C.8 Direct damage using the default physics constructor 201
C.9 Direct damage using the default physics constructor 202

C.10 Damage Classifications by Source 203
C.11 Damage Classifications by Source after 1 ns 203
C.12 Break complexity (N97 parameter set) 203
C.13 Break source (N97 parameter set) 204

Nomenclature

Roman Symbols

D_c	Diffusion constant
d_{DSB}	Maximum distance between two base pairs that can cause a DSB.
d_s	Maximum distance between two damage sites before they are considered as separate.
l	Bacterial lag time
N	Population size
n	Population size
p_{BD}	Base damage induction probability
p_{SSB}	SSB induction probability
R	Amount of resources available to bacteria
r	Amount of resources consumed by bacteria
s	Mutation supply rate
s_r	Relative mutation supply rate

Greek Symbols

κ	Rate of consumption of resources per bacterium
μ	Mutation rate
μ_B	Beneficial Mutation rate
μ_g	Per genome spontaneous mutation rate

τ Bacterial lag time

Acronyms / Abbreviations

bp Base Pair (as a unit of DNA length).

CFU Colony Forming Unit

DM250 Davis Medium enriched with 250 mg mL^{-1} glucose

DSB Double Strand Break

ESS Evolutionarily Stable Strategy

gen Generation (of a cell)

GPx Glutathione peroxidase

GSSG-Rx Glutathione reductase

GST Glutathione transferase

IRT Independent Reaction Times

LB Low Background

LET Linear Energy Transfer

LNT Linear-No Threshold

LPC Laboratoire de Physique Corpusculaire de Clermont-Ferrand

LSM Laboratoire Souterrain de Modane

LTEE Long Term Evolution Experiment

MMS Methyl methanesulfonate

ROS Reactive Oxygen Species

SB Standard Background

SOD Superoxide dismutase

SSB Single Strand Break

TA Tetrazolium-Arabinose

Summary

Right now, there is about one cosmic muon passing through your thumb every second. When this happens, it's slowed down by all the atoms it is passing through, and causing it to lose about 2 MeV of energy. You don't notice this of course, because such a small amount of energy is inconsequential in all but the most finely calibrated systems.

This thesis asks how this small amount of energy, and the small amounts of energy like it that make up the natural background radiation on earth can impact the delicate machinery that is DNA. And more specifically, can radiation, as an external, abiotic factor, influence the evolution of life? That is to say, if evolution is the result of an accumulation of genetic mutations, and ionising radiation is a mutagen, is evolution to a measurable extent dependent upon natural ionising radiation.

Basing our work around an experiment designed to measure whether background radiation impacts the evolution of *Escherichia coli*, we address this question in three ways. First, through simulation, we consider how often ionising radiation from background sources actually impacts cells. Secondly, we try and quantify what happens when ionising radiation strikes DNA: how much DNA is damaged when a certain amount of energy is deposited in DNA, and how likely is it that this damage induces a break in the DNA chain, a precursor for mutations. These two stages help us to make predictions about what will happen to bacteria when the natural radiation background is suppressed in an evolution experiment, our third approach in tackling this question.

Evolution experiments provide a way of observing evolution in a repeatable way. From a given starting genome and selection pressure, an organism is likely to respond in only a limited set of ways over the first few thousand or so generations that it grows in these conditions. While mutations can occur across the entire genome, those that are favourable and accessible are often quite limited, permitting experimentalists to know with a certain probability, what paths evolution is likely to take. By studying controlled evolution in two slightly different conditions, the impact of a certain treatment on the evolutionary behaviour of a model organism can be found. In this work, we consider the outcome of an evolution experiment across 500 generations in *E. coli*, conducted in two different radiation

environments, a reference environment at the Clermont-Ferrand Particle Physics Laboratory, and a low radiation environment at the Modane Underground Laboratory.

In Chapter 2, we quantify these two radiation environments, and detail some of the considerations necessary in order to control the radiation background in a low radiation biological experiment. The dose rate in Clermont-Ferrand was calculated to be 214 nGy hr^{-1}, while underground, conducting our experiment inside lead shielding, a background level of 26 nGy $hr^{-1}$was achieved. The vast majority of the radiation background underground was due to the β^- decay of ^{40}K inside the nutritive medium used in our experiments. The impact of a radiation environment on individual cells though is not easily found from the dose alone. This is because dose is a macroscopic concept, that smooths out across a volume the very narrow paths traced by ions and subatomic particles as they travel through space.

Using the Geant4 simulation toolkit, we were able to quantify the magnitude of this effect, finding that the tracks of ionising radiation cross over bacterial cells with a very low frequency. For a given cell population grown in Clermont-Ferrand, ionising radiation from the background will touch cells with a frequency of 6.0×10^{-5} day^{-1} cell^{-1}, while underground this frequency drops to 8.2×10^{-6} day^{-1} cell^{-1}. The median energy deposited in DNA when these events occur is 140 eV. Compared to the daily mutation rate of *E. coli* cells in our experimental conditions, mutations from sources endemic to the cell must occur at least one hundred times more frequently than ionising radiation induced mutations. While this is an argument suggesting that ionising radiation should play a minimal role in evolutionary behaviour, it does not provide any link between the number of mutations induced in *E. coli* per day from the radiation background, nor give any measure of what it means for a cell to be hit by radiation.

To investigate this, we developed a complete mechanistic DNA damage simulation as part of the Geant4-DNA project. The Geant4-DNA project is a subset of the Geant4 project which provides track structure codes for liquid water, and simulates the radiolysis of water following irradiation. It includes a module for simulating radiation chemistry up to 1 µs after irradiation. Chapter 3 describes this application, and the geometry of an *E. coli* cell we developed for the simulation. The application was designed to be generic, so that other groups could also use it to produce mechanistic DNA damage simulations for a variety of cell types and DNA conformations. We found that our simulations produced results largely consistent with experimental measurements of the strand break yield in bacterial DNA. Depending on the assumptions regarding radical simulations made in the simulation, we predict that for most electrons interacting with *E. coli*, between 0.006 and 0.010 DSB Gy^{-1} Mbp^{-1} are caused in the cell.

Yields of double strand breaks are a first step towards determining the mutagenicity of

energy depositions in cells from ionising radiation. The misrepair of single and double strand breaks is one mechanism by which mutations can enter the genome. *E. coli* has a genome length of 4.6 Mbp, thus given the median energy deposit in DNA from radiation, each interaction between the radiation background and a DNA strand will cause $\approx 6 \times 10^{-3}$ single strand breaks, and $\approx 5 \times 10^{-4}$ double strand breaks (median values). Work remains to be done associating strand breaks as well as radical damage of bases with mutagenicity, but these numbers provide an indication already that the majority of interactions between cells and the radiation background cause comparatively little DNA damage.

Experimental validation of these predictions is important. Basing our work upon previous *E. coli* evolution experiments, we grew *E. coli* across 500 generations in two different radiation environments, measuring their competitive fitness in each location at regular intervals. Previously, the rate of change of fitness has been shown to be correlated with the mutation supply rate, a measure of how many mutations can enter the population each generation. The experiment was designed such that if radiation was significantly responsible for mutations, the 7.3-fold reduction in the rate that cells interact with the radiation background between the high and low radiation background environments would manifest itself in the rate of change of fitness within 500 generations.

We conducted detailed comparisons between the multiple lineages studies after 0, 200 and 500 generations of growth. At each interval, no significant difference in the distribution of fitnesses was able to be measured between the radiation environments considered. Thus under the conditions of our experiment, we conclude that after 500 generations, the natural radiation environment does not have a significant impact on evolution in *E. coli* that is measurable by competitive fitness. This is in agreement with our predictions from simulation, that the radiation environment should not significantly impact evolutionary behaviour.

Despite this, there remain many cellular outcomes observed in response to suppression of the radiation background that merit exploration in an underground laboratory. These include changes in the growth rate, gene expression and antioxidant regulation. An implication of this work is that these observations are not caused by interactions between radiation and DNA, but rather must come from how cells sense and respond to background ionising radiation. Exploring these responses, in order to quantify how cells understand, measure and respond to low radiation environments could reveal a rich tapestry of cellular processes as yet unknown or poorly understood, that are possibly driven by epigenetics or intercellular communication. For both human spaceflight, and for understanding living systems in high radiation zones and nuclear disaster zones, it is important that the basic responses of cells to background radiation are well known. Such research may even have a significance that carries over towards understanding some of the more perplexing observations that occur at low absorbed

doses, such as the bystander effect and genomic instability.

Given the many ideas brought together in this thesis, each chapter attempts to introduce enough relevant background material so that it can be understood on its own. The interested reader is directed to Chapter 2 for a detailed discussion of dosimetry in our experiments, and a quantification of how frequently ionising radiation touches cells. Chapter 3 presents a thorough consideration of the DNA damage that occurs when the natural radiation background interacts with *E. coli* cells, while Chapter 4 presents and explores our biological results.

Synthèse

Pendant que vous lisez ce document, environ un muon traverse votre pouce toutes les secondes. A l'intérieur de votre pouce, il est ralenti par tous les atomes qu'il rencontre, ce qui lui fait perdre environ 2 MeV d'énergie. Vous ne le sentez pas bien sûr, cette infime quantité d'énergie est invisible sauf aux systèmes les plus sensiblement calibrés.

Cette thèse étudie comment ces petits dépôts d'énergie, et les autres petites quantités d'énergie qui viennent du bruit de fond radiatif naturel sur Terre peuvent impacter l'objet délicat qu'est l'ADN. Et plus spécifiquement, les radiations ionisantes peuvent-elles, en tant que facteur abiotique externe à la vie, influencer l'évolution des êtres vivants. Autrement dit, si l'évolution est le résultat de l'accumulation des mutations génétiques, et si les radiations ionisantes sont mutagènes, est-ce que l'évolution dépend, à un niveau mesurable, de l'exposition à la radioactivité naturelle ?

A partir d'une expérience conçue pour mesurer si les radiations naturelles ionisantes impactent l'évolution de la bactérie *Escherichia coli*, nous étudions cette question de trois façons. D'abord, par simulation, nous considérons la fréquence avec laquelle les rayonnements ionisants du bruit de fond naturel interagissent avec les cellules. Deuxièmement, nous essayons de quantifier exactement l'effet des rayonnements ionisants sur l'ADN : quelle quantité d'énergie est déposée et quelle probabilité un brin d'ADN de se casser du fait de ce dépôt d'énergie. Les cassures de l'ADN contribuent à la quantification de la mort des cellules et à l'apparition de mutations. Les résultats obtenus au cours de ces deux étapes nous aident à prédire l'impact de la diminution du bruit de fond sur une expérience d'évolution, troisième volet de notre étude.

Une telle expérience nous offre le moyen d'observer l'évolution d'une manière répétable. A partir d'un génome de départ et d'une pression sélective choisie, il y a comparativement peu de façons dont un organisme peut répondre pendant les premières quelques milliers de générations de l'expérience. Même si les mutations peuvent arriver n'importe où dans le génome, celles qui sont favorables et accessibles sont plutôt limitées. Cela permet aux experimentateurs de connaître, avec une certaine probabilité, quels chemins l'évolution va prendre pour l'organisme étudié. Faire ces expériences en ne variant qu'un paramètre,

toutes choses égales par ailleurs, permet d'isoler l'effet de ce paramètre sur le comportement évolutionnaire de l'organisme modèle. Dans ce travail nous considérerons les résultats d'une expérience évolutive à travers 500 générations chez *E. coli*, menée dans deux environnements radiatifs différents : un environnement de référence, le Laboratoire de Physique Corpusculaire de Clermont-Ferrand, et un environnement où le niveau de radiation est réduit, le Laboratoire Souterrain de Modane.

Après une introduction, nous quantifierons dans le chapitre 2, ces deux environnements et détaillerons comment comprendre et analyser le bruit de fond radiatif d'une expérience biologique aux très faibles niveaux de radioactivité. Selon ces estimations, le taux de dose absorbée à Clermont-Ferrand est de 214 nGy hr^{-1}, tandis qu'au Laboratoire Souterrain de Modane, il est de 26 nGy hr^{-1}. Au Laboratoire Souterrain de Modane, la majorité du bruit de fond souterrain vient de la décroissance β^- du Potassium-40 dans le milieu nutritif utilisé pour nos expériences. L'impact de l'environnement radiatif sur les cellules individuelles n'est pas directement proportionnel à la dose absorbée. Effectivement, la dose absorbée est une quantité macroscopique, qui correspond à l'énergie déposée par les rayonnements ionisants dans un volume comparativement large par rapport au volume des bactéries. En réalité, cette énergie est vraiment déposée le long de traces très fines, ne concernant qu'une petite fraction des bactéries dans le milieu.

A l'aide de la boîte d'outils Geant4, nous avons pu quantifier la magnitude de cet effet. Pour une population donnée, mise en culture à Clermont-Ferrand, les rayonnements ionisants touchent les cellules avec une fréquence de 6.0×10^{-5} $jour^{-1}$ $cellule^{-1}$. Dans l'environnement souterrain du Laboratoire Souterrain de Modane, ce taux est réduit à 8.2×10^{-6} $jour^{-1}$ $cellule^{-1}$. Dans les deux cas, l'énergie déposée médiane est de 140 eV. En comparaison avec le taux de mutation d'*E. coli* dans notre expérience, les mutations d'origine endémique se produisent cent fois plus souvent que les mutations radio-induites. Cela confirme l'hypothèse que les radiations ionisantes ne devrait pas jouer un grand rôle dans le comportement évolutionnaire d'*E. coli*. Cependant, ce résultat ne nous dit rien sur ce qui se passe chaque fois qu'une bactérie est touché par un rayonnement.

Pour cela, nous avons développé une application complète pour simuler les processus mécanistiques impliqués dans l'induction des dommages à l'ADN à la suite d'une irradiation. Ce travail s'inscrit dans le projet Geant4-DNA, qui ajoute à Geant4 des codes pour modéliser les structures de traces dans l'eau liquide, ainsi que la radiolyse de l'eau et les réactions chimiques qui la suivent, jusqu'à 1 µs après l'irradiation. Le chapitre 3 décrit cette application et la géométrie d'une cellule d'*E. coli* développée pour cette étude. L'application a été conçue pour être générique, pour que d'autres groupes de recherche puissent l'utiliser après son intégration au projet Geant4. Les résultats de ces simulations sont largement en accord avec des

mesures expérimentales déjà faites des dommages radio-induits à l'ADN chez *E. coli*. Selon les paramètres choisis pour modéliser le piégeage des radicaux chimiques dans la cellule, nous prévoyons que la plupart des électrons causent entre 0.006 et 0.010 CDB Gy^{-1} Mbp^{-1}[1] quand ils interagissent avec une cellule d'*E. coli*.

L'estimation du taux de cassures double brin est un premier pas vers la quantification de la mutagénicité des dépôts d'énergie dans une cellule soumise à des radiations ionisantes. La mauvaise réparation des cassures simples et doubles des brins d'ADN est une source de mutations dans le génome. Le génome d'*E. coli* compte 4.6 Mbp. Ainsi, en prenant en compte la médiane de l'énergie déposée dans une bactérie, chaque interaction entre le bruit de fond radiatif naturel et une cellule va induire, en médiane, $\approx 6 \times 10^{-3}$ cassures simples, et $\approx 5 \times 10^{-4}$ cassures doubles. Il reste du travail à faire afin d'associer des cassures d'ADN, ainsi que les dommages radio-induits aux bases de l'ADN, à la mutagénicité. Néanmoins, ces chiffres indiquent que la plupart des interactions entre les radiations ionisantes et les cellules n'ont pas l'énergie suffisante pour être responsables de dommages significatifs à l'ADN.

La validation expérimentale de ces prédictions est importante. A partir d'une expérience déjà faite sur l'évolution expérimentale d'*E. coli*, nous avons cultivé *E. coli* dans deux environnements radiatifs différents, nous avons comparé les fitness compétitifs de l'ancêtre et des souches évoluées à intervalles réguliers. Des travaux antérieurs ont montré que le taux de variation du fitness est corrélé au taux d'offre de mutations, ce qui est une mesure de la quantité de mutations qui peuvent entrer dans la population à chaque génération. Nous avons conçu notre expérience de telle façon que si les mutations radio-induites contribuaient significativement à l'évolution, la réduction d'un facteur 7,3 du taux d'interaction entre les radiations naturelles et les cellules impacterait le taux de variation de fitness à 500 générations.

Nous avons fait des comparaisons détaillées entre plusieurs lignées étudiées après 0, 200 et 500 générations de croissance. A chaque étape, aucune différence significative dans la distribution des fitness n'a pu être mesurée entre les environnements considérés. Ainsi, dans les conditions de notre expérience, nous concluons qu'au bout de 500 générations, l'environnement radiatif n'a pas un impact sur l'évolution d'*E. coli* qui se manifeste dans le fitness compétitif. Cette observation est en accord avec nos prédictions de simulation ; l'environnement radiatif, au niveau naturel, ne devrait pas significativement impacter le comportement évolutionnaire de l'*E. coli*.

Malgré ces observations, des effets cellulaires ont été observés en réponse à la réduction de bruit de fond radiatif qui méritent une recherche plus approfondie en laboratoire souterrain. Des changements dans le taux de croissance de bactéries, l'expression des gènes et la régulation des antioxydants cellulaires apparaissent en réponse à l'introduction des cellules à

[1] CDB : cassures double brin. Mbp : 10^6 pairs de base.

un environnement souterrain. Ce travail indique que ces observations ne sont pas causées par des interactions entre les rayonnements ionisants et l'ADN, mais qu'elles doivent quand même provenir de l'interaction entre les cellules et le bruit de fond radiatif. L'exploration de ces réponses, afin de quantifier comment les cellules comprennent, mesurent et répondent aux bas bruits de fonds radiatifs pourrait révéler un ensemble de processus cellulaires qui sont inconnus, ou peu compris. Il y a une forte probabilité que l'épigénétique et/ou la communication intracellulaire soient nécessaire pour que les effets observés puissent exister. Pour la compréhension des systèmes vivants dans les zones fortement irradiés, dans le contexte de l'exploration de l'espace, ou d'un désastre nucléaire, une connaissance des réponses fondamentales de la vie au bruit de fond radiatif est important. Ces connaissances pourraient même apporter un éclairage nouveau sur les observations précédemment réalisées aux faibles doses absorbées, notamment l'effet « bystander » et l'instabilité génomique.

Vu les idées rassemblées par cette thèse, chaque chapitre est rédigé de façon à être lu de manière quasi-indépendante. Pour une discussion détaillée des mesures de dose faites pendant cette thèse et une présentation des informations nécessaires pour comprendre un environnement biologique à très bas bruit de fond, le lecteur est dirigé vers le chapitre 2. Le chapitre 3 offre une description de notre application pour simuler les dommages radio-induits à l'ADN, ainsi que les résultats de nos simulations. Nos expériences biologiques sont décrites dans le chapitre 4, où nous présentons aussi nos résultats. Le manuscrit se finit par une conclusion et des perspectives.

Chapter 1

Introduction

This thesis considers one simple question: does ambient ionising radiation have a measurable impact on the pace of evolution? Evolution occurs as a result of genetic changes in organisms over time, and these changes come from many origins, including ionising radiation. In a sense it is a *fait accompli* that ionising radiation has triggered genetic changes in organisms in the past, but here we are looking for reasoning more significant than a Drake Equation-esque proof affirming that yes, there is a high probability that radiation has played a role in evolution somewhere along the tape of life. Answering this question thoroughly requires a quantification of how frequently ionising radiation interacts with cells, damaging DNA and inducing mutations. Is this rate comparable or not with the mutation rate in most genomes, or is the mutation rate largely independent of ionising radiation? And beyond this, can we conduct an experiment that should clearly show whether or not radiation at natural levels has a measurable impact on evolution?

To get a feel for some of the numbers and quantities that are important in this work, we can quickly consider how often the natural radiation background impacts cells[1]. The typical radiation background on earth is 100 nGy hr^{-1}, or 6.2×10^{11} eV kg^{-1} hr^{-1}. Typically, something around 20 eV of energy can cause an ionisation that damages a DNA strand, so a background rate of DNA damage from the environment could reasonably be approximated around 3.1×10^{10} kg^{-1} hr^{-1}. Taking a typical human cell with a nucleus whose diameter is 6 μm, the frequency with which the nucleus would conceivably be damaged is then around 3.5×10^{-3} hr^{-1}, or once every ten days or so. By comparison, the spontaneous mutation rate per genome (μ_g, given per genome duplication) in human cells is around $\mu_g = 0.16$ (Drake et al., 1998).

In reality, ionising radiation deposits energy in cells along correlated tracks which acts to decrease the frequency with which cells are touched by radiation. The point however of

[1] In print, Katz (2016) has shown similar logic applied to an experiment involving *Deinococcus radiodurans*.

the above calculation is to show, via a back of the envelope' calculation, that at the background radiation level, ionising radiation does not interact with cells all that frequently. Certainly not on the time scale where it could be considered a frequent or continuous process. Cells in humans may divide anything from several times per day to less than a few times per month, so depending on the cell line, one could say that radiation touches cells with a frequency comparable to the mutation rate.

In microbes, regardless of the genome size, a spontaneous mutation rate of $\mu_g \approx 0.003$ is observed. Considering an *Eschirichia coli* bacterium, its 1 μm^3 volume means that at most it would see ionising radiation around 1.3×10^{-4} hr^{-1}. When looking at this in comparison to the mutation rate per unit time (μ) for *E. coli*, $\mu \approx 0.06$ hr^{-1} (c.f. Drake, 1991, considering a 20 minute division time), it becomes clear that in many systems, the majority of spontaneous mutations cannot come from radiation.

This begins to highlight the difficulty of conducting research on the impacts of the natural radiation background. Quantitatively, its effects are negligible in relation to other biological processes, but this overlooks how potentially damaging ionising radiation can be to living organisms. Despite its relatively low level, natural ionising radiation is still observed to have measurable biological effects. Radiation hormesis (Calabrese, 2013; Calabrese & Baldwin, 2003), the bystander effect (Morgan, 2003a,b) and transgenerational radiation sensitivity (Dubrova, 2003) are likely all symptoms of this simple biological dilemma: ionising radiation interacts with cells rarely, but cells must be alert to it due to its potentially devastating consequences.

In this work, we want to see whether ionising radiation at the background level has a measurable impact upon the pace of evolution. Strictly following the numbers, ionising radiation should hardly impact the speed of evolution at all, but as an agent capable of wreaking great biological damage in a small number of cells, it is possible that it makes accessible evolutionary pathways that would otherwise be closed to endogenously driven mutations.

In the following few sections, we will present a brief background of radiation biology in the low dose regime, before summarising the work that has been done studying the impact of removing natural radiation background by conducting biological experiments in underground laboratories. Next, we'll introduce work that has been done up to now linking evolution and ionising radiation. Here, we introduce some theoretical concepts that will be important to understanding evolution later in the manuscript. We'll also discuss controlled evolution experiments, and the Long Term Evolution Experiment (LTEE) led by Lenski et al. (1991), which are the experimental tool we will use to probe the pace of evolution in different radiation environments. Finally, we introduce the strategies we will use in this thesis to

explore how ionising radiation impacts evolution. In simulation, we seek to precisely quantify how frequently radiation at the natural level impacts cells. This knowledge is applied to our experiments probing evolution, led in the *Laboratoire Souterrain de Modane* (LSM) and *Laboratoire de Physique Corpusculaire de Clermont-Ferrand* (LPC).

1.1 A Brief Introduction to Low-Dose Radiation Biology

In the preceding section, the relatively low frequency with which ionising radiation interacts with cells at the natural background level was introduced. This introduces a stochastic element to predictions of radiation damage and risk at low dosages. At high dosages this disappears and models of cell survival and cell damage are accordingly much better defined. The Linear-No Threshold (LNT) model for example describes radiation risk well for radiation workers, meanwhile the Linear Quadratic model is well adapted for predicting experimental cell survival data. In particular, the linear regions of these models are built on an assumption that at low radiation doses, radiation damage accumulates linearly from microscopic causes. Implicit in this is that there is no safe radiation dose. At low radiation doses however, both these models overestimate radiation risk, as cells have developed multiple responses to low levels of radiation (there is active debate in the community regarding this subject, e.g. Cohen, 2012; Feinendegen, 2005; Little et al., 2009; Tubiana et al., 2009).

In particular, the idea is emerging that the radiation background has a stimulatory effect on cells, promoting oxidative resistance and possibly growth. This idea is called radiation hormesis (Figure 1.1). Hormesis is an important phenomena when considering a range of toxins, where small amounts of them can be beneficial or stimulatory, whilst large quantities remain harmful (e.g. Calabrese, 2013; Calabrese & Baldwin, 2003). The mechanisms by which hormesis acts are often difficult to pin down, especially in regards to ionising radiation, where cells must conceivably be alert to the impacts of environmental radiation, whilst passing multiple generations without seeing its impact. In general, two methods are available to cells which can enable this, cell signalling, where information about the radiation environment is somehow passed between cells, with the ensemble of cells in a population acting as a much larger detector of radiation than any one cell alone, and epigenetics, whereby cells are able to pass down information about the radiation environment to their descendants.

The low dose response of cells can be studied experimentally by looking at both individual cells, cell populations and multicellular organisms. When probing individuals, microbeams in particular allow unparalleled access to the interior workings of cells (Prise & Schettino, 2011). Single ions can be fired with micrometre level accuracy to target certain cellular regions, enabling the study of cell death and repair. Microbeams have been used when

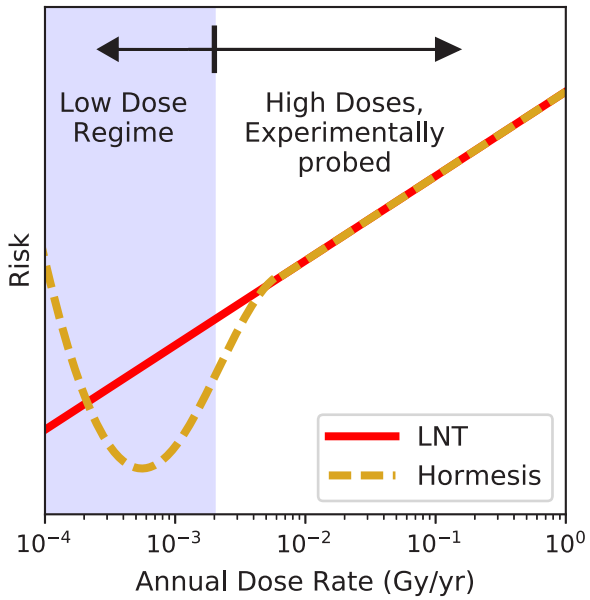


Fig. 1.1 The Linear-No Threshold (LNT) model, which describes the risk of cellular damage as linear with increasing dose is experimentally well validated for high doses, but is an extrapolation in the low-dose regime where biological responses are more difficult to probe. Alternative models, such as radiation hormesis, propose the hypothesis that small radiation doses are stimulatory and often beneficial.

studying tissues to provide evidence of bystander effects, where localised radiation dosages can be lethal to cells far from the location of the actual dose deposit (Figure 1.2), and have implicated not just DNA damage, but irradiation of the cytoplasm and mitochondria in the cell signalling that underpins the bystander effect (e.g. Zhou et al., 2008). They can also be used to explore epigenetic responses to low radiation dosages in larger organisms, in particular using the nematode *Caenorhabditis elegans* (e.g. Bertucci et al., 2009).

Epigenetics, the field which explores heritable traits that are not transmitted on the genome but rather in its regulation, is being seen as increasingly important in understanding radiation responses of organisms (Merrifield & Kovalchuk, 2013). Such changes are most visibly seen in the inheritable methylation of DNA strands, which changes the way they are read. In many cases where radiation leads to cell death, or cancer in larger organisms, it is difficult to mechanistically link a single radiation event with its later cause, as the two can be separated by months or even years, in the case of some cancers. The changes which may lead the descendant of an irradiated cell, many generations later, to become cancerous can be epigenetic in origin, carried by an inheritable genomic instability (Limoli et al., 1999).

The immediate environment around cells conspires with epigenetics to make the picture of low dose radiation effects yet more difficult to discern. It has already been said that cell signalling must be heavily implicated in observations of the radiation induced bystander effect. Numerous mechanisms have been proposed that can trigger cell death following the irradiation of neighbouring cells, from gap-junction transmitted information (Azzam et al., 2003) to reactive oxygen species (Mothersill & Seymour, 2004) and short RNa sequences (Ilnytskyy & Kovalchuk, 2011). Evidence for cell signalling as a mediator of the radiation response of cells has even been implicitly observed in population level studies of bacteria grown in underground laboratories (Castillo et al., 2015, 2016).

Detailed explanations of these mechanisms are beyond the scope of this introduction, but it is worthwhile to hypothesise why such responses might exist. There is no evidence that life has ever developed an ability to sense the radiation environment in a way that might encourage animals, plants or even bacteria to avoid high radiation regions. When cells are hit by ionising radiation, the damage that they succumb to is both far too localised, and far too abrupt to motivate a response from the organism to leave the environment. Instead, cells are required to, even in multicellular organisms, respond to radiation on an individual level, often via repair mechanisms (Steinhauser, 2015). On the other hand, cells do not see ionising radiation strike them very often - in the case of bacteria, cells could pass by tens or hundreds of generations without witnessing the stress of radiation damage. In this sense, when thinking about the low dose radiation response, cells are caught in a dilemma. Cells need to defend against the catastrophic impacts of radiation damage, whilst having

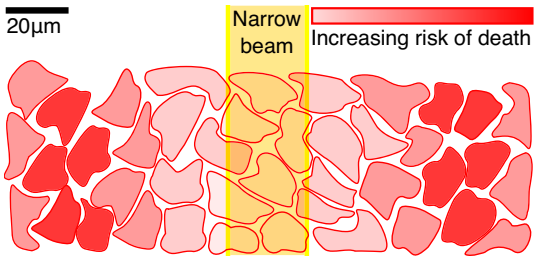


Fig. 1.2 The bystander effect describes a common experimental observation seen at low dosages, where a narrow radiation beam is focused on a cell medium, often a tissue. Paradoxically, in these experiments it is often cells far from the beam that are more likely to suffer cell death than those closest to the beam.

little exposure to radiation itself. The two ways cells can be aware of radiation though it infrequently affects them are by increasing either the volume in which they can be sensitive to radiation, or by increasing the time across which they can measure radiation.

It is plausible that the bystander effect, genetic instability and other epigenetic trends are a consequence of this dilemma. Epigenetics provides a mechanism by which cells can pass information to their descendants without a genetic change occurring. This increases the time over which cells can effectively sample the dose rate, and ensure they are primed to respond to radiation. Cell communication can theoretically increase the volume over which cells are sensitive to ionising radiation, again allowing cells to more effectively sample the background dose rate. A growing body of experiments conducted in underground laboratories hints that this contention is possible, however conclusive proof of this is difficult to obtain. Even in simple systems, the regulatory mechanisms that are invoked in these processes are difficult to quantify, and are likely very finely balanced.

1.2 Biology in Underground Laboratories

Low background radiation experiments typically study multiple cell populations treated in parallel, with equal numbers of cell lines grown at both a reduced radiation background, and a standard terrestrial background which serves as a control. The exact magnitude of the reduction in the radiation background varies between experiments, typically between 5- and 10-fold, once contributions to the background from ^{40}K in the nutritive medium have been considered. When comparing strains grown in different radiation environments, experimentalists typically target only a few measurements in order to gauge whether the cell populations have responded to the background. As ionising radiation manifests itself as oxidative cellular damage, these tests tend to analyse the presence and activity of proteins implicated in oxidative damage and stress. Mutation induction assays are frequently used,

as they allow the response of the cell to oxidative damage to be studied and for similar reasons, the response of cells to radiomimetic toxins is also studied. Finally, cell growth rate is also frequently measured, to see if the low background treatment of the cells has an impact on the performance and viability of the cell as a complete system. The majority of recent biological experiments conducted in underground and low background conditions has been summarised in Table 1.1, organised by organism type.

Under the hypothesis that removing ionising radiation from a growing cell population will reduce the need for scavengers of reactive oxygen species (ROS) in a cell, attention has been placed upon the superoxide dismutase (SOD) and catalase enzymes, as well as enzymes involved in glutathione regulation - glutathione peroxidase (GPx), glutathione transferase (GST), and glutathione reductase (GSSG-Rx). Growing V79 hamster cells in the Gran Sasso underground laboratory, Satta et al. (2002) showed that after 9 months culturing cells in both standard and reduced radiation background environments, catalase, GPx and GSSG-Rx were more dominant in the low background culture, whist SOD levels remained constant between the control and SB culture. Replicating the experiment across a ten month period though, Fratini et al. (2015) found equivalent SOD and catalase levels in the two populations, and significantly reduced levels of GPx. Across such a long duration experiment, it is foreseeable that culture ageing could have a stronger effect than the radiation background, explaining this discrepancy. The downregulation of GPx however in response to a reduced radiation background is supported by work in human TK6 cells, which have shown, over 6 months growth in a low background environment, a significantly decreased quantity of both GPx and catalase enzymes compared to cells grown at a standard radiation background level. Again here, the SOD abundance remained constant. In bacterial cells, qPCR analysis of *S. oneidensis* grown over 50 hours at the Waste Isolation Pilot Plant in New Mexico has shown that stress related genes, including those for catalase production, are up-regulated by exposure to low radiation environments (Castillo et al., 2015). The same experiments also found that exposure of *D. radiodurans* to reduced radiation backgrounds up-regulated the gene *dnaK*, responsible for the production of the heat shock protein HSP70. Intriguingly, the up-regulation of HSP70 has also been observed in bronchial epithelial cells and lung fibroblast cells as a result of growth at low backgrounds (Smith et al., 2011).

The number of mutants arising in cell populations following irradiation in low background biological experiments can diagnose whether the presence of the natural radiation background

[†]Satta et al. (1995)
[‡]Smith et al. (2011)
[§]Castillo et al. (2015)
[¶]Satta et al. (2002)
[‖]Fratini et al. (2015)
[**]Carbone et al. (2009)

Table 1.1 a selection of recent low background biological experiments, grouped by organism type. LB refers to the low background environment studied, and SB to the standard background used as a control.

Cell Type	Culturing	Experiment	Result
S. cerevisiae	120 gen. at LB 120 gen. at SB †	Mutation challenge from MMS	At high MMS concentrations, cells cultured in LB have impaired repair
D. radiodurans	75 hr at LB 75 hr at SB ‡	Cell growth rate Total Cell protein	Growth inhibited at LB compared to SB Proteins were reduced at LB
	50 hr at LB 50 hr at SB §	Cell growth rate qPCR	Reduced growth at LB compared to SB Upregulation of *dnaK* at LB
S. oneidensis	50 hr at LB 50 hr at SB§	Cell growth rate qPCR	Reduced growth at LB compared to SB Reaction of stress genes associated with exposure to UV and solar radiation to LB
V79 Chinese Hamster	9 mth at LB 9 mth at SB ¶	Growth curve Apoptosis following cyclohexamide exposure Antioxidant abundance Mutation induction after γ-irradiation	Growth unchanged between SB and LB. Increased apoptosis compared to control at LB and SB after 3 mth and 9 mth. Significantly increased apoptosis after 3 mth at LB compared to SB. Different modulation of antioxidant expression at LB compared to SB No increase in mutation induction at 3 mth compared to control, increased mutation induction relative to control and SB at 9 mth of LB.
	10 mth at LB then 6 mth at SB 16 mth at SB‖	Antioxidant activity Spontaneous mutation frequency	Downregulation of GPx activity in LB and upregulation of GPx activity in SB cells. Increased mutation frequency after 10 mth at LB, increasing further at 16 mth.
Human Bronchial Epithelial Human Lung Fibroblast	10 pass. at SB 10 pass. at LB‡	Protein expression	Upregulation of HSP 90B and HSP 70 in LB compared to SB
TK6 Lymphoblastoid	6 mth at LB 6 mth at SB**	Growth curve Micronuclei formation Antioxidant enzyme activity	No dependence on radiation environment More micronuclei formation in LB cells exposed to 2 Gy challenge compared to control. SB cells unchanged compared to control. Reduction in GPx and Catalase enzymes at LB compared to SB, no change in SOD abundance.

is important in the upkeep of biological processes related to DNA repair and prevention of oxidative damage. An increase in the number of mutants with the time spent in a low radiation level treatment relative to the control or standard background treatments would indicate that cells removed from the radiation background have lost some of their ability to resist oxidative damage, and suggests that the radiation background provides a stimulatory effect on these systems. Irradiating Chinese hamster V79 cells with up to 6 Gy of γ-radiation from a ^{137}Cs source, Satta et al. (2002) measured the number of mutants arising from mutations at the hypoxanthine-guanine phosphoribosyl transferase (*hprt*) locus, and found that after 3 months of culture in low and standard radiation background environments, no change in the number of mutants that appeared was apparent compared to a control measurement made at the beginning of the experiment. After 9 months however, the population grown at a lower background level showed a significant increase in the number of mutants that appeared, including a number of spontaneous mutants that arose even without irradiation. In a similar experiment, Fratini et al. (2015) found a higher rate of spontaneous *hprt* mutations after V79 had been cultured underground for 10 months. Upon reintroduction to a surface-level radiation environment, where the cells were cultured for another 6 months the number of spontaneous mutants increased again. This behaviour suggests a long term adaptive response to background radiation environments. Additionally, cells that had lost some capacity for repair and oxidative resistance were damaged by the higher oxidative stresses at the surface, to which they did not quickly habituate.

In human TK6 cells, the ability to resist and repair DNA damage was measured by subjecting cells to a 2 Gy dose of X-rays, and measuring the fraction of binucleated cells containing micronuclei following irradiation. Micronuclei formation is indicative of unrepaired or mis-repaired chromatin damage. Spontaneous micronuclei formation in populations of cells grown for 6 months in low and standard radiation backgrounds, and in a control population from the start of the experiment, shows little variation before irradiation, however after irradiation micronuclei formation is particularly elevated in the low radiation background population. This further supports the case for a drop in oxidative resistance following culturing of cells for extended time periods in low background environments.

The proportion of aberrant, damaged, or apoptopic cells in a population following exposure of cells to toxic agents can often serve as another indication of the ability of cells to recover from DNA damage. One study in the yeast *S. cerivisae* showed that cells grown at a low background in an underground laboratory for 120 generations showed a significantly lower ability to resist DNA damage than cells cultured at standard backgrounds in the same amount of time, when exposed to a high dose of methyl methanesulfonate (MMS), which induces DNA damage by stalling replication forks. In a later study, V79 cells were exposed

to cyclohexamide after 3 and 9 months at low and standard backgrounds. A significantly increased quantity of hypodiploid cells, indicative of eventual apoptosis, occurred in cells grown at both low and standard radiation backgrounds at both time points measured compared to the control sample (from zero culture time), however at 3 months the reduced background cells were significantly more likely to be hypodiploid than those grown in a standard radiation background. Echoing the results of past experiments, this supports the hypothesis that reduction of the ionising radiation background reduces the resistance of cells to stresses, though here this effect is possibly convoluted with a contribution from culture ageing.

Growth curves from cell cultures are an effective way to measure the impact of an environment upon a cell population. Planel et al. (1987) found the protozoa *P. tetraurelia* showed a marked decline in its generation time when grown at low backgrounds compared to the natural radiation level, whilst a stimulatory effect was observed upon growth when the radiation level was raised beyond the natural by growing cells at higher altitudes where cosmic radiation levels are elevated. This was replicated partially by Kawanishi et al. (2012), who, although unable to replicate inhibited cell growth immediately after cells underwent autogamy, did observe reduced growth rates in *P. tetraurelia* after it had grown at low backgrounds for 40 days.

Compelling evidence of reduced growth rates in cultures grown at low radiation backgrounds has been shown by Castillo et al. (2015) in bacteria, where both *S. oneidensis* and *D. radiodurans* exhibited reduced growth rates within 24 hr of being introduced to a shielded, underground, low radiation environment, compared to a parallel population grown in the same underground laboratory with a simulated natural radiation level environment. Additionally, the low background populations had lower maximum optical densities at the end of the exponential growth phase. It was also demonstrated in this work that the higher growth rate at the standard background level could be rapidly recovered by transferring the population grown at low background back to the standard background environment. Studies of growth rates in mammalian cells have not however indicated a clear difference in growth rate between cells grown at low and standard backgrounds. Neither TK6 cells (Carbone et al., 2009) nor V79 cells (Satta et al., 2002) showed a significant difference in doubling time after being cultured over months at different background radiation levels compared to the doubling time measured at the start of the experiment.

1.2.1 The Laboratoire Souterrain de Modane

This study is concerned with observing the evolutionary behaviour of bacterial cells across many generations in both underground and surface laboratories. This work is complementary to what has been shown so far in underground laboratories where a number of biological end

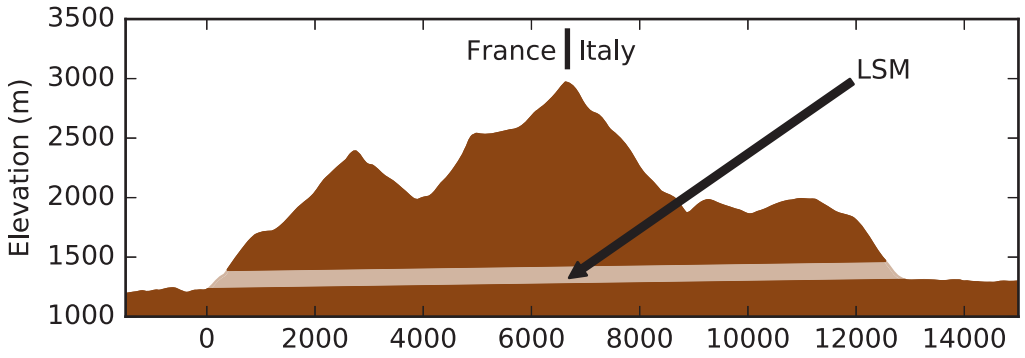


Fig. 1.3 The LSM is positioned in the Fréyjus road tunnel. The tunnel (light brown) runs through the Alps, under the French-Italian border.

points vary in response to the radiation environment, but no work has been done to show if the radiation environment causes this by directly impacting the genome. We assess this by looking, at first, at whether ionising radiation can impact evolution, which signals its role as both an evolutionary pressure and mutagenic agent that is significant at the background level.

We conducted our low background biological experiments in the *Laboratoire Souterrain de Modane* (LSM), located within the French Alps near the Italian border. The laboratory is built inside the Fréjus road tunnel near its deepest point, around 200 m west of the Fréjus peak. The lab itself is 1,700 m below the surface, and this rocky shield gives a radiation background equivalent to being 4,700 m under water (Figure 1.3). Beyond biological experiments, the laboratory hosts dark matter searches, neutrino science and is actively used for gamma ray spectroscopy.

1.3 Evolution and Ionising Radiation

Specific work has already been done trying to understand the impact of background radiation on evolution, though it has been hampered by the difficulties inherent in measuring what is typically a background effect. Two pistes have been explored in this research, the first considers whether higher levels of natural radiation can encourage extinction, and the second, more related to this work, explores the role of ^{40}K and artesian water as primordial gene irradiators.

Both the biodiversity present in the Chernobyl environment despite the elevated background, and the existence of radiation tolerant extremophiles on earth suggest significant increases in the radiation background don't prevent life, though they may hamper it. This hypothesis gathers some support from the fossil record, where both 62 Myr and 140 Myr

cycles are seen in fossil diversity (Rohde & Muller, 2005), which may be correlated with a 140 Myr cycle in the cosmic ray intensity on earth, caused by the solar system's passage through the spiral arms of the galaxy (Shaviv, 2002). This cycle does not greatly enhance the background cosmic ray flux above present levels however, estimates suggest that the cosmic ray background varies from 25% to 135% of current levels. It is very unlikely such a small signal impacts biodiversity, and the cyclic trends in fossil diversity may be due to other causes.

Potassium is an essential nutrient in biological systems. It is important in maintaining a correct electrolytic balance between cells and their environments, and across most cell types, the potassium is the main ion actively transported across the cell membrane. At present, radioactive ^{40}K makes up 0.012% of all natural potassium, having its origins in the formation of the solar system as a primordial radionuclide with a half life of 1.25×10^9 yr. Moore & Sastry (1982) hypothesised that the likelihood of mutagenesis from cellular ^{40}K was substantial given the yields of radicals produced from Auger electrons emitted following ^{40}K electron capture decay. Despite this, Gevertz et al. (1985) was unable to show, by measuring bacterial growth in the presence of varying ^{40}K levels, that Potassium has a mutagenic effect.

The idea of radiation as a mutagenic agent has merit beyond considering ^{40}K. Present high radiation environments are often dominated by a dose contribution from airborne Radon. ^{222}Rn and ^{220}Rn can be concentrated in groundwater supplies, and as α emitters, have a much larger biological effect than even the auger production of ^{40}K. The consequences of this were explored by Martell (1992), who suggested α emission could have a range of biological effects, especially early in earth's evolution when the primordial radionuclides ^{238}U, ^{235}U and ^{232}Th were significantly more abundant. The hypothesis though that these chemicals are essential, or at least important, in early evolution is difficult to test and has not largely been followed up.

It's to this end that we are conducting controlled evolution experiments in varied radiation environments. Controlled evolution experiments permit evolution to be studied in a controlled, repeatable and reproducible way. From the initial conditions of the experiment, the evolutionary behaviour of a species emerges as a dependent variable. Such experiments must take place over the long term, as evolution is a long term phenomenon. Specifically, we replicate the Long Term Evolution Experiment (LTEE, Lenski et al., 1991). The LTEE is a long duration experiment that has been measuring the evolution of *E. coli* for over 20 years in well defined experimental conditions. The very long baseline in time over which the experiment operates allows the evolution of the bacteria to be precisely studied.

1.3.1 Some Basic Evolutionary Dynamics

The dynamics of long term evolution experiments can become quite complex, and a few terms are necessary before tackling the idea. Two dominant ideas are going to be discussed here which illustrate and model the reasons species evolve in the ways they do. These are the fitness landscapes of Wright (1932), and the evolutionarily stable strategies (ESSs) of Maynard Smith & Price (1973).

Fitness landscapes were conceived before the discovery of DNA's structure by Watson & Crick (1953). Working under Mendelian genetics, Wright considered the enormous phenotypic variety that can arise when different alleles are combined. As an example, five genes, each being able to be in active or inactive states, can contribute to 32 (2^5) different genotypes. Expanding this, 1000 genes, each having ten different levels of activation means, to quote Wright's own understatement, "the number of possible combinations is 10^{1000}, *which is a very large number*"[8]. Linked of course to this is the fact that different gene combinations can lead two organisms having incredibly varied performances in different environments. As the genetic change that can occur between an organism and its children however is limited, the vast space of possible gene combinations is not well explored, and only small changes are permissible from one generation to the next. Nevertheless, the idea of a fitness landscape permits two somewhat quantitative parameters to be defined. The first is the fitness, and the second is a measure of the 'distance' between two genotypes. A rigourous definition of the fitness will be addressed later, but here, it can be thought of as a measurement of how well a genotype performs in an environment, with higher fitnesses corresponding to genotypes that are better adapted to an environment. Fitness as a function of genotype is an incredibly difficult parameter to quantify, given genotypes are overwhelmingly multi-dimensional. When considering fitness landscapes, simplifications are often made in favour of clarity, and the genome is simplified to only one or two dimensions, with genotypes that are only a few genes apart plotted close to each other, and more distant genomes drawn far away from each other.

A simple illustration of a fitness landscape is given in Figure 1.4, where a population inhabits a local maximum of the fitness landscape. The population, exchanging genes amongst itself could remain in the left local maxima indefinitely, but eventually, a spontaneous mutation would lead to genes that let the population migrate towards the second maxima. Such a change would be encouraged by high mutation rates, or large population sizes. Once in the second maxima, the population would continue to migrate towards the peak. In an environment where selection is weak, populations could exist in each peak, or between them. In an environment where selection is strong, it is likely that the population in the less fit peak

[8] Emphasis is mine.

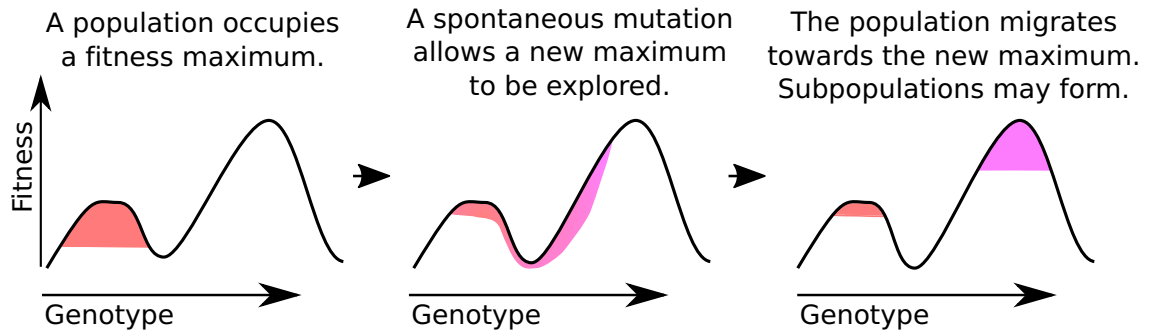


Fig. 1.4 Wright (1932) conjectured that evolution could be explained by fitness landscapes. As genotypes change, populations seek out local maxima of fitness. The quest to find different local maxima can lead to localised sub-populations of cells, or even speciation.

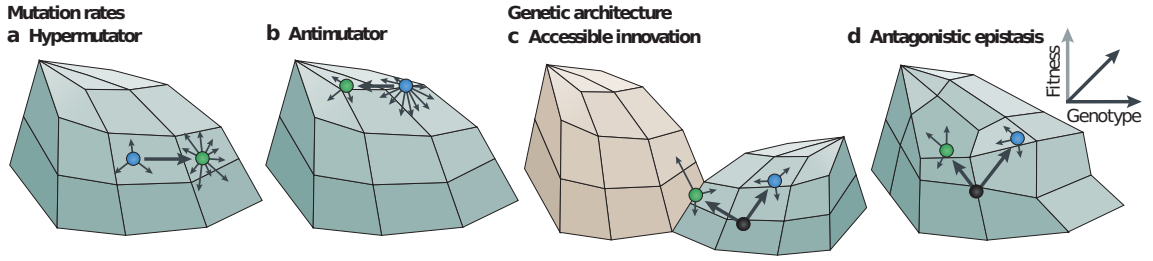


Fig. 1.5 Fitness landscapes can illustrate secondary aspects of selection (selection for traits that are not linked to an increase in fitness). Arrows around a point here represent the potential mutations from a point on the fitness landscape. Different mutation rates may arise that allow the landscape to be better explored (a), or encourage the species to remain near a peak fitness by discouraging further mutations (b). Some mutations may make accessible entirely new fitness spaces (c), or lead to antagonistic epistasis (d), where one mutation (in blue) limits further fitness gains more than another (in green). From Barrick & Lenski (2013).

would reach extinction (or at least, the genes responsible for that peak would become extinct in the population).

At a deeper level, fitness landscapes explain a wide variety of evolutionary phenomena, including the appearance of mutator genes, and rare innovations. This is illustrated in Figure 1.5 (Barrick & Lenski, 2013), which illustrates how one mutation may limit or promote the adaptation of a species. In relation to mutation rates, a single mutation can permit or deny to a genotype the ability to rapidly explore the environment. So called 'mutator genes', which substantially increase the mutation rate of a species, often by decreasing the efficiency of repair processes or the fidelity of gene transcription, can be favourable in certain instances, as they allow the fitness landscape to be explored rapidly. Often when they are of no more use, they will be selected against, as the advantages they offered in rapidly increasing the fitness disappear, outweighted by the increased cell death they can engender (as even simple mutations can have lethal consequences). Certain mutations can also open

up new parts of a fitness landscape, whilst others may close off avenues to further fitness increases, leading to evolutionary niches and cul-de-sacs.

Despite saying that fitness landscapes are complicated due to the multi-dimensionality of a genome, they can be simplified by considering fitness as a function of only two or three phenotypic properties. This is a common property of evolutionary algorithms, where a complex model is measured against a few performance criteria before being slightly modified, to see if the updated model parameter set is more advantageous than the precedent set. Niklas (1994) used this to great success to predict the most common structures of trees. Whilst the structure of a tree is complex, containing a potentially infinite number of branches, of varied lengths, at myriad positions, the success of a particular geometry is dependent upon a tree's ability to gather sunlight, spread seeds, and resist breaking. By considering a fitness landscape for trees, Niklas showed that a wide variety of tree-like geometries that can be observed in nature are reproducible by maximising fitness (the ability to gather sunlight, spread seeds and resist breaking). As this thesis is concerned with bacterial growth, in Section 1.3.2 we consider a simple fitness landscape for bacteria, a small application of these ideas.

ESS's provide a complementary method of understanding evolution, by addressing a danger in the interpretation of fitness landscapes whereby they may be perceived as an evolutionary march across a static landscape to the highest adaptive form available to an organism. Whilst the idea of a fitness landscape doesn't preclude a dynamic landscape, the changing dynamics of a competing population that alters its landscape as it evolves are fundamental with the ESS framework.

Before continuing, it's worthwhile to note the language that the fitness landscape and ESS frameworks use to describe themselves. A fitness landscape seeks to classify the fitness of a species as a function of either phenotype or genotype, depending on the study. The surface created by this highly multivariate fitness function is the fitness landscape. Changes that increase fitness are characterised as climbing up the fitness landscape, whilst detrimental changes move down the landscape. Populations tend to occupy peaks in the fitness landscape, as the consequence of climbing up the landscape is a migration of the population toward peaks. In many cases, the fitness landscape becomes metaphorical and statements can be made describing speciation.

While fitness landscapes may seem akin to an optimization problem, ESS's adopt the language of game theory. Here, evolution seeks strategies which win 'games'. A 'game' occurs any time to organisms interact. The strategy they choose defines how they interact. To illustrate this, we can think of simple behaviours animals may choose when they interact[9].

[9]these examples follow closely the explanations given by Maynard Smith & Price (1973) and Dawkins

Take for example a population of animals that may be genetically programmed to behave always as an *aggressor* or as a *pacifist*. Every time an *aggressor* meets another animal it will always attack, until the other animal is substantially wounded or dead. The *pacifist*, conversely, never attacks but will leave if the animal it encounters attacks. If a *pacifist* encouters another *pacifist*, one will eventually give up, after waiting a certain amount of time. In this game, we could assign a value to each outcome of the game. Winning the game, because your competitor is wounded or retreats, would yield a significant pay-off, let's say +50. Losing due to wound would have a large negative pay-off, suppose -100. Retreating instead of being attacked would naturally have no cost, and perhaps retreating after a long time competing would have a small negative pay-off, say -10.

In this scenario, if all the players are *aggressors*, the average pay-off each time to competitors meet would be very low, at -25. Even though a competitor wins half the time, receiving a pay-off of +50, when it loses it suffers greatly, with a pay-off of -100. A community of only *pacifists* however would have a much higher average pay-off of 20. Whenever two *pacifists* meet, they stare each other down under these rules until one leaves, suffering a penalty of -10, whilst the 'winner' takes the pay-off of +50. Whilst this may make it seem like evolution will favour the higher average pay-off strategy of the *pacifists*, this solution is not evolutionarily stable. This means that a population entirely composed of *aggressors* would easily be invaded by *pacifists*, who would preferentially always never suffer whilst the *aggressors* fought amongst themselves. Equivalently, in a population consisting entirely of *pacifists*, a mutation that caused the appearance of an *aggressor* would be favourable. Given the numbers we have chosen, there is an evolutionary stable number of *aggressors* and *pacifists*, but neither strategy, to be always aggressive or always passive, is favourable. What this example highlights is that the most competitive behaviour, and by extension phenotype, often changes as the population changes. The game theory model of ESS's provides a mechanism for exploring how shifting levels of a trait or behaviour in a population change the potential benefit of those traits.

Such an interpretation is not typically necessary to understanding the dynamics of our work, due to the sample studied and the short duration of the experiment. In different circumstances however, this can become very necessary, especially as time goes by and evolutionary niches are created. Over long time periods, complex behaviours can arise, including unstable oscillatory patterns where over time the population shifts from predominantly expressing one genotype to another. Computational modelling offers excellent advances in understanding these behaviours through the lens of game theory (Nowak & Sigmund, 2004).

(2006).

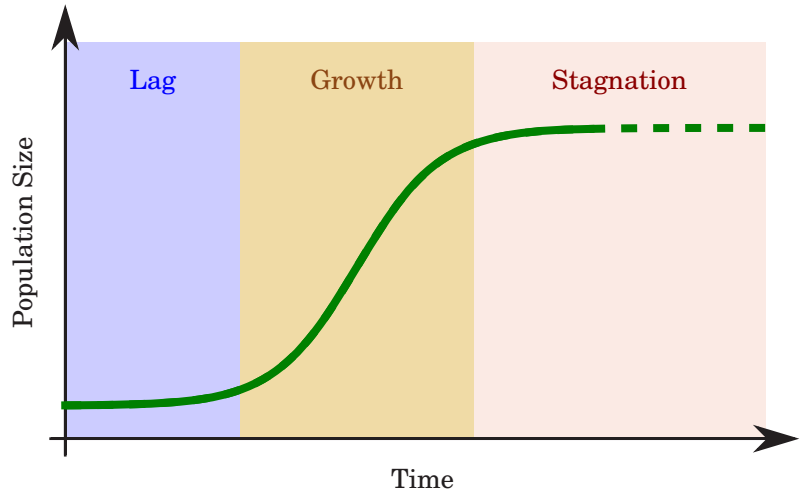


Fig. 1.6 Bacterial growth in a nutrient limited medium has three distinct phases: lag, in which bacteria adapt to the fresh nutritive levels; growth, a period of exponential growth; stagnation, where cells enter a stressed state as the medium is depleted of nutrients.

1.3.2 a fitness landscape for bacteria

Fitness landscapes can motivate which parameters should change in evolution experiments. Here, we consider for example how the parameters in a bacterial growth curve (similar to Figure 1.6) can lead to a fitness landscape, and how this can predict how selection should change those parameters. Bacterial growth can be described by three gross parameters, the lag time l a bacterial cell waits before division starts when it adapts to a new solution, the *e-growth time*[10], τ, and the rate at which a strain consumes nutrients κ. We consider a fourth parameter R_0, the resources available in the medium at the beginning of a growth cycle. Given this, a simple fitness landscape could be constructed by considering the amount bacteria grow before all resources disappear.

To derive this, we'll consider a general case of n competing lineages, before simplifying the derivation to the case of a single lineage. We define the set of latency times as $\mathscr{L} = \{l_1, l_2, \ldots, l_n\}$, the e-growth times as $\mathscr{T} = \{\tau_1, \tau_2, \ldots, \tau_n\}$ and the set of nutrient consumption rates as $\mathscr{K} = \{\kappa_1, \kappa_2, \ldots, \kappa_n\}$ for each competing strain. The number of bacteria in each strain is given by the set $\mathscr{N}(t) = \{n_1(t), n_2(t), \ldots n_n(t)\}$.

We model bacterial growth by an exponential curve following a stationary lag phase as follows for the m-th strain:

[10]This, the time it takes for a population to grow by e, is related to the doubling time, given by $\tau \ln(2)$

$$n_m(t) = \begin{cases} n_{m,0} & t < l_m \\ n_{m,0} \exp\left(\frac{t-l_m}{\tau_m}\right) & t \geq l_m. \end{cases} \tag{1.1}$$

Our aim is to find the fitness of each strain. Fitness is often expressed in a relative way, being the ratio of one strains growth rate to another's. The fitness of the m-th strain here can be given relative to the growth of the entire population as

$$F_{r,m} = \frac{\log\left(n_m(t_f)/n_m(0)\right)}{\log\left(n_{\text{tot}}(t_f)/n_{\text{tot}}(0)\right)}, \tag{1.2}$$

where t_f is the time when the level of nutrients reaches zero, and N_{tot} is shorthand for the total population, the sum of all members of $\mathscr{N}$.

The total amount of resources consumed by each strain, $r_m(t)$ can be modelled as follows, based on the rate of resource consumption being proportional to the number of cells.

$$\begin{aligned} \dot{r}_m(t) &= \kappa_m n_m(t) \\ r_m(t) &= \kappa_m \int_0^t n_m(t)\, dt. \end{aligned} \tag{1.3}$$

Given an ensemble of strains, the global nutrient level is given by the initial level of nutrients minus the sum of the individual consumptions, $r(t) = \sum_{m=0}^{n} r_m(t)$, and hence the global level of nutrients with time is

$$\begin{aligned} R(t) &= R_0 - \sum_{m=0}^{n} \int_0^t \kappa_m n_m(t)\, dt \\ &= R_0 - \sum_{m=0}^{n} \left(\int_{l_m}^t \kappa_m n_{m,0} \exp\left(\frac{t-l_m}{\tau_m}\right) dt + \int_0^{l_m} n_{m,0}\, dt \right) \\ &= R_0 - \sum_{m=0}^{n} n_{m,0} \kappa_m \tau_m \left(\exp\left(\frac{t-l_m}{\tau_m}\right) - 1 + \frac{l_m}{\tau_m} \right) \end{aligned} \tag{1.4}$$

Rearranging Equation 1.4 to find t_f, the time when the resources available drops to zero is non-trivial when multiple strains exist. The single strain case is however solvable, and is quite instructive, though it's worth noting that the presence of multiple strains can strongly impact the shape of the fitness landscape. Considering though only one strain, the subscripts

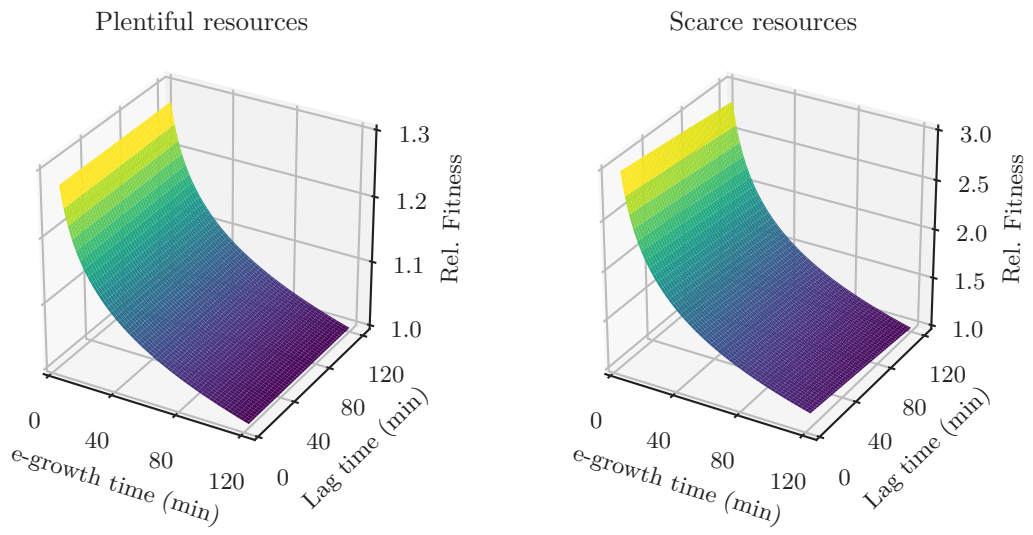


Fig. 1.7 Here we simulate a simple bacterial fitness landscape. When resources are plentiful (left), there is little incentive for a cell to adapt mutations that reduce its latency phase, however when resources are scarce, reductions in the latency phase can have important implications for fitness.

can be dropped and we find

$$R_0 = \kappa n_0 \tau \left(\exp\left(\frac{t \quad l}{\tau} \right) \quad 1 + l/\tau \right) \tag{1.5}$$

$$t_f = \tau \ln\left(\frac{R_0}{\kappa \tau n_0} + 1 \quad l/\tau \right) + l \tag{1.6}$$

leaving

$$n(t_f) = \frac{R_0}{\kappa \tau} + (1 \quad l/\tau) n_0. \tag{1.7}$$

In the single strain case, there isn't a candidate strain against which a relative fitness can be considered, so instead we consider just the numerator of Equation 1.2, which equates fitness to growth:

$$F_r = \log\left(\frac{R_0}{\kappa \tau n_0} + (1 \quad l/\tau) \right) \tag{1.8}$$

Noting that $R_0 > l n_0 \kappa$ so that $t_f > l$.

Equation 1.8 is a little hard to distinguish a general rule from, however the relative fitness ought to be inversely proportional to both the latency time and the e-growth rate. This is seen in Figure 1.7, which plots Equation 1.8, across a range of e-growth times and

lag phase durations. In the case of plentiful resources considered here, there are 200,000 times more resources than in the case of scarce resources, where the amount of resources is only a few times more than is necessary to survive a 120 min lag phase. In particular, when resources are scarce, selection favours more strongly reductions in the lag time than it does when resources are plentiful. Similarly, when resources are scarce, even small changes in the growth parameters can lead to rapid fitness gains, highlighting the impact that environmental factors can have on evolutionary behaviour.

These landscapes look incredibly simple, but they hide the real domain over which species evolve, namely the genotype. Bacteria cannot smoothly move across this landscape as the paths that are permitted by the mutations possible in a bacterium are limited. Mapping genotype to any single macroscopic parameter is an immense task, however fitness landscapes like the one shown above do permit us to rationalise the mutations observed in real life.

1.4 The Long Term Evolution Experiment

Having seen a small amount of the theory that motivates evolutionary biology, selection and competition, this section introduces the Long Term Evolution Experiment (LTEE) run out of Michigan State University by R. E. Lenski (Barrick et al., 2009; Lenski et al., 1991; Lenski & Travisano, 1994). The experiment is the longest running controlled evolution study, and offers an immensely detailed view of how the subject organism, *E. coli* evolves under fixed conditions. The LTEE is an ongoing experiment, having now grown 12 independent *E. coli* lineages for over 65,000 generations. In particular, it has been used to guide the development of models of competition in clonal populations (Gerrish & Lenski, 1998), study the appearance of mutator genes (Sniegowski et al., 1997), explore the links between fitness, mutations and phenotype changes (Elena et al., 1996), understand the targets of selection (Travisano & Lenski, 1996) and observe the variation in genome dynamics across multiple parallel populations (Barrick et al., 2009; Tenaillon et al., 2016).

The 12 independent lines grown by Lenski are subjected to a daily adaptation-growth-starvation cycle (as in Figure 1.6) in a nutrient limited medium. At the end of each cycle, the bacteria have grown by 6.6 generations. Over time, bacteria adapt to the daily growth cycle, which acts as an evolutionary pressure. As each lineage faces the same pressures, and the same conditions, the reproducibility of different evolutionary paths can be studied. Mutations arise randomly on genomes, and selection itself has a random component, but by repeating the same evolutionary initial conditions multiple times, trends can be observed, alongside which mutations are more favourable than others.

By making evolutionary studies reproducible, reproducing an LTEE-like experiment

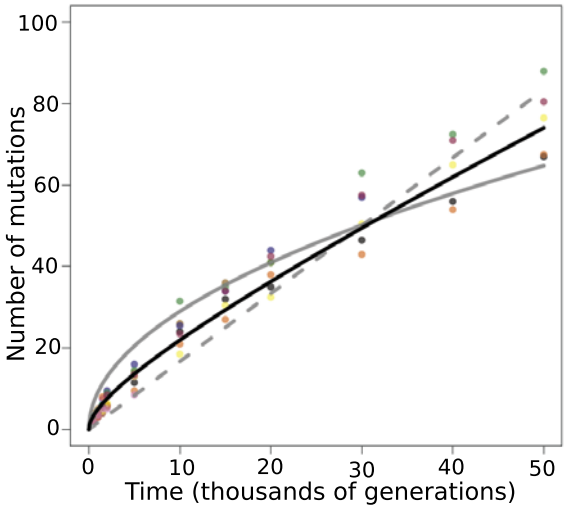


Fig. 1.8 Across an evolution experiment, both beneficial mutations and neutral mutations accumulate. Here, independent lineages with no mutator genes are shown with the number of mutations they accumulate as they adapt (a different symbol is used for each independent line). Neutral mutations accumulate linearly with time (dotted grey line), while beneficial mutations accumulate more slowly as time passes, and can be modelled as proportional to the square root of time (grey line). The black line best fits the data, by considering both beneficial and neutral mutations (number of mutations, is fitted via $n_m = at + b\sqrt{(t)}$). From Tenaillon et al. (2016).

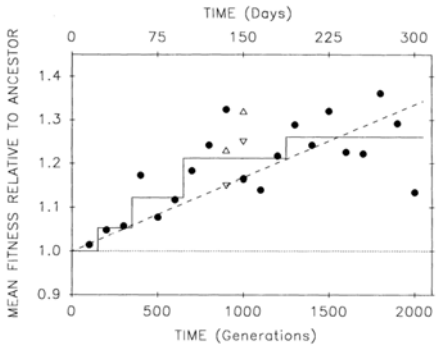


Fig. 1.9 Here, the fitness of an individual *E. coli* lineage is shown measured every 100 generations. A step-model was fitted to the data, showing that changes in fitness occur in discrete steps. This is due to the fixation of beneficial mutations in the populations. From Lenski et al. (1991).

allows us to measure the impact of natural radiation on the speed of evolution. Across the LTEE, many different endpoints have been observed, but for this work fitness is the most relevant. In the LTEE, the fitnesses were measured for every one of the 12 independent lineages at 500 generation intervals. Recent work has seen the genomes of the bacteria at these intervals sequenced, which allows the evolution of beneficial and neutral mutations to be observed with time. These studies have permitted a quantification of the number of mutations observed in *E. coli* with time. Two types of mutations generally become noticeable in these experiments, neutral and beneficial mutations. Overly harmful mutations tend to be suppressed as they are selected for only under specific circumstances[11]. Neutral mutations accumulate linearly in a population, whilst beneficial mutations accumulate most rapidly at the beginning of the fitness experiment, when the genome has the most potential to adapt (Figure 1.8). The relatively rapid beneficial mutation rate at the beginning of a fitness experiment can be used to determine whether different environments change the evolutionary behaviour of *E. coli* as here, changes in fitness are particularly sensitive to changes in the mutation rate (see also Section 4.1.2).

In the early stages of a fitness experiment, the population dynamics of a clonal population lead to a somewhat jagged development of the fitness of a bacterial population, linked to the 'fixation' of mutations in the population. Fixation denotes that a mutation has passed through the entire population[12]. Fixation of a gene can take place in 100 generations, and can manifest itself then as step like increases in the fitness of a population when the fitness

[11] Detrimental mutations, commonly mutator genes, can 'hitch-hike' to fixation in a population, when they occur alongside a beneficial mutation, by benefiting from the beneficial mutation whose genome they share.

[12] Interested readers here are referred to the article *Genome dynamics during experimental evolution*, by Barrick et al. (2009). In particular, Figure 2 in their paper aptly describes the different genome dynamics that arise in experimental evolution in clonal populations.

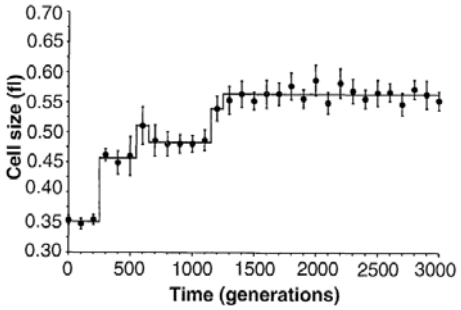


Fig. 1.10 In the early stages of controlled evolution experiments, sudden 'punctuations' are observed, where a phenotypic parameter changes rapidly in response to the rapid fixation of an advantageous mutation. Here, the size of *E. coli* cells shows a punctuated increase within the first 500 generations of the LTEE. From Elena et al. (1996).

is sampled at 100 generation intervals. The fitness of an individual *E. coli* line is shown in Figure 1.9. Steps in the growth of fitness are observable in the fit to the data made by Lenski et al. (1991), however it ought to be noted that a linear model can also explain the growth of fitness.

Rather, step like changes in biological parameters can be more easily seen in phenotypic parameters. Elena et al. (1996) showed the cell size, in particular, exhibits a step like change due to the fixation of cell membrane related genes in the studied population. This behaviour harkens back to the idea of punctuated equilibrium (explored by Eldredge & Gould, 1972; Gould & Eldredge, 1977, and equivalent in some ways to rapidly moving to a new part of a fitness landscape, or a new ESS becoming prevalent), that the fossil record shows speciation occurs in rapid jumps rather than gradual changes. Figure 1.10 shows an example of this at a simple level. The exploration of the genetic landscape that the LTEE bacteria underwent when they were placed in their new experimental conditions allowed them to access a new mutation with a significant benefit. This mutation rapidly swept through the population, causing the mean cell size to change in a binary, rather than gradual manner.

1.5 Simulating the Impact of Radiation on DNA

Any change in the mutation rate in response to the radiation environment must be caused by radiation interacting with cells. A large part of this work consists of modelling and understanding the impact of environmental radiation on cells. We perform this at both a dosimetric level and a DNA level. Using dosimetry, we can understand the frequency with which radiation interacts with cells, important in quantifying the impact of ionising radiation on individual cells rather than just the population. Simulations can also reveal

the mechanistic underpinnings of radiation damage, building a model that passes from energy depositions, to radiation damage, and finally to its biological effect.

The impact of radiation on DNA damage is often modelled in one of two ways. Commonly, a 'top-down' approach is used, where dosimetric measurements are made, and a relationship is inferred between dosimetry, radiation type and cell type to estimate the biological effectiveness of a radiation treatment. At a crude level, even equivalent dose is an example of this, by considering the elevated likelihood of biological damage following irradiation by high linear energy transfer (LET) radiation sources. Current trends favour the local effect model (Elsässer et al., 2008; Elsässer & Scholz, 2007; Scholz et al., 1997) as a 'top down' way of describing the biological damage caused by different radiation sources. Such models can even be combined with empirically derived repair models (e.g. McMahon et al., 2016), to provide a complete model of radiation induced biological damage, or enhanced to improve their predictive power for nanoparticle based treatment (Brown & Currell, 2017; Lechtman et al., 2013). In our work, such models are not appropriate as they are designed to work in the high dose regimes dealt with in cancer radiotherapy (> 1 Gy). To explore the impact of the radiation background on mutation rates, we need rather a model that is built from the 'bottom up', which mechanistically tries to model all the processes involved in cellular damage. In particular, most top-down models break when considering damage caused by low radiation dosages, as this is significantly beyond their domain of validity.

Modelling radiation-induced biological damage requires three elements, a Monte Carlo track structure code that is accurate at molecular resolutions, a model for radiochemistry and chemical reactions that can induce DNA damage, and a model for DNA. Monte Carlo codes capable of some or all of these functions have previously been reviewed by Nikjoo et al. (2006) and El Naqa et al. (2012). Most important in this is the modelling of physics at the molecular level. Many Monte Carlo codes are 'condensed history codes', which in order to improve performance, do not simulate every individual physical reaction a particle makes, permitting particles to move relatively large distances in a single computational step. Discrete track structure codes do not make this approximation, they attempt to simulate every interaction made by a tracked particle until it reaches thermal energies. From the physics stage, energy depositions from a transported particle in a track structure code can be correlated spatially, to give indications of biological damage, even in the absence of a geometrical DNA model (Francis et al., 2011), and similarly, analytic chemistry models can be used to replace a full simulation of radiochemistry to calculate biological damage (Liang et al., 2016).

There are two major existing platforms that combine the simulation of physics, chemistry and geometrical DNA models. These are the PARTRAC (Friedland et al., 2011) and KUR-

BUC (Nikjoo et al., 2016) platforms. This work extends the Geant4-DNA project (Bernal et al., 2015; Incerti et al., 2010), which already possesses a number of physics models and a chemistry module (Karamitros et al., 2011, 2014), so that generic DNA geometries may be simulated. Within these platforms, DNA damage must be modelled and evaluated. Such damage can typically be broken down into direct and indirect damage. Direct damage is caused in the physical stage of the simulation, and is linked to DNA molecules being directly impacted by radiation. Indirect damage occurs subsequent to this, and is caused by reactive oxidative species created through the radiolyis of water chemically reacting with DNA, thus damaging it and inducing strand breaks.

1.5.1 The Geant4-DNA Project

Geant4-DNA was developed to extend the Geant4 Monte Carlo simulation toolkit (Agostinelli et al., 2003; Allison et al., 2006, 2016) to molecular level simulations. It leverages the flexible architecture of Geant4 to allow users to conduct a wide variety of simulations at energies and scales relevant for microdosimetry, nanodosimetry, and at energies relevant in assessing biological damage. Geant4-DNA comprises three distinct sections for modelling the radiolysis of water. The physics stage comprises a choice of models for electron and ion track structures in liquid water, based on either the dielectric response function of liquid water (Emfietzoglou et al., 2005), or the CPA100 model set, notably derived by using the binary encounter Bethe approach to model ionisation (Bordage et al., 2016). The physico-chemical stage then occurs, where radical species and excited water molecules are created in the simulation. Finally, a chemical stage simulates the diffusion of these molecules and their chemical reactions.

Regardless of the physics model chosen in Geant4-DNA, all interactions between electrons and ions with their surrounding medium are discrete. That is to say that each step made by these molecules passing through the medium corresponds to an interaction with it, be it elastic scattering, ionisation or electron excitation. This permits the simulation to achieve a very high spatial accuracy, down to the nanometre scale, whilst the statistical nature of Monte Carlo simulations ensures quantum limits are not violated (Liljequist & Nikjoo, 2014). The physics modelling in Geant4-DNA takes place rapidly compared to other events, and is typically said to have terminated within 1 fs. The propagation of tracks takes place in liquid water, though validation data is frequently based upon data coming from the vapour or solid phases of water, due to the experimental difficulties of measuring interaction cross-sections in liquids. Particles are traced down to the energies where they are solvated or thermalised (around 8 eV), whereupon the physico-chemical stage commences.

The aim of the physico-chemical stage is to simulate the interactions that take place in

water between 1 fs and 1 ps after irradiation. Here, excited water molecules dissociate into new radical species, such as e^-_{aq}, $^\bullet OH$, H_2, H_3O^+ and $H^\bullet$. These radicals go on to interact amongst themselves to produce other species, notably OH^- and H_2O_2. It is not unusual for these molecules to form solvation cages, where water molecules orient themselves to surround the (often non-polar) radicals. This is notably the case for e^-_{aq}, which represents a solvated electron, that is to say an electron within a shell of bound water molecules.

Following the creation of radical species and the de-excitation of water molecules, radicals diffuse and interact. In Geant4, the processes simulated in the physico-chemical stage, as well as diffusion and reaction rates for the chemical stage are based on those used in PARTRAC, described by Friedland et al. (2011) and Kreipl et al. (2009). The chemical stage diffuses and reacts all radicals in the system simultaneously, across a number of time steps. Two time stepping algorithms have been implemented in Geant4-DNA. The first is a traditional method based upon using adaptive time steps, and solving the Smoluchowski diffusion equation to propagate molecules. Chemical reactions are realised when molecules are in close proximity to each other. A newer method is under development and has been used in this work based on the Independent Reaction Times (IRT) modelling approach (Green et al., 1990). This method offers a vastly accelerated simulation of chemical reactions and diffusion by the times of chemical reactions independently of their diffusion.

Geant4-DNA also contains a library of geometries for DNA level simulations provided through examples. In particular, a whole human cell has been modelled, filled with chromatin, (Dos Santos et al., 2013), and the capability of importing Protein Data Bank geometries has been developed (Delage et al., 2015). The toolkit allows for flexible modelling of geometries by building from the constructive solid geometry library within Geant4. Geometries have also be imported into Geant4-DNA simulations from the DnaFabric application, a tool for generating realistic human cell geometries (Meylan et al., 2016).

1.5.2 DNA Damage and the Radiolysis of Water

DNA damage is measured in simulations separately for the direct and indirect damage pathways. The direct pathways depend only on the physical stage of the simulation and attempt to model the likelihood that DNA molecules have been ionised or excited, and that this has lead to a strand break. Indirect pathways consider the chemical reactions between ROS and DNA that can induce a strand break. In both cases, the fundamental quantities considered are Single Strand Breaks (SSBs) and the Double Strand Breaks that occur when two SSBs are in close proximity to each other. Secondary parameters often considered are the amount of times bases are damaged in relation to the sugar and phosphate molecules that make up the backbone of the DNA chain.

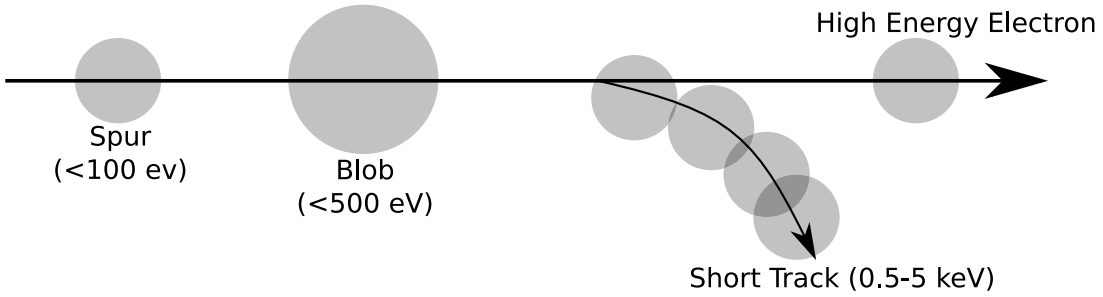


Fig. 1.11 a primary electron travelling through water creates radicals by first creating δ-rays, low energy electrons released following excitation or ionisation of water molecules. Low and medium energy δ-rays create blobs (r 7 nm), or more commonly spurs (r 4 nm), of radical species as the δ ray thermalizes in water. More energetic δ-rays can create short tracks, which themselves yield (often overlapping) blobs and spurs.

Before addressing SSB and DSB induction, it is worthwhile to take some time to explore the radiolysis of water, which at a basic level is important in understanding what is being modelled. When an electron passes through water, at high energies, its LET is of course low, and as it is slowed in the medium, its LET increases. All this time, the electron is losing energy by exciting or ionising atoms along its path, and this creates low energy secondary electrons, often called δ-rays. δ-rays bear the primary responsibility for the radiolysis of water and biological damage, as these electrons deposit energy in relatively confined volumes. Following an image first proposed by Samuel & Magee (1953), these energy deposits occur along the length of the path of a high energy electron as 'spurs'. δ-rays with a typical energy of $\lesssim 100$ eV have such a small path length in water that they tend to create small pockets of radicals called spurs (Figure 1.11). More energetic δ electrons can create larger blobs, or escape the trajectory of the mother track as their own short track.

The addition of DNA to a water medium complicates this picture, at the physics level, as it is difficult to model the interaction cross-section of DNA molecules in a way that is useful in a Monte Carlo simulation. Furthermore, a bulk cross-section doesn't always represent the dynamics of how excitation and ionisation occur in a single molecular chain. Electrons with energies of 5 eV, below the ionisation level of DNA (7.5 eV) have been shown experimentally to induce SSBs in DNA molecules (Boudaiffa et al., 2000), and theoretical models have shown that even less energetic electrons can induce breaks via shape-resonances (Barrios et al., 2002). Incident electrons of between 5-15 eV can also cause DNA strand breaks, typically by the dissociative excitation of phosphates and sugars, as well as molecular resonances, while more energetic electrons can induce strand breaks via direct electronic excitation of dissociative states (Orlando et al., 2008). Water molecules around the DNA can

also introduce new resonances into the system which lead to breakages.

Rather than simulate this level of detail in the Monte Carlo physics stage, the criteria used to determine whether a strand break has been induced in DNA is the accumulation of a threshold energy in a certain region around the DNA. Such an approach has statistical merit, whilst departing from the complicated reality of excitation and ionisation induced strand breaks. Additionally, as electrons are typically tracked down to 8 eV in Monte Carlo simulations, the small amounts of damage induced by low energy electrons are very difficult to model.

DNA damage by radicals near DNA by chemical processes relies on a smaller degree of approximation, in some respects, as it is based on well known chemical reactions, with defined chemical reaction rates. Some level of approximation does need to be made here, as the DNA geometry can block certain chemical reaction sites (Balasubramanian et al., 1998). SSBs resulting from chemical attack though can be easily found by analysing which chemical reactions are able to proceed in a simulation. In addition to explicitly specifying chemical reactions, it can be emulated by seeing with what likelihood relevant ROS species, notably $^{\bullet}$OH diffuse towards DNA molecules, and by counting diffusion into the DNA region as a chemical reaction (e.g. Nikjoo et al., 1997).

1.6 Putting it all together

This introduction has brought together many disparate threads across biology and physics. This thesis is really about replicating the Long Term Evolution Experiment in both a standard, above ground radiation environment, and the low background environment provided by the Modane Underground Laboratory. We spend a lot of time however trying to understand these two radiation environments as they apply to biology, in order to be sure that any changes we see between the two environments are able to be motivated by changes in the radiation environment.

Chapter 2 is intimately linked to this, exploring the many considerations that are necessary when conducting biological experiments in underground environments. We describe the dosimetry conducted in each environment, and the simulations we have performed in order to quantify the frequency with which cells in different radiation environments interact with the radiation background. Chapter 3 continues the theme of simulation, but now we try and quantify the rates of DNA damage in underground environments. Here, we present a new method for flexibly implementing a DNA geometry in Geant4, developed as part of the work of the Geant4-DNA collaboration. In particular, the method allows different geometries to be implemented, whilst taking away the burden of developing a scoring system for physical and

chemical DNA damage.

In Chapter 4 we address our biological experiments, explaining our protocols and motivating the experimental method we have chosen. We take as a starting point that the results of a controlled evolution experiment change when the environment changes in a way that evolution is sensitive to. To this end, our replication of the LTEE in both the LPC and the LSM is used to indicate whether evolution, in *E. coli* at least, is sensitive to the radiation background, or whether it is primarily driven by endogenous causes.

Chapter 2

Biological Experiments in Low Radiation Environments

This chapter addresses a fundamental question in low background radiation studies. What is the fundamental impact of different radiation environments on cells? Quantifying this is essential in designing and developing experiments at low dose, especially for quantifying the scale of damage and cellular disruption triggered by radiation compared to that which occurs due to cellular processes.

Radiation exposure is typically described in terms of absorbed and effective dosages. That is, a measurement of the amount of energy deposited in a target volume by weight. Effective doses introduce a scaling parameter that describes the likelihood of different particles to induce biological damage. At each of these levels, experimentalists withdraw slightly from the reality of radiation exposure in an effort to aid modelling. First, the conversion from an energy deposit by an ionising particle along a track that is discrete in space to an energy deposited in a volume obscures the spatially origin of radiation events. Secondly, conversion to effective dose converts the energy deposited into a new parameter that better describes a risk factor than an energy deposit. These parameters are both very valuable in their domains of applicability, but here we seek to understand both different radiation environments as well as their impact on individual cells. This dilemma is illustrated in Figure 2.1, where an incident electron travels through a field of cells. Only a few cells are traversed by the electron, and the energy deposition in each cell is likely far from uniform.

This chapter is divided into three sections, which present varying aspects of what is required to quantify radiation backgrounds in underground biological experiments. The first section provides an overview of common considerations for radiation environments in biological experiments. Unlike low background physics experiments where background events need to be almost eliminated, or restricted to a few events per day, the sensitivities

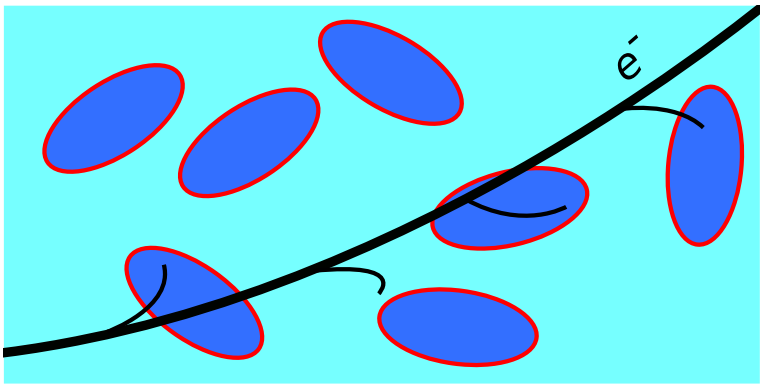


Fig. 2.1 From absorbed dose, ionising radiation deposits energy roughly uniformly in a volume. However, ionising radiation really deposits energy along a track. For small volumes, such as cells, when few particles traverse a volume, the approximation of a continuous energy deposit loses validity when considering the impact of radiation on single cells. In the above image, few cells, drawn as ellipsoids, are actually traversed by the passing electron, whilst the cell containing a black spur, representing a secondary electron, is likely to have a significantly higher energy deposited within it than other cells which were hit.

required for biology aren't as restrictive. This is largely a consequence of a biological experiment consisting of thousands to millions of cells, each of which can be regarded as a detector, whilst physics experiments typically have comparatively few detectors. In particular, for biological experiments, we are interested in seeing how many cells pass from seeing an energy deposit within them due to radiation in a given time period, to seeing no energy deposit at all in the same time period within their cell wall.

We begin this chapter by discussing different radiation environments in biology, and some general principles that need to be considered when designing an experiment. This discussion, presented in Section 2.1, focuses upon the impact of ^{40}K and ^{222}Rn in biological experiments at low background. In particular, we note here that the radiation dose coming from the biological media used in the experiment is important in assessing the absorbed dose for a cell sample, and to this end we consider the activity of a range of commonly used growth media.

A dosimetric investigation of the different environments we consider is presented in Section 2.2. As previously mentioned, absorbed doses paint however an incomplete picture of how a radiation environment affects cells, especially when considering the long term effects of radiation where the key parameter concerned is the chance that an individual cell is intersected by the track made by an ionising particle. In the experiments we conduct, this is particularly relevant as individual cells may pass many generations without 'seeing' any radiation. For the different radiation environments considered in our experiments, we have conducted an investigation into the frequency with which ionising radiation actually

impacts *E. coli* cells in Section 2.3. These events occur stochastically, and understanding their frequency is important in predicting the physical consequences upon cells of changing the radiation environment, in a far more complete way than dosimetric measures can.

2.1 Radiation Environments in Biology

Conducting our experiments in physics laboratories, work has already been done to classify the sources of radiation present, especially in the LSM. These measurements have been made though with physics experiments in mind. In low-background physics experiments, a single α decay may, for example, lead to a false detection and bias an experiment. When conducting biological experiments, this level of sensitivity isn't necessary as trace quantities of radio-elements in biological growth media contribute to a background that limits the maximum possible dose reduction.

Keeping this in mind, Table 2.1 presents a brief overview of the relevant radiation background sources in surface and underground radiation environments. In addition to radio-elements in the nutritive medium, at the surface, biological systems are affected by terrestrial radiation sources and the cosmic background. In underground laboratories, the cosmic component of the background is greatly reduced, though the quantity that does penetrate the laboratory is still significant enough to affect sensitive experiments.

For the LSM, a small amount of cosmic muons penetrates into the laboratory, and a thermal neutron flux is measurable coming from spontaneous fission of radio-elements and reactions in the surrounding rock (Bettini, 2014). A gamma background is created from the decay chains of ^{40}K, ^{232}Th and ^{238}U within the laboratory concrete and walls (Chazal et al., 1998). Measurements from Malczewski et al. (2012), show the gamma ray flux in the LSM over the energy range $7.4 - 2734.2$ keV varies as one moves closer to the walls. Low numbers of cosmic muons penetrate the lab, with a vertical flux of 5.4 ± 0.2 m^{-2} day^{-1} (Schmidt et al., 2013), a reduction by a factor of over 10^6 compared to the surface level.

Table 2.1 The radiation sources dominant in biological experiments in underground and surface laboratories are marked with ticks. For most experiments, while construction materials can emit α and β particles, the short penetration distance of these particles greatly reduces their impact on experiments.

	Terr. γ	Terr. α,β	^{222}Rn	^{40}K,^{14}C	Cosmic γ,μ	Cosmic n
Surface	✓	-	✓	✓	✓	✓
Underground	✓	-	✓	✓	-	-

Thermal neutrons are also present in the LSM from spontaneous fission and (α, n) reactions, with a flux of $\phi_n = 1.9 \times 10^{-6}$ cm^{-2} s^{-1} (Savvidis et al., 2010). This flux is over one thousand times smaller than the surface level neutron flux from cosmic sources.

Trace radio-elements in biological media are typically ^{40}K and ^{14}C. Whilst this quantity varies between experiments, it along with the contributions to the radiative background from Radon and the terrestrial γ background, typically dominate the radiation dose received within a biological experiment. When conducting a low background experiment, many of these sources can be suppressed further. Biological experiments may be conducted inside incubated lead shielding to reduce the terrestrial γ rays. Radon suppression presents a greater obstacle and can be minimised by using de-radonised air, or alternatively, a radon-minimising experimental apparatus.

2.1.1 Radon in biological experiments

In the LSM, the radon level is maintained between 5 Bq m^{-3} and 15 Bq m^{-3} (Piquemal, 2012), and is dominated by ^{222}Rn due to its 3.8 day half life (whereas all other natural Radon isotopes have half-lives less than a minute). This can have a significant effect on both physical and biological experiments due to Radon's decay chain. A daughter of naturally occuring Radium, Radon enters the atmosphere by dissolving through rocks after its creation. The dominant decay chain of the gaseous atom is the following:

$$
\begin{aligned}
^{222}\mathrm{Rn} &\rightarrow {}^{218}\mathrm{Po} + \alpha & T_{\frac{1}{2}} &= 3.8 \text{ d} \\
^{218}\mathrm{Po} &\rightarrow {}^{214}\mathrm{Pb} + \alpha & T_{\frac{1}{2}} &= 3.1 \text{ m} \\
^{214}\mathrm{Pb} &\rightarrow {}^{214}\mathrm{Bi} + \beta^- & T_{\frac{1}{2}} &= 27 \text{ m} \\
^{214}\mathrm{Bi} &\rightarrow {}^{214}\mathrm{Po} + \beta^- & T_{\frac{1}{2}} &= 20 \text{ m} \\
^{214}\mathrm{Po} &\rightarrow {}^{210}\mathrm{Pb} + \alpha & T_{\frac{1}{2}} &= 160 \text{ µs} \\
^{210}\mathrm{Pb} &\rightarrow {}^{210}\mathrm{Bi} + \beta^- & T_{\frac{1}{2}} &= 22 \text{ y} \\
^{210}\mathrm{Bi} &\rightarrow {}^{210}\mathrm{Po} + \beta^- & T_{\frac{1}{2}} &= 5.1 \text{ d} \\
^{210}\mathrm{Po} &\rightarrow {}^{206}\mathrm{Pb} + \alpha & T_{\frac{1}{2}} &= 140 \text{ d}.
\end{aligned}
$$

Immediately upon decaying, ^{222}Rn ceases to diffuse gaseously and falls. It poses significant problems in low background experiments as its immediate daughter nuclei decay rapidly, releasing two α and two β^- particles before the comparatively stable ^{210}Pb is reached. In physics experiments, this problem is typically solved by using de-radonised air in experiments. This is not always practical though in biological experiments as the process of removing radon from air tends to dry it significantly, which may alter biological outcomes.

Given that in biological experiments the complete absence of radon is not as necessary as in physics, it is often better to construct experimental geometries which reduce as much as possible the quantity of radon to which a sample is exposed.

Radiation from radon may affect a biological experiment in one of two ways. First, radionuclides in water, or radon dissolved in water from the air may decay inside the sample. Secondly, airborne radon can decay and fall on to a sample, subjecting it to radiation. In the first case, the two relevant factors are the biological sample size and the propensity of radon to accumulate in water. Rather than attempting to estimate the quantity of radon in water based on the atmospheric density of radon and its subsequent diffusion in water, it is more instructive to consider the recommended maximum α emission activity of French tap water, which is 0.1 Bq L^{-1} (Caamano et al., 2011). In a 1.5 mL sample, this corresponds to a maximum of 13 α decays per day. Using pure distilled water in our experiment further reduces this as it eliminates contributions to the radioactive background from water-borne radionuclides such as ^{226}Ra.

In the second case, the amount of airborne radon falling on a sample following its decay grows with the height of the air column above a sample. Reducing this height by raising the sample in an incubator can effectively reduce the radon induced dose. Keeping cell samples wrapped in aluminium foil virtually eliminates this exterior radon dose as falling radon daughters land upon the aluminium, which will block the vast majority of α and β particles, effectively removing the radon background.

Based upon these two cases, it is safe to assume that the impact of Radon in biological experiments at low background can be ignored when suitable measures are undertaken to reduce its impact. More generally, there are experimental geometries where Radon will impact experiments, notably experiments that have large air columns above samples, or water sources with high radionuclide abundancies. These scenarios are less likely to occur however, as most biological practices favour covered nutrient media and distilled water.

2.1.2 Potassium in biological systems

Given that the impact of radon in biological experiments can be easily managed, ^{40}K rests as the sole natural radioisotope whose impact must be mitigated in low background experiments. The impact of ^{40}K on mutation rates in simple cellular systems has been considered in the past. Moore & Sastry (1982) considers that ^{40}K could have played a role as a primordial gene irradiator. The elemental potassium is located inside all living cells, playing an essential role in cellular processes. β^- emission from the element could trigger mutations with a high likelihood, given the particularly high LET with which some electrons are emitted from the radionuclide. This hypothesis was partially tested by Gevertz et al. (1985), who found

that no change was observable in the mutation rate of K12 strain *E. coli* when it was grown in media containing different levels of ^{40}K, a measurement they achieved by using potassium salts at least partially depleted of ^{40}K (and consequently enriched in ^{39}K).

Practically, low background biological experiments need to accurately quantify the radiation dose received from internal ^{40}K as this is often the dominant source of radiation exposure. For certain biological growth media the exact chemical composition of the media is known, and thus the activity of ^{40}K inside the media can be accurately determined. For other media, this is more difficult as the two major potassium sources, tryptone and yeast extract, are purified products that result from the digestion of organic compounds. Here, gamma spectroscopy can be used to determine the ^{40}K concentration of these mixtures.

There always remains the possibility to use biological media depleted of ^{40}K through physical separation processes. This process is expensive, but it does permit one to conduct experiments where virtually all background radiation is suppressed. At this level, further challenges are encountered, as the radiation background can become dominated by unexpected sources, such as metals in the experimental apparatus, trace radionuclides in plastics, and even the radioelements present in inks used in to mark samples. As such, any experiment that seeks to reduce the background radiation level below the level that naturally occurring ^{40}K forces in an experiment requires an accounting for sources that goes beyond the scope of this thesis.

2.1.3 Activity measurements for biological media

In our experiments, *E. coli* is grown in glucose-enriched Davis Medium (Carlton & Brown, 1981). In order to be able to compare the activity of Davis Medium to other commonly used biological media, we measured by gamma spectrometry the potassium concentrations of two common ingredients in biological experiments, yeast extract and tryptone. In particular, these are derived from biological processes rather than a mixture of chemicals, so their potassium concentration is not derivable analytically. Yeast extract (Sigma-Aldrich 70161) is a nutrient rich medium containing autolysed yeast cells. Tryptone (Sigma-Aldrich T9410) is a peptide mixture made from the digestion of casein by the tryptase enzyme. Inside a gamma spectrometer, we measured a 2.65 g sample of yeast extract for 190,450 s and 2.77 g of tryptone was measured for 239,602 s. In both measurements the potassium concentration was found from the size of the 1.46 MeV gamma ray peak in the spectrum, and was corrected for the detector background, geometry and efficiency.

In Table 2.2, we show the concentration of potassium in the various biological powders, and the associated activity of ^{40}K. In both yeast extract and tryptone, there is significantly less potassium than in Davis Medium powder. Whilst the potassium concentration of a biological

Table 2.2 Measured concentrations, c, of potassium and activity, A, of ^{40}K within one kilogram of standard biological powders from gamma spectroscopy. The measurement from Davis medium is calculated from its chemical composition.

Medium (as powder)	c(K) (mg g^{-1})	A(^{40}K) (Bq kg^{-1})
Yeast Extract	63.3 ± 1.2	$1.96(.04) \times 10^3$
Tryptone	2.5 ± 0.2	7.7 ± 0.6
Davis Medium	351	1.084×10^4

culture depends on the levels of yeast extract and tryptone demanded by the recipe, large reductions in concentration can be made by using media other than the Davis Medium. For example, 1 L of Lysogeny Broth (containing 10 g L^{-1} tryptone, 5 g L^{-1} yeast extract) contains 341.5 mg L^{-1} of potassium, a reduction in the potassium concentration by a factor of 10.9 times compared to Davis Medium (containing 10.6 g L^{-1} powder typically).

2.2 Evaluating Dose in Different Environments

We address dosimetry in the different environments we consider in two parts, reflecting that we have used a combination of measurements, where available and simulation, where measurements were not possible. We first present the dosimetric measurements we were able to take, followed by our dosimetry simulations. The results of both these sections are summarised together in Section 2.2.3.

2.2.1 Simple dosimetry

In both the LPC and LSM, measurements of the gamma dose and spectrum were made. Dose measurements made in the LSM were all taken within the recently added biology room with the detector in the centre of the room, or in cases noted as being "shielded" measurements, within a lead shield consisting of an interior layer of 5 cm of copper, surrounded by 10 cm of lead. Dose measurements of the background at the LPC were made in the microbiology laboratory attached to the LPC.

Measurements of the gamma spectra in each location (Figure 2.2) were made using the NaI (Tl) detector of a handheld IdentiFinder Ultra-NGH. The detector chamber in the device is 36.0 mm in diameter and 50.8 mm long. Integration times for each observation varied from 7.46×10^4 s at the LPC to 2.61×10^3 s and 7.87×10^4 s in the unshielded and

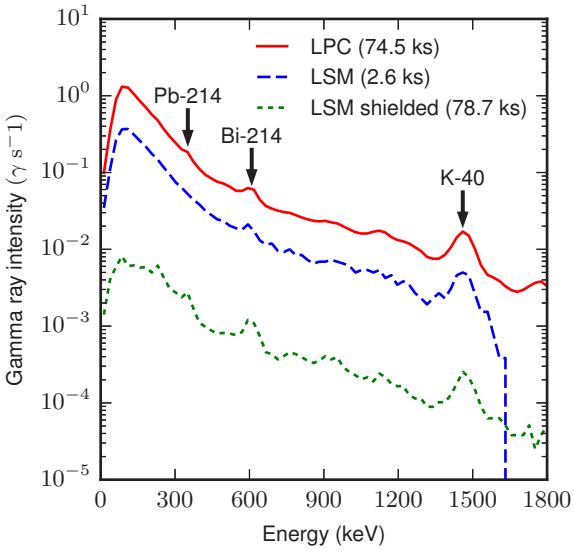


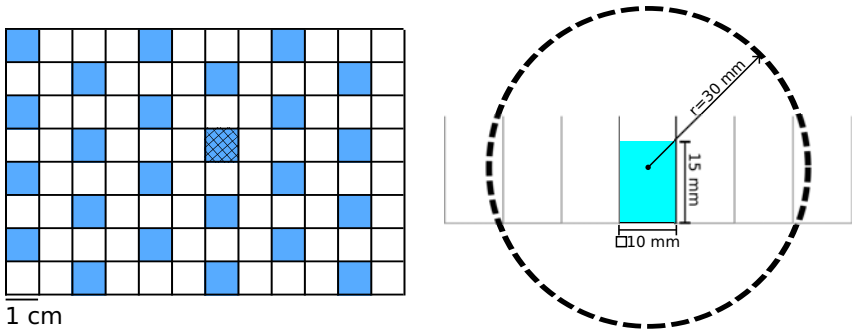
Fig. 2.2 Comparison of gamma spectra at Clermont-Ferrand (LPC, solid red), the LSM biology lab (LSM, dashed blue), and inside the experimental lead chamber (LSM shielded, dotted green) measured with a NaI (Tl) detector.

shielded LSM environments respectively. The measurements were made in Modane on March 18 and 19, 2015 and in Clermont-Ferrand from March 19-20.

The absolute count rate at the LPC was significantly higher than the count rate in the LSM, at 10.57 $\gamma\,s^{-1}$ compared to 3.16 $\gamma\,s^{-1}$ and 0.092 $\gamma\,s^{-1}$ in the LSM biology room and lead shielding respectively. A steep drop is observed in the LSM biology room at counts above 1,650 keV as the integration time was not sufficient to observe gamma rays at this energy. The curves for the LPC and shielded LSM environments do extend beyond this to 3,068 keV, however there are very few counts in this region (in both cases, $< 0.6\%$ of total counts). Across the entire range of energies sampled, the gamma background in Clermont-Ferrand remains approximately 100 times greater than that in the shielded LSM environment, and 3 times greater the unshielded environment.

Whilst this measurement lacks the spectral resolution of a high purity germanium crystal, gamma peaks corresponding to some elements are evident and have been marked on Figure 2.2. These are all traceable to the primordial radionuclides that surround the laboratory. Most notable is the 1460 keV peak from ^{40}K, and the 609.3 keV peak from ^{214}Bi, a radon daughter. An excess is observed in the curves near 338-362 keV, which is likely due to the 352.0 keV peak of ^{214}Pb (also a radon daughter).

The gamma dose was measured using the dosimeter built into the IdentiFinder probe in each location. All dose measurements made were corroborated by secondary measurements from a Canberra Radiagem 2000 Personal Dose Rate and Survey Meter with a Very Low



(a) 96-well microplate from above. (b) Individual well from the side.

Fig. 2.3 The geometry of the 96-well microplate used (a), highlighting the central well that was chosen to be used as a detector (hatched), alongside an individual well filled with 1.5 mL of water (b). The individual well is drawn surrounded by a radius 30 mm circle, representing a spherical source used to simulate isotropic irradiation.

Dose Rate Probe attached. The ambient equivalent doses measured by these devices were converted to absorbed doses using the radiation weighting factor for gamma rays of 1.

Dosimetry reveals the significant decrease in the gamma dose achieved by conducting experiments in the underground lab. Using the IdentiFinder gamma dosimeter, the gamma dose in the LSM biology room was found to be 20 ± 5 nSv hr ^{1}, whilst in the Clermont-Ferrand microbiology laboratory the gamma dose was significantly higher, at 150 ± 10 nSv hr ^{1}. Measurements made inside the lead chamber with the door ajar were not possible as the dose rate was below the detection threshold of the dosimeter for the integration period. The IdentiFinder detector was limited to a precision of 10 nSv hr ^{1}.

2.2.2 Dosimetry simulations

In order to corroborate our gamma dosages, and to obtain the absorbed doses in the biological sample from both cosmic background sources and the internal emission from radionuclides in the nutritive medium, we used Geant4[1] to model the transport and energy depositions of each of these radiation sources. A simulation geometry was defined based upon the 96-well microplates within which bacteria grew on a daily basis, shown in Figure 2.3. For these simulations, electromagnetic physics was simulated using the "Livermore" physics constructor. The simulations were repeated using the default "StandardPhysics" constructor to ensure that results were consistent across different physics models. The results obtained using the Livermore models are by large presented here, as these models are better adapted to low energy electromagnetic physics.

[1] version 10.1.patch01 was used throughout §2.1

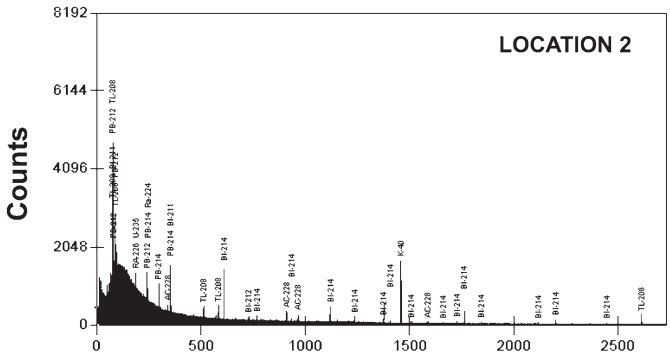


Fig. 2.4 The γ spectrum measured using a high purity germanium spectrometer in the centre of the LSM grand hall, using a collimator around the spectrometer. Identifiable emission lines are marked with their mother radionuclei. From Malczewski et al. (2012).

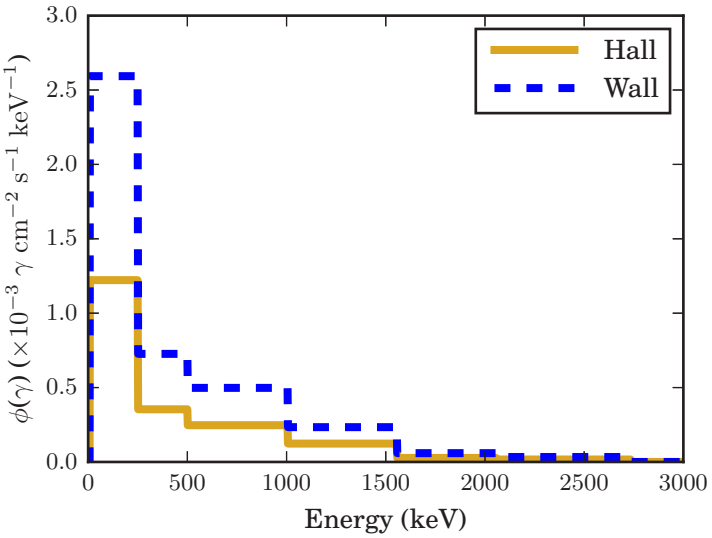


Fig. 2.5 Binned γ spectra based upon Table 2.3 for measurements made by Malczewski et al. (2012) at both a wall in the LSM and the centre of the grand hall.

Table 2.3 Gamma fluxes in $\gamma\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ measured by Malczewski et al. (2012) within the LSM, with a detector near the centre of the main hall and flush against a wall.

	7.4-249.8 keV	250.2-500.4 keV	500.8-1005.2 keV	1005.6-1555.8 keV	1556.2-2055.8 keV	2056.2-2734.2 keV
Hall	0.185	5.36×10^{-2}	3.74×10^{-2}	1.87×10^{-2}	4.10×10^{-3}	2.50×10^{-3}
Wall	0.389	0.109	7.47×10^{-2}	3.51×10^{-2}	8.89×10^{-3}	5.01×10^{-3}

Choosing one well towards the center, we measured the energy absorbed within this well when it was surrounded by an isotropic gamma source, emitting with the flux defined in Table 2.3. These binned fluxes are based upon spectra (Malczewski et al., 2012) measured with a high-purity germanium crystal detector (Figure 2.4) allowing more distinct radio-elements to be identified above the background than in Figure 2.2. The γ-spectrum at the centre of the hall is shown in Figure 2.4. This resolution of course disappears when the binned spectrum is considered (Figure 2.5), however as the spectrum is background-dominated this has little effect on the final result. The isotropic source was defined as a sphere of radius 3 cm around the centre of the well (Figure 2.3b). Using each spectra, we simulated 10^7 gamma rays, corresponding to 0.524 days for the spectrum at the wall and 1.08 days for the spectrum in the great hall. Physics was simulated using the "Livermore" physics constructor. The simulations were repeated using the default "StandardPhysics" constructor to ensure that results were consistent across different physics models. A secondary production cut of 1 μm was used in both cases.

The Davis minimal broth solution (Carlton & Brown, 1981) that provides nutrition for the bacteria in the experiment has an irradiative effect on the bacteria. We used simulation to measure the dose received from this source, based on the activity of the Davis medium solution found from its chemical composition. The DM250 solution we use (Davis medium without dextrose, sourced from Sigma-Aldrich product 15758 with glucose added) carries a radioactivity from its potassium concentration of 115 Bq L^{-1}. This activity was confirmed using a Germanium gamma spectrometer at the LPC Clermont-Ferrand. ^{14}C within the DM250, in its standard abundance carries a minor contribution to the activity of 5.17×10^{-2} Bq L^{-1}. To simulate these sources, we kept the same simulation geometry as in Figure 2.3b except the water-filled well served as the particle source, for both beta electrons (in the cases of ^{40}K and ^{14}C) and for gamma rays (^{40}K only). For ^{40}K, particles were simulated across two simulations. The first simulation measured the received dose from internal gamma emission, arising from the 1.46 MeV photon emitted when ^{40}K undergoes electron capture. The second simulation modelled internal electron emission from β^- decay, using the emission spectrum measured by Cameron & Singh (2004), shown in

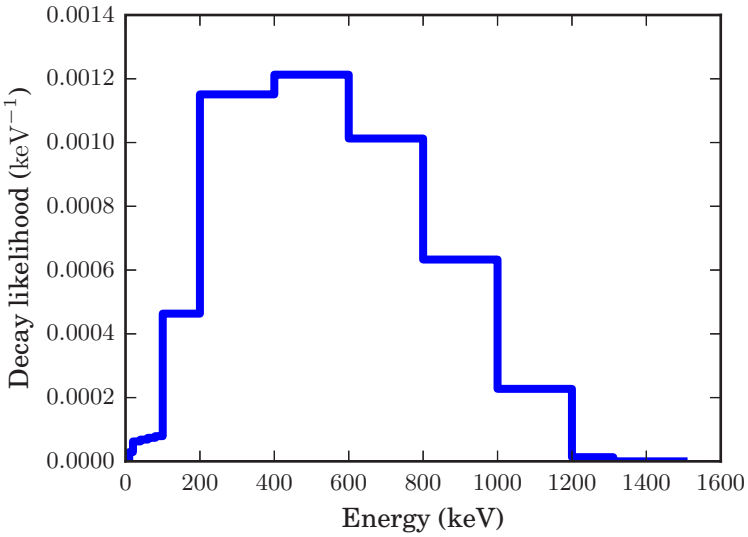


Fig. 2.6 β spectrum measured of ^{40}K, binned according to Cameron & Singh (2004). The mean emitted electron energy is 511 keV.

Figure 2.6. The 10^7 primary particles simulated corresponded to simulation time periods of 6.16×10^3 days (γ source) and 7.33×10^2 days (β source). For ^{14}C, electrons were simulated matching the element's beta spectrum (Tilley et al., 2012), corresponding to 1.49×10^7 days of experiment. Physics processes were modelled using the same methods as for the gamma background.

As part of providing a dose comparison between the LPC and LSM, it is necessary to estimate the radiation dose due to cosmic rays above ground at the LPC. The absorbed dose from charged cosmic rays and photons is reasonably uniform across most latitudes, and a good estimate is provided by the standard annual effective dose, which for these particles is identical to the absorbed dose (UNSCEAR, 2000). Following Bouville & Lowder (1988), this value needs to be increased by 6% to account for the 400 m altitude of the LPC. The absorbed dose from neutrons is more difficult to calculate, and while standard effective dose numbers at sea level are available, the relationship between absorbed and effective neutron dose is a function of neutron energy. In order to estimate the absorbed dose in an experimental well above ground, we simulated the energy absorbed by a well in our experiment again using Geant4, keeping the same well geometry as in Figure 2.3b but with a modified source geometry. The modified source was a circular surface 150 mm in radius, 30 mm above the base of the well, centred directly above the well considered. Particles were generated along the down-facing side of this surface with their direction specified by an isotropic angular distribution. This is equivalent to simulating neutrons arriving at the well from an angle of 11.3° above the horizon up to the zenith. The particles arrive at the well with an

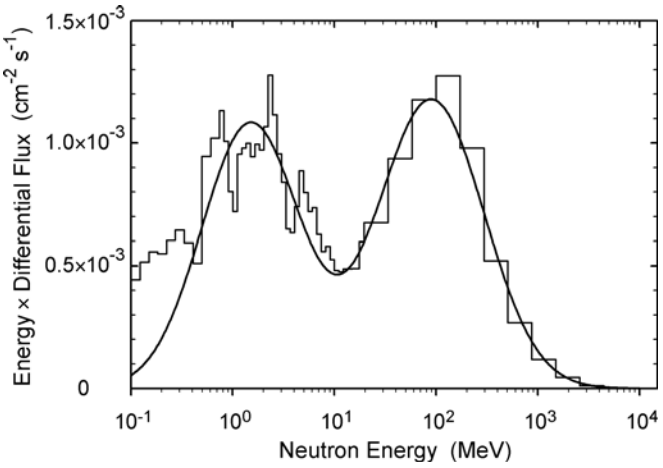


Fig. 2.7 The analytic neutron energy spectrum we simulate (curve) alongside its measurement (histogram). Our simulations consider neutrons with energies up to 500 MeV. (Reproduced from Gordon et al., 2004).

angular distribution of where is the angle from the zenith. 10^7 particles were simulated, corresponding to 4.37 days of real time. The algebraic approximation of the neutron spectrum in New York (Gordon et al., 2004) was used between 100 keV and 500 MeV in the simulation, giving a neutron flux of 5.96×10^{-3} cm $^{-2}$ s $^{-1}$ (Figure 2.7), and physics was modelled using the QGSP_BIC_HP physics list and a 1 mm secondary production cut.

2.2.3 Results

Simulations of the energy deposited in water by the gamma background were made using energy-binned flux measurements taken at a wall and in the centre of the LSM great hall. The results of these simulations are shown in Table 2.4, for the StandardPhysics and Livermore physics constructors in Geant4. Both physics constructors used gave results in agreement with each other. The absorbed dose measured using the spectrum from the great hall is within the precision of that measured by dosimetry, whilst the spectrum against wall overestimates the measured dose.

The major internal contributions to the radiation dose are also presented in Table 2.4. Firstly, the ^{40}K gamma line has a very minor impact on radiation dose, and the radioactive decay of ^{14}C is negligible. Rather, the internal dose is dominated by the emission of beta particles from ^{40}K. It bears noting that the average energy of a ^{40}K beta particle is 560 keV, however as electrons may escape the simulated well, an average of 426 keV is deposited in the well per event. The particles that escape may interact with another well. Whilst this depends strongly on the position of the well on the microplate (as not all wells have

Table 2.4 Simulated absorbed dose rates from the gamma background in the LSM and the radiation sources in the nutritive medium. Doses were calculated using the Livermore and StandardPhysics constructors within Geant4. The gamma background was simulated using spectrum measured against a wall, and towards the centre of the laboratory hall.

Source	Livermore nGy hr^{-1}	Standard nGy hr^{-1}
Gamma (wall)	47.3(4)	47.9(4)
Gamma (hall)	23.3(2)	23.3(2)
^{40}K(β)	25.865(5)	25.939(5)
^{40}K(γ)	0.1080(3)	0.1088(3)
^{14}C(β)	$1.4757(3) \times 10^{-3}$	$1.3762(3) \times 10^{-3}$

the same number of neighbouring wells), an approximation of the radiation dose in one well from another can be made by considering the angular space occupied by a neighbouring well. By considering the maximum solid angle that a neighbour may occupy and also assuming that little energy is lost in the air between wells, an upper bound can be found for the energy deposited in each neighbouring well per event of 6.1 keV. That is, the dose is underestimated by 1.4% per neighbouring well. Thus, considering the geometry of a well in the centre of the microplate with six nearby, filled, neighbours (see Figure 2.3a), the size of the underestimation in the ^{40}K beta dose is no larger than 7%.

The summed doses for each environment are presented in Table 2.5, where we have also marked the estimated neutron dose above ground (4.4 ± 0.3 nGy hr^{-1}) from simulation. The presence of internal ^{40}K sets a limit upon the maximum dose reduction achievable following a protocol that uses Davis Medium. Merely conducting experiments in an underground laboratory reduces by a factor of 4.6 the background radiation level compared to the LPC. Removing the gamma background from the LSM using shielding allows a further reduction, such that the dose compared to the LPC is reduced by a factor of 8.2.

2.2.4 Discussion

These dosimetric results show the suitability of Modane for low background biological experiments. We show that it is possible underground to reduce the background radiation from ambient source to the 1 nGy hr^{-1} level, using relatively simple means. Nevertheless, a significant radiation background remains from ^{40}K in nutritive media. Biological experiments introduce a unique nuance into the standard way of running a low background experiment, where typically it is necessary to eliminate as many background sources as possible, to as large a degree as possible. Practically, biological experiments bring with them a level of

Table 2.5 Doses relevant to biological experiments at the LSM, relative to measurements made at the LPC Clermont-Ferrand. Dosimetry measurements exhibited a time variation on the order of 20% on both sides of the value cited. The background for cosmic rays was taken from UNSCEAR (2008). Doses from potassium and carbon inside the well were simulated using the Geant4 simulation toolkit.

Source	Method	LPC Clermont (nGy hr^{-1})	LSM (nGy hr^{-1})	LSM (shielded) (nGy hr^{-1})
γ background	Dosimeter	150	20	<1
Cosmic rays (μ, e^-, p, γ)	UNSCEAR	33	$\ll 1$	$\ll 1$
Cosmic rays (n)	Simulation	4.4	$\ll 1$	$\ll 1$
^{40}K (γ)	Simulation	0.13	0.13	0.13
^{40}K (β)	Simulation	26	26	26
^{14}C	Simulation	$\ll 1$	$\ll 1$	$\ll 1$
Total		214	46	26

radioactivity in their natural chemical composition, by virtue of the important biological role of potassium. Methods of reducing the potassium dose in biological experiments have not been widely discussed, indeed some experiments have forgotten to include the contribution of potassium within growth media (e.g. Castillo et al., 2015). Media selection also plays an important role in reducing the potassium background. We have presented measurements of the potassium concentration of one manufacturer's yeast extract and tryptone in Section 2.1.3. These measurements can serve as a guide to assessing the radiative impact of different media. Indeed, there is a ten-fold decrease in the potassium concentration, and thus ^{40}K activity going from Davis Medium to Lysogeny Broth. Further reductions in the radiation background could be achieved by using ^{39}K rather than natural potassium in experiments, as was performed by Gevertz et al. (1985). In this case, new dosimetric measurements would be required in the LSM to quantify the dosage below the 1 nGy hr^{-1} level, as portable dosimeters are not sensitive to such low radiation levels.

Working in underground laboratories, the major significant external radiation source is the gamma background. This gamma background can be effectively suppressed using standard lead shielding. While above ground, lead shielding can reduce the gamma background, underground laboratories shelter biological experiments from the far more difficult to eliminate cosmic sources of radiation, typically neutrons and muons. The underground laboratory itself is responsible for a reduction in the absorbed dose by 35 nGy hr^{-1}, according to our quantification of the neutron dose, and the standard estimate of the cosmic background dose from charged particles and photons. A higher reduction is often achievable given that most underground laboratories include other measures to reduce the background radiation

level, such as low radioactivity concrete and low radon air supplies.

It should be noted that our measurement of the gamma background at the LPC are significantly higher than the standard terrestrial gamma background of 60 nGy hr^{-1} (UNSCEAR, 2008). Whilst the measurement we made does convolute the terrestrial and cosmic gamma backgrounds, as the gamma dosimeter could not distinguish the two, this is not a significant factor in the large background measurement as the cosmic background dose is dominated by muon interactions. Instead, the high gamma background at the LPC is caused by its location, in the 'Massif Central' region of France, which is known to have a particularly high background radiation level.

We have not specifically considered ^{222}Rn and its daughters as an important contributor to the internal dose in these measurements. Whilst the energy contained in from decaying ^{222}Rn is large, and can significantly impact a biological system, the experimental geometry can play a large role in varying this dose. Lungs, or a lung like system for example will receive a significantly larger radiation dose from radon, due to the frequent gas exchange, than a glass test tube containing bacteria in a nutritive medium, and little air. Even without elaborate measures to limit exposure, the impact of ^{222}Rn is relatively small, for example, in the Gran Sasso laboratory, Fratini et al. (2015) find its contribution to be less than 1% of that of ^{40}K (0.17 nGy hr^{-1} in a 5 Bq m^3 environment). In the LSM, we judge the contribution of Radon to be small also, based on the low ambient Radon concentration (≤ 25 Bq m^{-3}, or $\leq 5 \times 10^{-5}$ Bq $well^{-1}$), and the geometry of our system, where an aluminium sheet prevents falling products from decaying, airborne radon from irradiating our samples.

Our simulation measurements of the gamma background in the main hall are in agreement with gamma dosimetry measurements, and as such would serve as an appropriate input for future simulations of the gamma background in Modane. Conversely, the gamma spectrum taken next to the wall overestimates the gamma dose, which is unsurprising given that the geometry of the measurement is heavily directional. Whilst using an energy binned spectrum as a simulation input causes the spectrum to lose the distinctive peaks that characterise a gamma spectrum, this has little effect on the underlying result as the majority of counts are in the continuum part of the spectrum.

Assessing the biological impact of a background reduction from the 214 nGy hr^{-1} absorbed dose an organism would receive on the surface in Clermont-Ferrand with the much lower 26 nGy hr^{-1} dose in the LSM is difficult due to the poorly determined response of living systems in ultra-low background environments. Daly (2012) suggests that double strand breaks (DSBs) from radiation exposure typically occur with a frequency of 0.004 DSB Gy^{-1} Mbp^{-1}. If a linear model for induced radiation damage with dose holds true, then there would be an 8-fold difference in the number of DSBs occuring within organ-

isms outside an underground laboratory, compared to those studied within it. Following this logic, cells grown underground would be considerably less damaged than those grown above ground. Whether this is observable however requires a better quantification of how radiation interacts with cells, compared to the biological processes acting upon cells. Moreover, such a simple assumption ignores how regulatory mechanisms and repair functionalities act within cells at low radiation doses, which is likely far from linear.

2.2.5 Conclusions and Perspectives from Dosimetry

Dosimetry is an inherently macroscopic concept. At high radiation doses, it can lead to good predictions of biological outcomes, however at low doses its usefulness is limited. Models of the biological effect of radiation, such as the Linear-No Threshold model that are built from extrapolations from high dosage data, break down at low doses (Feinendegen et al., 2004; Tubiana et al., 2009). Our dosimetric studies here have highlighted to what extent different radiation sources can be removed from the environment. A key finding is that potassium in biological media sets an approximate limit on the amount of radiation that can reasonably be removed. Dosimetry is also important for providing a simple comparison between multiple environments. A difference in dose corresponds loosely to a difference in radiation-induced damage events. The utility of this measurement depends highly on each individual cell's response to radiation.

To better understand the impact of radiation on cells though, it is necessary to quantify that impact on the level of individual cells. This is the object of the next section, where rather than providing dosimetric evaluations of radiation environments, we use simulation to work out the frequency with which radiation tracks traverse individual cells.

2.3 The impact of radiation environments on cells

When considering the impact of ionizing radiation on cellular systems from the environment, the absorbed radiation dose is considered by experimentalists. While this is appropriate in high dose regimes, it is less applicable in low background biological research. Absorbed dose measures a continuous energy deposition per unit mass, when in reality energy is deposited by ionizing particles along tracks. For low doses, these tracks do not always cross a significant proportion of cells in the populations studied in a biologically relevant time period. In this section we use Geant4 simulations to replicate our biological experiments and their radiation environments, in an effort to quantify how frequently radiation tracks traverse cells.

The need for such studies is highlighted by a recent study conducted at the Waste Isolation Pilot Plant in New Mexico where a reduction in the growth rates of both *Shewanella oneidensis* and *Deinococcus radiodurans* was observed when they were grown for 48 hr in an extremely low background environment compared to a reference environment (Castillo et al., 2015). As the bacteria were given no time to adapt to the radiation environment, and upon re-introduction to the reference environment, the bacteria immediately recovered their former rate of growth, one is left wondering how the bacteria could so rapidly sense the change in the radiation level. Quantifying the frequency with which radiation tracks traverse cells permits this, though even simple estimations of the frequency with which ionisation events occur in cells suggests it is difficult for ionising radiation alone to mediate this population wide response (Katz, 2016).

Using simulation, the stochastic impact of the radiation background in biological experiments can be constrained, by calculating the frequency with which particle tracks intersect cells and deposit energy within them. Monte-Carlo based particle track simulation packages have seen wide use in simulating the impact of radiation upon cells in radiotherapy (El Naqa et al., 2012; Nikjoo et al., 2006) and are easily applicable to cellular dosimetry (Douglass et al., 2012; McNamara et al., 2012). Going further, Monte Carlo codes can simulate both direct and reactive oxygen species induced damage caused by radiation sources, both through explicit simulation (Friedland et al., 2011) and analytical modelling of the chemical processes induced by radiation (Liang et al., 2016). This is beyond the scope of this chapter, but is discussed in Chapter 3.

Continuing from the previous section, we present a method in this section in which the Geant4 simulation toolkit (Agostinelli et al., 2003; Allison et al., 2006, 2016) is used to accurately calculate the frequency with which ambient radiation sources interact with bacterial cells (all simulations in this section were run using Geant4.10.2.p01 MT). We apply this method to our LTEE in both the LSM and LPC. We show how many cells are impacted per unit time by the radioactive background, placing bounds on the maximum rate of mutations triggered by the ionizing background. More generally, these numbers are interpreted in the light of short term low background experiments, giving a physical quantification of the extent to which bacterial cells may indeed be able to ‘sense’ the radiation present in their environment.

As part of the Geant4-DNA collaboration, the source code used for the simulations in this section has been released publicly[2].

[2]The curious reader is referred to Lampe (2016), and http://github.com/natl/multiscale.

2.3.1 Biological Conditions

Our simulations are designed to mimic the LTEE we are conducting in the LPC and LSM. A more complete description of the experiment is provided in Chapter 4. In both the ambient radiation environment in the LPC and the low background environment provided by the LSM, *E. coli* has been grown across hundreds of generations in 24 wells of a 96-well polypropylene microplate. Each well in the microplate was filled with $1 \times 1 \times 1.5$ cm^3 of Davis minimal broth (Carlton & Brown, 1981) enriched with 250 mg L^{-1} glucose. This is essentially the same geometry that was simulated in Section 2.2 (see also Figure 2.3. Every day, the liquid culture was agitated constantly at 170 RPM at 37°C for 24 h. At the end of each 24 h cycle, 5 μL of bacteria was transferred to a new microplate. The bacterial concentration begins at the start of each daily cycle at 1.7×10^6 cells mL^{-1} which rises to 5.0×10^8 cells mL^{-1} at the end of the growth phase.

Our simulations consider two levels. The macroscopic level considers environmental sources interacting with the liquid medium in the microplate itself. At the microscopic level, we consider the charged particles created within the well and how frequently they traverse cells. In this second, microscopic simulation, 4000 cells are simulated in a cube of side length 200 μm^3, which replicates the maximum cellular density observed in experiments. This two-stage simulation and the repeating boundary condition at the microscopic level where chosen to enable us to efficiently place over 10^8 cell geometries into memory for the simulation. In hindsight, a more effective way to run this simulation is to use multiple placements of a small repeating unit equivalent to our microscopic simulation geometry into Geant4, thus eliminating the need to run two levels of simulations, and develop a repeating boundary condition.

2.3.2 The Macroscopic Level

The simulations at the macroscopic level consider the environment of the experiment. Following the dosimetric simulations, we modeled a 96-well microplate using polypropylene (Fig 2.3a) in Geant4. A Davis minimal broth solution was modeled as water that was enriched elementally by the chemical composition of the Davis minimal broth (specifically matching the composition of Sigma-Aldrich product 15758), as the trace presence of these constituents, notably potassium cause a $\approx 10\%$ increase in the neutron absorption cross section of the well (c.f. Varley, 1992). The composition used is shown in Table 2.6.

The aim of this simulation was two-fold, we measured the dose deposited in the well in addition to recording the charge carriers created within the well, either directly from an internal source, from a charge carrying particle entering the well from the outside, or from

Table 2.6 Simulated composition of Davis Medium

Molecule	Concentration (%)
H_2O	98.9147
N	0.0212
H	0.0163
S	0.0269
O	0.4430
K	0.3702
P	0.1700
Mg	0.0020
Na	0.0117
C	0.0222

physical processes where charged particles were created within the well by neutral particles that entered the sensitive volume (ie. photons and neutrons). By storing only charged particles created, and not the secondaries that these particles in turn created, we preserved the spatial correlations between all electrons created by ionization from any recorded charged particle within the sensitive volume, as they were 're-created' at the microscopic level. Particles were still tracked after they were saved in order to measure the energy that they deposited in the well and to observe whether secondary particles were later created that re-entered the well even after a given particle had left. When particles were saved, their positions, momentum directions, species' and energies were recorded. At the conclusion of the run, these were placed into a binary data file formatted following the ROOT format. This file served as the input for microscopic level simulations.

2.3.3 The Microscopic Level

Within the previous simulation, a phase space file was produced that contains the positions, energies and directions of each proton and electron created within a wells of a microplate. This data is used as the input to a new simulation run at smaller length scales, which allows the impact of these particles upon cells to be evaluated. By simulating individual cells in their nutritive medium, the frequency and magnitude of energy deposits in cells from a radiation source can be found, and the exact energies and types of particles that enter cells to interact with DNA can be quantified. The energy spectrum of particles entering cells can be placed into a phase space file with each particle's direction and point of entry into the cell in order to seed a yet lower level simulation where the interactions between these particles and DNA is measured, so that the DNA damage induced by a particular radiation source may be assessed.

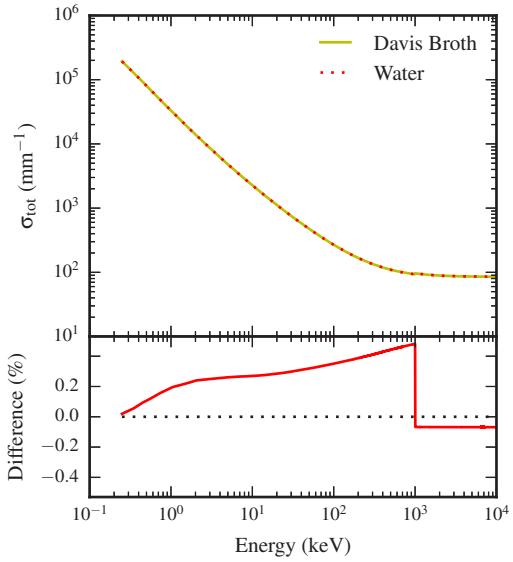


Fig. 2.8 The top panel shows the total electron interaction cross section (σ_{tot}, given per volume, for $\rho = 1$ g cm^3 solutions), which determines the likelihood of electron interactions occurring for a given electron energy, in both water and Davis broth. The difference between these curves is shown in the bottom panel. Across the range of electron energies considered in our simulations, water approximates the Davis broth solution to within $< 0.5\%$. The comparison here was made using the Penelope physics models in Geant4.

The geometry used within Geant4 at this level is a cube of side length 200 μm filled with 4000 bacterial cells. This domain is significantly smaller than a well in a microplate, and is easily escaped by even low energy electrons as 100 keV electrons have a mean penetration distance of 140 μm in water (Berger et al., 1998). A periodic boundary condition is thus placed on the domain so it may adequately represent the larger domain.

Particles are generated inside this domain. Macro commands allow the user to specify the location where particles are generated and whether they are generated in a uniform direction or isotropically, if a phase space file is not being read. The capacity to use an input energy spectrum following a user histogram is also available via macro commands. When generated particles enter cells, the energy they deposit in each cell is saved, as well as their position, direction and energy at entry. The number of cells hit in each event is also saved.

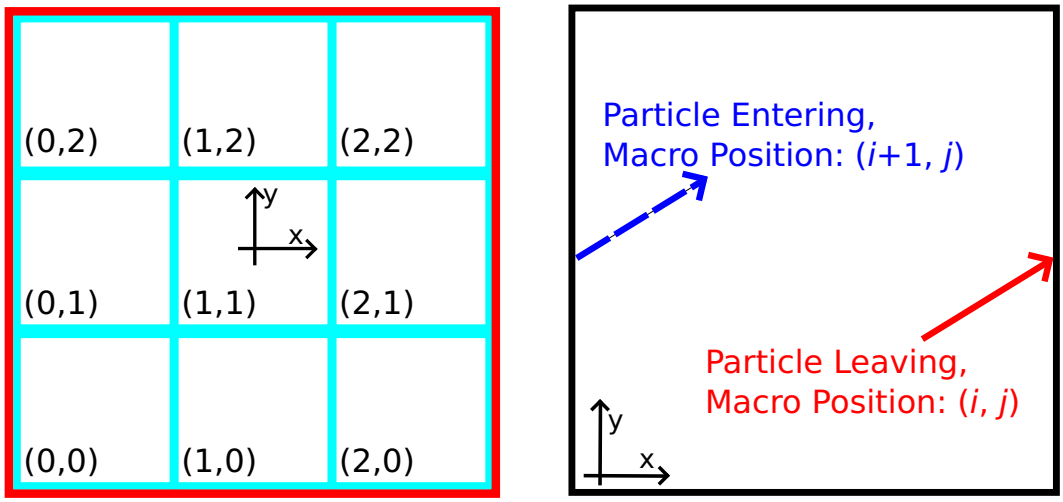
Each cell itself was modelled as a cylinder of length 2 μm capped at each end by $r = 0.5$ μm hemispherical end caps. Unlike at the macroscopic level, cells were simulated as being composed of water, in a water medium. The density of water inside the cells was however set at 1.10 g cm^{-3}, a compromise between the few measurements of the internal *E. coli* density we found in the literature Baldwin et al. (1995); Godin et al. (2007). Approximating the Davis Medium as water has little impact here, as the electron cross section is not substantially different between these two media (Figure 2.8 shows the maximum difference over the energies we are most interested in to be $< 0.5\%$).

At least 10^6 particles were simulated in each run at the microscopic level, which was enough for the simulation results to converge. As the inputs at the microscopic level are dependent on the macroscopic level, the time normalisation for each microscopic simulation run was determined from the rate at which secondary particles were created (following normalisation) in the macroscopic simulation.

Implementation of a repeating boundary condition

The repeating boundary condition was implemented by adding a counter to G4Track objects that tracks the location of the particle in a larger, fictitious parent domain, so that particles may leave the simulation if they travel far enough in any direction. In this way, particles are identified by their identity not just within the box, but also within the larger domain that the repeating boundary replicates.

Using the repeating boundary conditions, we introduce two simulation regions. The microscopic region is that which is actually simulated, a small domain containing a reduced number of cells. The parent region however is made of tessellated microscopic regions and describes the entire simulation region. This region allows a particle's position in real space to be calculated. The relationship between these two regions is illustrated in Figure



(a) (b)

Fig. 2.9 **a**) The repeating boundary condition allows the simulation area at the macroscopic level to be broken up into a series of microdomains with $(x,y) \in [-100\ \mu\text{m}, 100\ \mu\text{m}]^2$. Each microdomain contains the same geometry due to the repeating boundary, however each track additionally stores the index of its microdomain domain, here labelled in the bottom left of each small cube. **b**) When a particle leaves a microdomain, it re-enters at the other side, due to the periodic boundary. When this occurs, the track updates the index of its microdomain counter. This allows the position of the track in the overall parent domain to be recovered.

2.9. The microscopic region tracks position within its domain boundaries centered about an origin at the box center. The parent domains are counted from $(x,y,z) = (0,0,0)$ up to $(x,y,z) = (n_x,n_y,n_z)$ where n_x (n_y, n_z) is the number of repeated microscopic regions needed to fill the macroscopic region in the $x-$($y-$, $z-$) direction. Whenever the repeating boundary is used, the particle updates its position inside the parent domain to account for this change. When the particle leaves the bounds specified by the domain, it is killed.

The implementation of the repeating boundary condition is handled in Geant4 using a custom physics process, a parallel world constructor, the user information classes for both tracks and events, and the post user tracking action (Lampe, 2016). The relevant classes inherit from existing Geant4 classes and are called *MicroPeriodicBoundaryProcess* (inherits from *G4VDiscreteProcess*), *MicroParallelWorldConstruction* (inherits from *G4VUserParallelWorld*), *MicroEventInformation* (inherits from *G4VUserEventInformation*), *MicroTrackInformation* (inherits from *G4VUserTrackInformation*) and *MicroTrackingAction* inherits from *G4UserTrackingAction*). When a primary vertex is created within Geant4 (inside the implemented 'Primary Generator Action'), its location in both the microscopic and macroscopic domains are set from an initial position in the macroscopic domain that

is broken down into two components, an index (i_x, i_y, i_z) corresponding to the space that the microscopic domain would occupy within the larger space, and a position (p_x, p_y, p_z) within that microscopic cube.

The processes involved in transporting particles and ensuring that particle positions in the parent domain are valid are illustrated in Figure 2.10. The *MicroTrackingAction* class is responsible for assigning the correct *MicroTrackInformation* object to each track within Geant4. This action is called each time a track is taken from the stack in Geant4, and *MicroTrackInformation* stores a record of the tracks position in the larger 'macroscopic' spatial domain. While the object is being tracked, the track information may be changed to reflect the passage of the particle through the periodic boundary. We ensure the periodic boundary process is always triggered when it is hit by a particle, regardless of what is in the main Geant4 world, by defining its region in a parallel world. To simplify the processing of trajectories when a particle crosses the boundary, the particle crossing the boundary is killed, and a clone is created to replace it, which is placed in the vector of secondaries created by the track. When the track ends, the tracking action acts on all the secondaries created along the track to ensure that they have the appropriate *MicroTrackInformation* object assigned to them. If the particle has crossed a boundary, the track information is used to identify any changes necessary to the secondary's position, so that it is correctly moved to the opposite side of the periodic domain.

2.3.4 Background sources and simulation parameters

In the previous section we discussed the relative contributions of different background sources to biological experiments and presented measured dosages pertinent to these experiments in both above and below ground environments. Here, however, the complete spectrum of each particle source is needed, rather than just dosimetric measurements and UNSCEAR recommendations. We simulate the γ and ^{40}K backgrounds in both the LSM and LPC, as well as the cosmic μ and neutron backgrounds in the LPC. The γ, ^{40}K and neutron backgrounds are similar to those presented in Section 2.2.2, whilst the cosmic μ background has been rederived for the 400 m altitude of the LPC.

At both simulation scales hadronic processes were modeled using the QGSP_BIC_HP physics list and the 'Option 4' standard EM physics list was used to simulate EM processes with a low energy production threshold of 100 eV. Option 4 was chosen in particular due to its more accurate Compton scattering model, which may improve simulations of the gamma background (Brown et al., 2014). At the macroscopic level a secondary production cut of 1 μm was used, and at the microscopic level the secondary cut was reduced to 10 nm. We also set the maximum step size to 10 nm for these simulations.

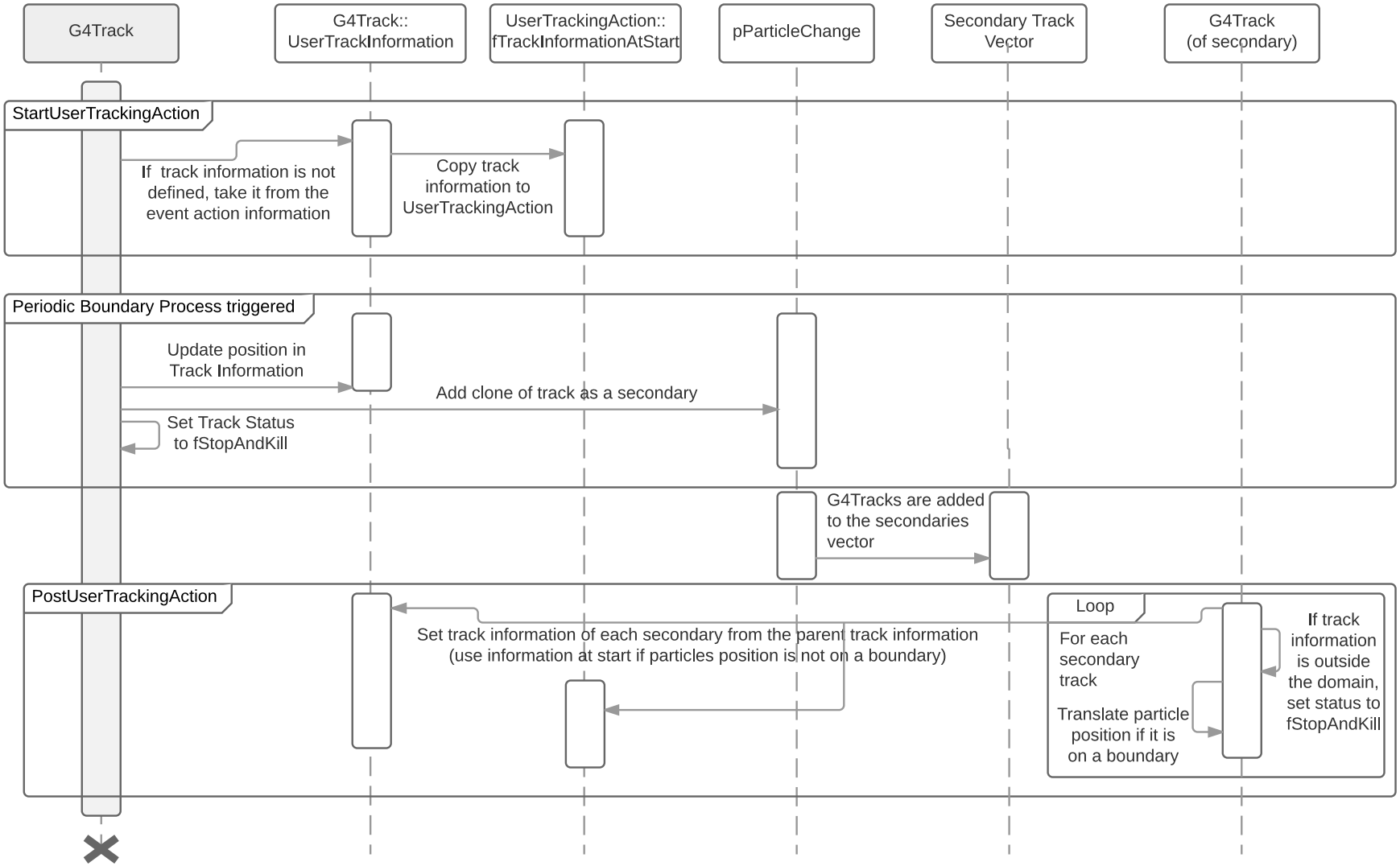


Fig. 2.10 Sequence diagram indicating the interactions between the various classes that implement the repeating boundary condition and the G4Track object.

The choice of the secondary production cut at the microscopic level is very important, as energy deposition measurements are made within cells with a characteristic length of around 1 μm. In Geant4, the accuracy with which energy deposits can be localised is determined in part by the secondary production cut. The cut stops the production of secondaries that have a range shorter than that specified in the cut. The energy of these secondaries is instead recorded as part of the energy lost along the step. By insisting on a secondary production cut significantly smaller than the cell size, spatial errors in energy deposition are minimised in the simulation.

The gamma background

Recalling Section 2.2.2, we have based our simulations of the terrestrial gamma background upon measurements made within the LSM using a high purity Germanium spectrometer (Malczewski et al., 2012). The γ background in the LPC, when measured using a Thallium-doped Sodium Iodide handheld gamma spectrometer was shown to be very similar to the LSM spectra, albeit significantly reduced in intensity (c.f. Figure 2.2). Accordingly, the LPC spectra used was obtained by scaling the underground spectrum by 6.4 (thus ensuring a surface dose of 150 nGy hr^{-1}, consistent with measured values). The tabulated underground spectra is shown in Table 2.3. At the macroscopic level, the energy binned gamma fluxes were simulated as isotropic, by considering the source to be an $r = 3$ cm gamma-emitting sphere around a central well in the microplate, which was chosen to be the detector. Incoming gamma rays create electrons by the Compton and photoelectric effects, as well as occasional positrons by pair production. The positions, directions and energies of these secondary particles created in the chosen well was then used to seed the simulation at the microscopic level. 10^8 primary gamma rays were simulated at the macroscopic level, equivalent to 10.8 days of exposure in the LSM (after scaling, this is 1.68 days in the LPC). At the microscopic level, 2×10^6 secondaries were simulated, randomly selected from those created at the macroscopic level.

The nutritive background - Potassium-40

The radiation background from the nutritive medium in biological experiments is dominated by the contribution from β^- emission by ^{40}K, which dominates the absorbed dose from ^{40}K γ-emission and ^{14}C β-decay by over two orders of magnitude. To model the secondaries that enter each well in our experiment, we chose a central water-filled well in the microplate geometry to be the sensitive region, and defined that well and its six closest filled neighbors to be sources. Within each of these seven source wells, electrons were created with a uniformly

random distributions of position and emission direction. The energy spectrum (see Figure 2.6) was defined by the β spectrum of ^{40}K (Cameron & Singh, 2004). 10^7 events were simulated at the macroscopic level, corresponding to 105 days of real time. At the microscopic level, 10^6 events were simulated, drawn randomly from the phase space file created at the macroscopic stage. For this source, the phase space file consists solely of electrons with an energy spectrum very similar to that at the macroscopic level, as all electrons created within the sensitive region are saved as soon as they are created.

The cosmic neutron background

As in Section 2.2.2, the neutron background was simulated using an algebraic expression for the neutron differential flux in New York (Gordon et al., 2004) between 100 keV and 500 MeV (Figure 2.7). Within this range, the differential flux is $\phi_n = 5.96 \times 10^{-3}$ cm^{-2} s^{-1}. In simulation, this source was modeled as a disc of radius 10 cm situated 30 mm above the center of the microplate emitting neutrons uniformly along its surface with an isotropic angular distribution. This causes the particles to arrive at the well with an angular distribution of $\cos^2\theta$ where θ is the angle to the vertical. Neutron interactions within the well create primarily free protons, electrons, alpha particles and ^{16}O ions, while many other ions are created in small quantities. Given the high kinetic energy of the incident neutrons, these particles often also have a high kinetic energy, sometimes a few hundred MeVs. All these particles are saved at the end of the simulation, to be read into the microscopic simulation. At the macroscopic level, 10^8 particles were simulated corresponding to 98.4 days of real time. 10^6 particles were then simulated at the microscopic level, drawing randomly from the list of particles created at the macroscopic level.

The cosmic muon background

We simulated the cosmic muon background based upon the spectrum predicted by Chirkin (Chirkin, 2004), which provides the differential muon flux as a function of both muon momentum and direction. A correction to the spectrum was made corresponding to the altitude of the LPC, provided by Neiss (2016). While this derivation is for muons with momenta above 600 GeV/c, it models low energy to an acceptable level of accuracy for our simulations. To generate the distribution of muons, we sampled the differential flux distribution 10^6 times, generating a series of pairs of zenith angle and muon energy. We considered muons with momenta between $0.1 - 50$ GeV/c, and simulated 54% of muons as μ^+, with the rest as μ^-. The total integrated flux within this range was 3.4×10^{-2} μ cm^{-2} s^{-1}, in reasonable agreement with the accepted sea level muon flux (Olive et al., 2014). The energy

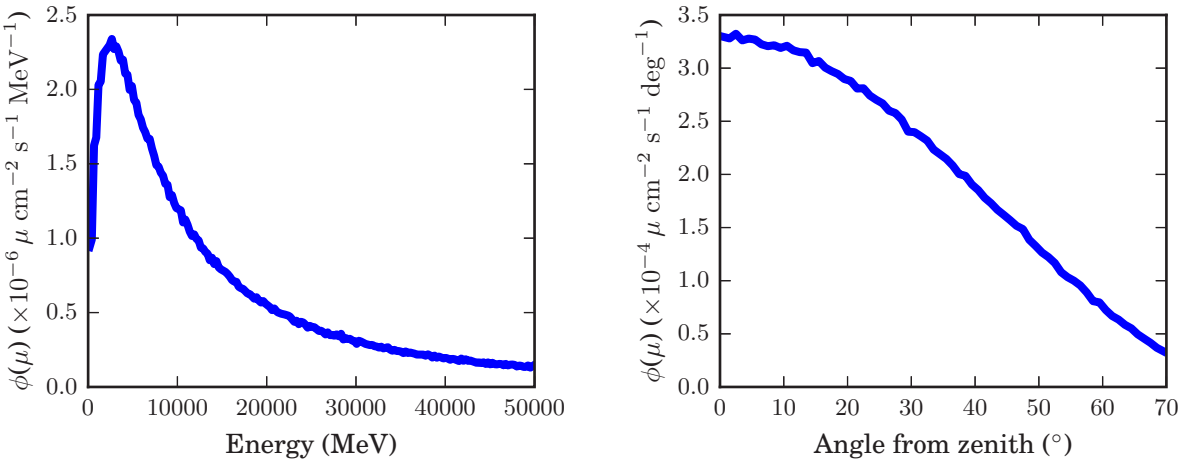


Fig. 2.11 The cosmic muon energy spectrum is angle dependent, as low energy muons are more likely to be absorbed by the atmosphere at large angles from the zenith, due to the thicker atmospheric width in these directions. The above figures decouple this dependency, showing histograms of muon number with energy (left) and zenith angle (right).

spectrum of muons is strongly dependent on zenith angle, as muons with low elevation travel through a thicker atmospheric column before reaching the earth's surface, cutting off the contribution of low energy muons. Figure 2.11 shows histograms of muon abundance for different energies and angles respectively.

We simulated 10^8 μ as the primary source at the macroscopic level, created at random positions in an $r = 15$ cm disc positioned 3 cm above the microplate, with emission angles based upon the integrated differential flux formula. This is normalized to 48.7 days of real time. At the microscopic level, 10^6 particles generated at the macroscopic level were randomly drawn and tracked to measure the interactions between muons and electrons from muon-disintegrations and bacterial cells. The energy of these particles spanned the same range as the input muon spectrum, given that the energy loss of muons travelling through air is small.

2.3.5 Results

Model Selection Using a Monte Carlo particle transport simulation at such small length scales, we were nevertheless concerned that some continuous multiple scattering models employed could introduce errors into our analysis. To better understand the impact of our choice of physics models on the final simulation results, we simulated the transport of 10^6 200 keV electrons in the same geometry as the microscopic simulation, with the only change being that the repeating region continued indefinitely. The aim of this simulation was to

Table 2.7 Variation between simulation outputs for different physics models. Energy deposits are given at the 25th, 50th and 75th percentile of the distribution.

EM Physics Constructor	Cells hit	E_{dep} (eV)		
		25th	50th	75th
Geant4-DNA Option 2	541568	108	212	426
Standard Option 4	534138	152	218	414
Penelope	531943	125	222	446
Livermore	570801	114	191	396

compare different physics models, to give an estimate of the variation of the meaningful simulation outputs when different approximations of particle transport processes were used. We compared the interactions of these electrons with cells using the standard electromagnetic Option 4 physics constructor, as well as the low energy electromagnetic Penelope, Livermore and Geant4-DNA option 2 physics constructors (Geant4 Collaboration, 2015). Notably, the Geant4-DNA Physics models (Bernal et al., 2015) provide fully discretized low energy electromagnetic processes. This allows us to see whether the approximation of scattering as a continuous process, which is made in most other models, noticeably impacts our results.

As the choice of physics models used in any simulation impacts the outcome, we first present our brief comparison of physics models for the microscopic level simulation. The distribution of energy depositions in cells and the number of cells that had energy deposited in them are summarized in Table 2.7. The spectrum of energy depositions follows a Landau distribution as it is effectively a sampling of the energy deposited by a decelerating charged particle. Accordingly, we present percentiles of this distribution rather than a mean, as the mean carries little meaning for this type of distribution. Between each model, the distributions of energy deposition are significantly different (a Kolmogorov-Smirnov test between any two models shows the distributions are dissimilar to $> 5\sigma$), however the parameters relevant to our study show broad agreement between models. The number of cells hit in the simulation (those experiencing at least one energy deposit) agrees within 10% between models, and the measures of energy deposited, while being different distributions, differ by $\approx 10\%$ beyond the 50th percentile. These two observations allow us to approximate the errors in our results coming from the physics models chosen to $\approx 10\%$.

Simulation Results Table 2.8 indicates the frequency with which cells are subjected to a radiation induced energy deposition for each source we considered. The hit frequency is normalized by the total number of cells considered in the study, giving a quantity that corresponds to hits per cell per day, or alternatively, the chance that any given cell is hit in

Table 2.8 Frequency with which ionizing radiation from background sources interacts with *E. coli* cells, the median and modal energies deposited in each interaction. These quantities correspond to the surface environment at LPC.

Source	Dose Rate (nGy hr^{-1})	Hit frequency (day^{-1} $cell^{-1}$)	$E_{\text{dep, median}}$ (eV)	$E_{\text{dep, mode}}$ (eV)
γ background	150	3.6×10^{-5}	140	100
^{40}K β-decay	26	8.2×10^{-6}	120	120
Cosmic μ	45	1.6×10^{-5}	110	120
Cosmic n	4.4	1.4×10^{-7}	1.2×10^{3}	670
Total	225	6.0×10^{-5}	-	-

a one day period. The specific distribution of energies deposited in a cell per day is shown in Fig 2.12.

Studying just the dosages found using the inputs available, it is evident that the gamma background dose is elevated when compared to the standard population weighted average dose of 60 nGy hr^{-1} (UNSCEAR, 2000) . This is caused by naturally higher radiation levels at the LPC due to the soil composition. Similarly, the muon dose is 27% higher than modeled values would predict for a site at our elevation (400 m), where the predicted dose is 33 nGy hr^{-1} (Bouville & Lowder, 1988).

The total number of interactions between an *E. coli* cell per day and the radiation background at the surface is 6.0×10^{-5} day^{-1}, indicating that on average roughly 1 in 20,000 would be expected to interact with ionizing particles from the radiation background on a given day. Underground, the 6.4-fold reduction in the gamma background reduces the frequency of interactions per day for a given cell to 1.37×10^{-5} day^{-1}. Suppressing the gamma background entirely leaves only the contribution of ^{40}K, giving a 7.3-fold reduction in the cellular hit rate compared to the background of 8.2×10^{-6} day^{-1}.

To better understand the nature of energy deposition induced by each source, in Fig 2.13 we show the distribution of energy deposits per 10^6 simulation events. Energy depositions correspond with what one would expect based on the particle transport characteristics of each input source, that is to say that higher LET sources deposit energy according to a flatter Landau distribution, whilst the exact quantity of cells hit is determined by the mean distance particles would travel through the water medium simulated.

Comparing the backgrounds from the gamma background and the nutritive medium, for the same input dose, electrons from beta decay of potassium in the nutritive Davis minimal broth impact more cells, albeit with a lower median energy deposition (Table 2.9). This is consistent with the expected comportment of electrons: higher energy electrons will have

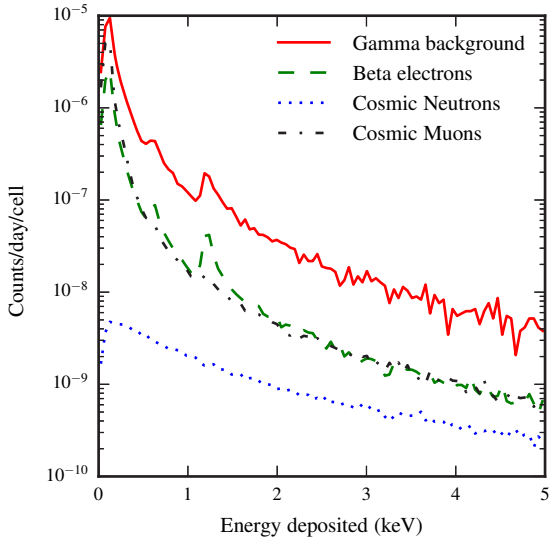


Fig. 2.12 Each source deposits energy in cells according to different Landau-like distributions. Energy depositions are normalized to the hit rate, indicating for each source the chance a specific amount of energy is deposited in it in a day. The peaks near 600 eV and 1.2 keV in the γ-background and β-electron spectra are related to the emission of one or two short-traveling Auger electrons emitted by Oxygen atoms within cells, in addition to the energy deposited by other processes.

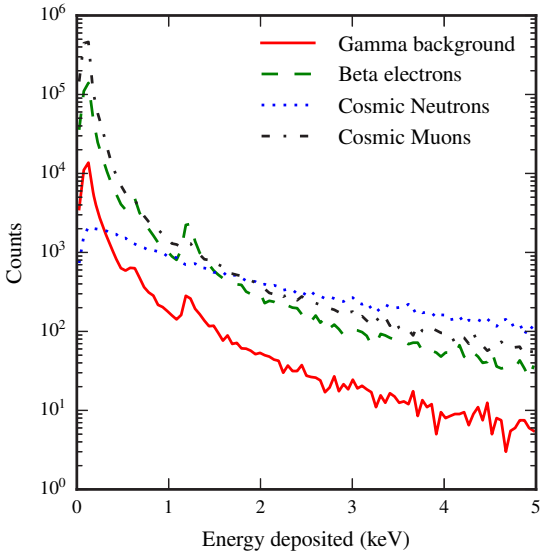


Fig. 2.13 When the energy depositions are normalized to 10^6 primary events, the characteristics of each source become clearer. Sources that travel further through the medium impact more cells, whilst the significantly higher LET from neutron-induced ions is reflected in the flatter distribution of energy deposits from this source.

Source	Expected interactions (nGy^{-1})	Mean source energy (keV)
γ background	1.0×10^{-8}	7.4×10^{-2}
^{40}K β-decay	1.3×10^{-8}	5.11×10^{-1}
Cosmic μ	1.3×10^{-8}	1.34×10^{7}
Cosmic n	1.3×10^{-9}	1.1×10^{4}

Table 2.9 Expected radiation interactions in one *E. coli* cell per unit dose from natural background sources. These quantities are sensitive to both the geometry of the cell considered and the experimental geometry.

longer paths relative to their energy given linear energy transfer is inversely proportional to a particle's energy, and the mean beta electron energy from ^{40}K is 511 keV in comparison to 74 keV for electrons created from the gamma background. Secondaries created from the neutron background, being predominately ions, tend to have very high LET values leading to both a large energy deposit when cells are hit, and a relatively small quantity of cells hit given the short path neutrons traverse.

It is interesting to note that both tracks from ^{40}K beta decay electrons and cosmic muons traverse the same amount of cells per unit dose of radiation. Additionally, the spectrum of energy depositions from muons strongly resembles that caused by beta decay electrons. This is tied to the similar LETs of 511 keV electrons (2.04 MeV cm^{-1}, c.f. Berger et al., 1998) and muons (2.55 MeV cm^{-1} for 14 GeV muons, c.f. Groom et al., 2001), however based upon this alone muons ought to interact with comparatively fewer cells per unit dose deposited in the material. The additional component in the muon interactions comes in part from a contribution to the interactions made by electrons that enter the well as the children of decaying muons.

2.3.6 Discussion I: Physics

By conducting these simulations, we have sought to quantify the impact of background radiation in biological experiments, and in doing so, guide the interpretation of the growing body of low background biological experiments. Measurements of the frequency with which the radiative background interacts with cells provides an upper bound for the size of most radiation induced effects. Such effects are not limited to genetic damage induced by radiation, as radiation may also affect cells and induce death by damaging proteins directly and through oxidation (Daly, 2012; Krisko & Radman, 2010).

Physically, the spectrum of energy depositions from each source (Figs 2.12 and 2.13) reveals a significant amount of information about the nature of the energy being deposited

within cells. Clearly evident in the spectra are peaks related to the Auger emission spectrum of water (near 500 eV), emphasizing the significance of this process in low energy cellular dosimetry. This is especially important considering the otherwise low energy deposits that occur in the absence of Auger effects. For lepton backgrounds, the modal and median energy deposits are near 100 eV per cell hit, meaning that the emission of just one Auger electron in a cell has the capacity to significantly affect the energy that would otherwise be deposited. It is worthwhile to give some consideration to the impact of these energy deposits in terms of the volume they impact through the radiolysis of water. In the domain we are interested in, a majority of the primary electron's energy is deposited in water as 'spurs' along the electron's path. Spurs are bead-like regions where a $40-100$ eV energy deposit ionizes and excites water molecules, which react and diffuse in a cloud with a diameter of 4 nm (Ward, 1988). Considering that the median energy deposit from lepton sources is at its highest 140 eV, the majority of cells impacted by radiation contain only one to three such regions. Emission of a single Auger electron, which has a maximum track length of 11 nm (Mozumder & Magee, 1966) significantly contributes to the energy deposited, and also to the volume within the cell impacted by radiation. More importantly perhaps, this energy is deposited over a well-localized region in space, while spurs can be separated by several hundred nanometers. It would be interesting to pursue simulations further to quantify the precise impact of this on cellular structures.

2.3.7 Discussion II: Interpretation for an LTEE

These simulations were conceived to help predict the outcome of our long term evolution experiment in the LSM. Between the reference radiation environment at the LPC and the reduced radiation environment at the LSM, cell cultures are grown experiencing either 6.0×10^{-5} interactions cell^{-1} day^{-1} or 8.2×10^{-6} interactions cell^{-1} day^{-1}. From these figures, we seek to determine whether the evolutionary behavior of *E. coli* ought to change between these regimes. Such an evaluation is comparatively simple: the upper bound on the point mutation rate of *E. coli* across the first 20,000 generations of a long term evolution experiment is 7.4×10^{-4} mutations per generation (Barrick et al., 2009). Given we grow 8.23 bacterial generations per day, the upper bound on the point mutation rate in our experiments is 6.1×10^{-3} mutations per day: 102 times higher than the frequency with which radiation interacts with cells. The significance of this comparison indicates that, following the assumption that radiation should not produce mutations that differ significantly in their effects to those of biological processes, in an *E. coli*-focused long term evolution experiment, the radiation background should not significantly affect the evolutionary behavior of the population, due to the relative infrequence with which it impacts bacterial cells at the surface radiation

level. With the caveat that a comparison between the daily mutation rate and radiation interactions per day is a comparison of two upper bounds, the evolutionary behavior of bacterial cells should not be significantly different in an underground environment compared to a surface-level laboratory, as the impact of the radiation background is less than 1% of that from biological processes.

The independence of radiation induced interactions, and thus, possible radiation induced mutations from the point mutation rate is not surprising. The mutation rate is a biological parameter that is subject to selection, which is optimized according to the dynamics of the host population rather than by a uniform oxidative stress. Typically, its value is limited by random genetic drift: the variation in the frequency of different alleles within a population that comes from random sampling of a population. This limit arises because while proteins could theoretically be synthesized in the cell to reduce the mutation rate, after a point this becomes disadvantageous when the cost to the organism of having such proteins does not significantly outweigh the gains from reducing the total amount of genetic variation between generations (Lynch, 2010). Certain situations may also favor the appearance of vastly higher mutation rates, as is often marked by the appearance of mutator alleles in evolution experiments (Sniegowski et al., 2000, 1997). The existence of radiation-tolerant bacteria such as *D. radiodurans*, and experiments forcing the evolution of radio-resistance in *E. coli* (Harris et al., 2009) indicate that when oxidative stresses are considerable, species evolve mechanisms to protect themselves from oxidative damage. In many ways this is both a by-product of the cell evolving mechanisms that allow it to survive oxidative stress as well as the cell attempting to select a mutation rate that is optimal for its environment, as each of these goals are mutually compatible. Nevertheless, there remains no significant reason why such a mutation rate should be particularly correlated with the radiation environment given the other forces involved in selection.

Exploring these responses in the context of controlled increases in the background radiation does present further avenues of future study. Long term evolution experiments show that a three-fold increase in the mutation rate caused by transfecting cells with a mutator gene can produce observable changes in the fitness trajectory (de Visser et al., 1999). Simulations such as those performed here can be used to determine which radioactive sources best increase the quantity of cells impacted by radiation so that this rate approaches or exceeds the mutation rate. For the sources considered here, an increase in the background rate to ~ 20 μGy hr^{-1} would be sufficient to cause the rate at which cells are impacted by radiation to be near equivalent to the point mutation rate. Whether this increased radiation level would favor mutations linked to radioprotection rather than the fitness experiment itself needs careful evaluation. One study from the Chernobyl environment showed that background absorbed

dose rates of up to 75 μGy hr^{-1} do not seem to encourage the formation of radio-resistant sub-strains (Zavilgelsky et al., 1998), however a more recent study showed that resistance to γ-radiation was augmented in bacteria living in bird feathers that grew in radiation environments only a few times above the standard background (450 nGy hr^{-1}), compared to bacteria found in feathers at both standard and significantly elevated (2.9 μGy hr^{-1}) backgrounds (Ruiz-Gonzalez et al., 2016). Controlled, long term low-dose evolution experiments could even elucidate whether different radioprotective mechanisms evolve in different radiation environments.

There remains scope for the possibility that radiation may interact with biological systems in ways that conflict strongly with the assumptions made in the preceding paragraphs. Much as one type of mutation may become more or less likely depending on the genome of a cell[3], one could propose the idea that radiation could act as a trigger for less likely mutations. Measuring and quantifying this would be challenging, however this does leave a mechanism by which the radiation background could impact the evolutionary behavior of a population. Our measurements of competitive fitness in different radiation environments are designed to elucidate this, and are discussed in Chapter 4. Even if the evolutionary behavior of a cell population shows no dependence on the radiation environment in the first thousand generations of an LTEE however, this does not eliminate the potential for radiation to play a subtler, longer term role in LTEEs. As the cell population becomes increasingly well adapted to its environment, measurements at much later generation times could potentially show a fitness dependence on radiation environment were radiation responsible for rare mutations, as the supply of non-radiation induced mutations could become exhausted. Whether this is possible is debatable, given that even after growing 50,000 generations of *E. coli*, the measured fitness of the bacteria continues to grow seemingly without bound (Wiser et al., 2013) .

2.3.8 Discussion III: Implications for Low Background Experiments

Cast in the light of other low background experiments, the relatively low frequency of interactions between the radiation background and cells challenges existing assumptions about the mechanisms by which bacterial cells have seemed to 'sense', in a relatively short amount of time (days up to a week), that they have been transferred to a low background environment. In introducing this section, we presented Castillo et al.'s measurement of impaired bacterial growth in a low background environment compared to a reference environment after just 24 hours growth with the radiation background suppressed (Castillo et al., 2015). Repeating

[3] e.g. the *mutY* allele in *E. coli* increases G:C to T:A transversions (Radicella et al., 1988)

our simulations for the experimental and cellular geometries used in their experiment[4], we estimate that the chance a radiation track deposits energy in a *D. radiodurans* cell in a day is 1.3×10^{-4}, significantly lower than the upper bound on this figure given by Katz (2016), largely due to our consideration of track structures. Note that despite the lower dose in this experiment as compared to our *E. coli* simulations, there are more interactions per day due to the larger size of *D. radiodurans* cells compared to *E. coli*, and the higher LET of secondary electrons induced from the γ-background used by Castillo et al., compared to a standard terrestrial background spectrum. Another way of interpreting these figures is then to say that Castillo et al. notice a population wide effect when only $\approx 0.01\%$ of the cell population is actually able to notice a decrease in the radiative background in a one day long period. While bacterial cells can communicate, for example through the secretion of outer membrane vesicles (Kulp & Kuehn, 2012), the emergence of a population wide effect coming from such a small fraction of the cell population is startling and warrants further investigation. While the decrease in growth rate is consistent with the hypothesis that the radiation dose response is hormetic, the speed with which the change occurs remains to be explained.

Beyond considering the rapidity with which a population level change has been reported to occur in a low background environment, it's worth also noting that the time scale on which cells interact with the background is significantly longer than their division time. Assuming for a cell that the likelihood of interacting with the radiation background scales with its surface area, an animal cell ($r \approx 15$ μm) is hit 900 times more often than an *E. coli* cell ($r \approx 0.5$ μm). Thus even larger cells are still hit relatively rarely (on average, once every 25 days). In long duration experiments across both yeast and mammalian cells, evidence of a hormetic response to radiation has been observed in low background experiments (Carbone et al., 2009; Fratini et al., 2015; Satta et al., 1995, 2002), where a small level of radiation seems to stimulate oxidative response mechanisms. Yet the mechanism by which information about very infrequent radiation energy deposition events is passed through the cell population or down the cell lineage remains to be understood.

2.3.9 Perspectives

This study in some ways poses more questions than it answers. In line with previous estimates, we have shown that the frequency with which radiation tracks interact with bacterial cells is incredibly small. That the radiation background and its related suppression has been shown to have an effect, often rapidly, on simple cellular systems is astounding, and it highlights the need for further investigation in this area. In this section, not much comment

[4]Where *D. radiodurans* cells were simulated as spheres of radius $r = 1.5$ μm exposed to 71.3 nGy hr^{-1} from an isotropic 1.46 MeV γ-ray source, and a 7.2 nGy hr^{-1} exposure from internal ^{40}K β^--decay.

has been made regarding low-dosage radiation effects, such as the bystander effect and genomic instability. This study suggests that the rapid changes observed in cellular systems to reductions in the radiation background are another unorthodox result that can be observed in low dose regimes.

Typically, one wouldn't expect an LTEE conducted at low backgrounds to bear any changes compared to one conducted at the surface. This is shown numerically in the very low frequency of cells hit per day, even at a standard background level. The odd behaviour of cells at low backgrounds suggests that this prediction is not guaranteed (the experiment itself is the subject of Chapter 4).

There is also scope to continue this investigation in simulation. Little work has been done simulating at a genome level the impact of radiation on bacterial systems. Whilst the frequency of interactions between a radiation source and a cell provides useful upper bounds on the amount of damage that could be incurred, we are currently unable to clarify the nature of this damage. Chapter 3 considers this in detail, as given an appropriate geometrical description of a cell, detailed simulations combining physics and radiation chemistry can enable the quantification of single and double strand breaks of DNA from radiation sources. Importantly, these simulations we have presented provide a means of obtaining the input spectrum of ionizing particles that interact with cells from an environmental source.

2.4 Conclusions from Dosimetry and Microdosimetry

From the work in this chapter, we have quantified the LPC and LSM environments for low background biological experiments. Radiation acts upon cells at the level of individuals, and we have shown a method which quantifies this, extending the general principle of dosimetry to a level where one can consider how a radiation environment actually impacts a cell. For *E. coli* at a standard background, the chance in a day that a cell is touched by a radiation track in a single day is 6×10^{-5}. This makes interactions between radiation and cells incredibly rare events. It's entirely possible that the rarity of these events is what creates the driver that allows cell populations to respond rapidly to them. If radiative interactions are both rare and damaging, a collective response to them that lasts across generations has some evolutionary sense.

Substantial work needs to be done to show this however, and it can advance in two directions. Showing that a particular mechanism, such as population-level sensitivity to radiation, is evolutionary favourable may be accomplished through simulations of simple cell populations via evolutionary games and cellular automata (Maynard Smith & Price, 1973). On the other hand, the magnitude of cellular damage induced by radiation passage

events needs to be better established. This is a difficult task, though experimentally it can be realised with nuclear microbeam studies looking at the appearance of γH2AX foci following irradiation, which in cells that form chromatin is indicative of double strand breaks. This can also be explored in simulation.

A number of Monte Carlo codes can already simulate the direct and indirect DNA damage induced by ionising radiation. They provide an indication of the magnitude of cellular damage by quantifying the number of single and double strand breaks induced by radiation tracks. This is can be of particular importance in low dosage simulations where the discrepancy between mechanistic predictions of biological effects, and the observed effects can best be quantified by accurate models. Currently, no open source method exists though of running these simulations. In the following chapter, we present a Geant4 application that extends the Geant4-DNA toolkit to the simulation of direct and indirect radiation damage for flexible cellular geometries.

Chapter 3

Evaluating radiation damage with Geant4-DNA in bacteria

When ionising radiation interacts with a living cell, it can trigger a range of events. Ionisation of molecules composing the DNA strand, or of water molecules very close to the DNA strand can directly damage DNA. Additionally, ionising radiation radiolyses water, producing reactive oxygen species which can damage DNA, particularly hydroxyl radicals $^{\bullet}OH$ and solvated electrons e^{-}_{aq}. Ionising radiation induced DNA strand breaks play a role in both mutagenesis and cell death, with double strand breaks, damage events triggered by two nearly coincident strand breaks often causing apoptosis when left unrepaired (Friedberg, 2003; Friedburg et al., 2005).

Quantifying radiation induced damage from different environments allows better limits on the extent to which different radiation levels can impact the evolution of simple cellular systems. More broadly, the end-goal of quantifying radiation induced DNA damage is to have better predictive models of cellular responses to radiation, built from mechanistic predictions. Work towards a complete mechanistic model of irradiation, and of subsequent biological responses, has been underway since the early 1990s. Notably, the PARTRAC project (Friedland et al., 2011) and the simulations of Nikjoo et al. (for a review see Nikjoo et al., 2016) have made significant progress in this regard. Both these projects however are closed-source, and difficult to build upon for non-experts.

In this chapter, we present a framework for modelling radiation induced DNA damage based on a geometrical DNA model as part of the Geant4-DNA project. Previous work within the Geant4-DNA project has used clustering approaches to measure DNA damage (Francis et al., 2011) from discrete Monte Carlo simulations, as well as correlating energy depositions in water with a spatial DNA model (Bernal et al., 2013). Dos Santos et al. (2013) also presented a geometric model of a whole nucleus modelled in Geant4, however all

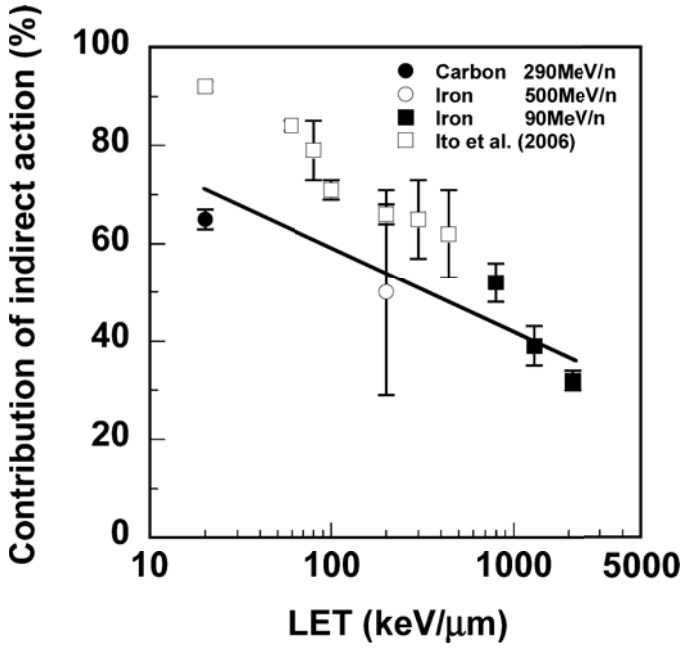


Fig. 3.1 Indirect DNA damage, that is damage caused by water-based radicals created by low energy radiation tracks, is responsible for the majority of radiation induced DNA damage in cells. Here, the contribution of indirect effects to cell death is considered, as a function of dose-averaged LET. For low LET radiation, indirect damage is a significant contributor to cell death. Indirect damage decreases as LET increases (trendline). Not shown are X-rays (with a dose- (track-) averaged LET of 9.4 (1.7) keV μm ^{1}) where indirect damage contributes at a level of 76 ± 5 % to cell death. From Hirayama et al. (2009).

these models lack an interface to the chemistry module of Geant4-DNA (Karamitros, 2013; Karamitros et al., 2011, 2014). The inclusion of chemical reactions, to facilitate the simulation of indirect damage effects, is important because indirect damage can contribute up to 80% of radiation induced DNA damage (Figure 3.1), in the case of low LET irradiation. Prior to this work, at least one program had already been developed that permits Geant4 to model radiation damage in DNA geometries, across both the physical and chemical stages of DNA damage, by linking multiple Geant4 applications that consider the physical and chemical stages DNA damage separately (Meylan, 2016; Meylan et al., 2016).

In the following section, we introduce the requirements for our simulation. Namely, we seek to create a simulation that can simulate a variety of DNA geometries, and their interactions with the physics and chemistry of Geant4. This simulation will be demonstrated with a bacterial genome, linking the simulation work here to the work in Chapter 2 and our LTEE. The requirements for the simulation application are described in Section 3.1. Section 3.2 describes the basic organisation of a bacterial genome, which underpin our geometric model. We then detail the implementation details of our simulation in Section 3.3, before

presenting a parameter study of the model in Section 3.4 as a form of benchmarking. We then apply our model to bacteria in Section 3.5 before presenting our conclusions.

3.1 Requirements for the simulation

While other applications exist to mechanistically model DNA damage in Geant4, they are tailored towards human cells, and lack a simple means of re-using their interfaces for bacterial cells. We have sought to create a simulation platform for DNA geometries integrated into Geant4 which can flexibly model arbitrary DNA types, plasmid structures and common test geometries, whilst also remaining well contained within the limits of physical memory on modern computers.

The requirements for our application can be broken into a few categories, namely: *flexibility*, the application should allow as many physical parameters as possible to be varied so that multiple values can be tested; *extendability*, the application shouldn't restrict itself to one DNA conformation; *performance*, the application needs to run on available hardware in a reasonable time frame; and *compatibility*, the application should remain compatible with Geant4.

The need for flexibility in the simulation design is assured through the use of macro based application commands in Geant4. An advantage of simulation is that it allows the impact of changing parameters on a result to be studied: this is best facilitated by having easily changed parameters. Similarly, the simulation ought to be extensible, permitting multiple varied geometries without recompiling the application. This can be assured by either having a robust geometry specification language incorporated in the application that can generate varied geometries, or by allowing data files that specify geometries to be read by the application.

In terms of performance, the simulation ought to be processor limited rather than memory limited. Simulating all 4.6 Mbp of a bacterial genome could naively require 20 GB of memory if each base pair is placed separately, which is towards the memory limit of most modern computers. Such a model is furthermore inextensible as a similar implementation of human DNA would require terabytes of RAM. Instead, the application should take advantage of the repeating patterns present in DNA in order to economise memory. Despite the need to ensure good memory performance, the strength of the Geant4 framework is its generality, and this shouldn't be compromised to implement the simulation. Therefore, the application should be written without the addition of any custom Geant4 classes.

To this end, we place a constraint on the geometry, in that it be definable in terms of simple repeating units. Via macro commands, we will specify one text file, which defines the position of these repeating units, and a number of secondary text files which describe

the position of DNA related molecules inside these repeating units. The way these files are read influences how geometries are generated, and how they are then placed inside Geant4, so for this reason we detail the file specification inside the simulation requirements. Ideally, to simulate any new simulation geometry, all that should be necessary is a series of files defined according to the standard laid out in the following section.

Specification of input file formats At least two files need to be generated in order to run a DNA level simulation using Geant4. These allow the breakdown of the DNA geometry into a large scale structure which specifies the placement of many different placement volumes. Positions are given in relative units so that spatial scalings of the geometry do not require a new definition file. Each placement volume is described by a file that contains a molecular specification of the DNA chain. Files are single-space separated text files, where commented lines are marked by a hash.

Files are selected using the Geant4 macro interface of the simulation application[1]. Two commands are provided for this. `/dnageom/definitionFile` sets the location of the file that defines the locations of placement volumes. The command `/dnageom/placementVolume` sets the name and location of each placement volume. The name chosen is the name referenced by the definition file when specifying where to put each placement volume. An optional boolean parameter allows the user to flag that the strand undergoes a 180° twist from start to end, necessary when joining two strand elements together properly.

For the large scale structure the aim of the file is to specify the position of each individual placement volume together with its rotation. Positions can be specified in units relative to the size of an individual placement volume and then scaled via the command `/dnageom/fractalScaling`. This requires a file with the following columns:

1. **Index**, *integer*, the index to be assigned to the placement volume. It should increase from zero by unity for each placement. The DNA strand is assumed to pass continuously from the placement with index i to the placement with index $i+1$.
2. **Kind**, *string*, a string specifying the name of the placement volume. Names are set for a given placement volume used when that placement volume is defined in a macro command.
3. **Position-X**, *double*, the x-position where the specified placement volume will be centered, in relative units.
4. **Position-Y**, *double*, the y-position where the specified placement volume will be centered, in relative units.

[1] A detailed guide to all commands available in the simulation is provided in Appendix A

5. **Position-Z**, *double*, the z-position where the specified placement volume will be centered, in relative units.
6. **Angle-ψ**, *double*, rotation about the x-axis to transform the global axis into the local axis, in radians.
7. **Angle-θ**, *double*, rotation about the y-axis to transform the global axis into the local axis, in radians.
8. **Angle-ϕ**, *double*, rotation about the z-axis to transform the global axis into the local axis, in radians.

Rotations are defined by the three Euler angles (ψ, θ, ϕ) about the (x, y, z) axes. As there are many definitions of Euler angles, we note here that the rotations, when applied to the identity matrix I should create a rotation matrix $R = R_z(\phi)R_y(\theta)R_x(\psi)$ where the columns refer to the local axes inside the placement volume. Because this is an internal transformation, we note that when these Euler angles are read in to Geant4, Geant4 will apply the transformation $R = (R_z(\phi)R_y(\theta)R_x(\psi))^{-1}$ to the placement volume. This convention was chosen because a reference to the local axis is useful when generating fractals using 'turtle graphics' methods as we do here.

Placement volumes are specified similarly to the large scale structure file, in a text file with space separated values, where hashes can be used to comment lines. They are placed in Geant4 inside a box with dimensions specified by the command `/dnageom/placementSize`. Rotations are defined in a similar way, where the Euler angles $(\phi,\ \theta,\ \psi)$ rotate the standard $(x,\ y,\ z)$ axes into the local $(x,\ y,\ z)$ axes of the molecule. The positions $(x,\ y,\ z)$ that are specified are the position of the molecule within the volume, in angstroms.

1. **Name**, *string*, the name of the molecule being placed, either "Phosphate", "Sugar", "Guanine", "Adenine", "Thymine" or "Cytosine".
2. **Shape**, *string*, a string specifying whether the molecule should be modelled as a circle or ellipse (as of the current implementation, this is ignored in Geant4 and all shapes are treated as ellipses, though this flag may be useful in visualisation engines).
3. **Chain-ID**, *integer*, starting from zero from the first chain, identification number of each chain when multiple chains of DNA are present in a volume.
4. **Strand-ID**, *integer*, the index of the strand, either zero or one, to identify which side of the DNA strand a molecule is on.
5. **Base Pair Index**, *integer*, the index of the base pair, starting at zero for the first base pair and increasing by one as pairs are added to the chain.
6. **Position-X**, *double*, the x-position where the specified molecule will be placed, in angstroms.
7. **Position-Y**, *double*, the y-position where the specified molecule will be placed, in

angstroms.

8. **Position-Z**, *double*, the z-position where the specified molecule will be placed, in angstroms.
9. **Angle-ψ**, *double*, rotation about the x-axis to transform the global axis into the local axis, in radians.
10. **Angle-θ**, *double*, rotation about the y-axis to transform the global axis into the local axis, in radians.
11. **Angle-ϕ**, *double*, rotation about the z-axis to transform the global axis into the local axis, in radians.

There are a few additional restrictions to observe in the structure of this file. Each molecule in a base pair will interact with those around it, and so the order of molecules is important. Each base pair needs to be written as the following six rows (noting that the distinction between strand zero and strand one is arbitrary provided it remains internally consistent):

1. Phosphate on strand zero
2. Sugar on strand zero
3. Nucleotide on strand zero
4. Phosphate on strand one
5. Sugar on strand one
6. Nucleotide on strand one

Each chain must be completed before the next one commences. The application expects the specified placement volumes to contain either one, four or eight chains (given cubes have a four-fold rotational symmetry). Additionally, the placement volumes should not include molecules placed outside the size specified by the macro command `/dnageom/placementSize`.

3.2 Generating Geometries

Before discussing the implementation of the Geant4 simulation, we turn to the generation of cellular geometries[2]. Generation of a cellular geometry can be separated into a discussion of the large scale and small scale orderings DNA takes. At a large scale, DNA forms one or multiple chromosomes, with DNA folded and compacted into the nucleus. In different phases of the cell cycle, organised chromosomes may be visible as characteristic 'X' shapes, while at other times the chromosome may resemble something of a messy ball of DNA. At

[2]It is noted here, and also later in this section that Python routines to generate the geometries described here are published online at http://github.com/natl/fractaldna. Routines have also been written to visualise these 3D geometries using the Blender open source 3D graphics suite (www.blender.org).

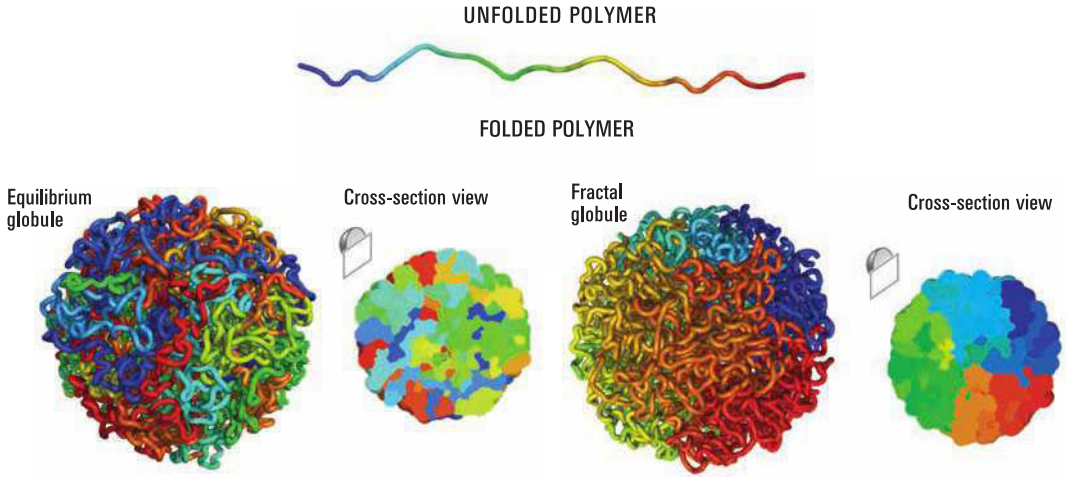


Fig. 3.2 Chromosomes fold into a geometry defined more by a space-filling curve (similar to a fractal globule), than a random, equilibrium globule. This has the effect of ensuring genes that are linearly close on the DNA polymer are also close in their folded, 3-D conformation. From Lieberman-Aiden et al. (2009).

smaller scales DNA in eukaryotic organisms folds into different conformations of chromatin, which itself possesses a level of order. Bacterial systems compact their DNA into the cell through a mixture of supercoiling at small scales and the formation of loop domains at large scales.

Whilst there can seem as though there is a lack of order to the packing of DNA, at large scales, DNA packs itself into a roughly fractal geometry. This satisfies two naïve requirements for DNA packing: first, DNA must be able to unfold itself without tangling so that it can duplicate; and secondly any packing scheme for DNA should require as little information as possible. Space-filling curves, a form of fractal, allow this, by permitting a 1-dimensional fractal curve to be densely compacted without forming knots. Additionally, space-filling curves require very little information to describe mathematically as they can be described recursively. This behaviour seems to be favoured in natural systems, and it has been shown that DNA compacts following the behaviour of a space-filling curve, which also ensures that points that are linearly close along the DNA chain are spatially close (Lieberman-Aiden et al., 2009).

It should be noted there is significantly more complexity to DNA packing at the large scale than just a space filling curve, though this serves as a good method of approximating spatial relationships between genes. As it is relevant to this work, we note that in bacteria, the packing of DNA can be described with the same ‘high-level’ accuracy as eukaryotic DNA by a space filling curve. The close localisation in space of linearly close genes on the DNA

Fig. 3.3 Polymer folding studies of *E. coli* reveal that linearly nearby segments of bacterial DNA cluster together in three-dimensional space (image from Fritsche et al., 2012). This trait is shared by space-filling curves.

has been demonstrated in relaxation simulations of DNA polymers (Figure 3.3). In bacteria, a range of different proteins are responsible for bending, binding and wrapping the DNA, as well as twisting it to enable further compaction by supercoiling. This is most active when the cell is in its exponential growth phase (Dillon & Dorman, 2010), otherwise the packing of the DNA tends to be more relaxed.

Modelling all these levels is difficult, so we commence with a simple model where the bacterial DNA follows a large scale fractal 'Hilbert curve' geometry. The Hilbert curve is broken into a series of straight and 90° turned DNA segments, each containing four DNA strands fixed distances from the centre. Different densities of DNA can be accommodated by changing the turning radius of the DNA sections.

3.2.1 Generating Fractal Geometries

Fractals are commonly found in nature as they allow dense packing to be efficiently achieved using very few instructions. Typically, recursive rules are used to generate fractals. Common methods include iterative functions, strange attractors and L-systems. L-systems in particular lend themselves well to generating path shaped fractals, and accordingly we have implemented an L-system based fractal generator in Python. The generator produces a path specified in terms of turns about an axis, and steps forward, which are then converted into rotated straight and turned 'voxels' in 3-dimensional space that are appropriate for reading by our application.

Fractal L-systems

L-systems construct fractals from strings of symbols that are iterated according to a series of rules. Certain symbols in the string carry a geometrical meaning that can be interpreted to construct a fractal. In two dimensions, a set of instructions may be as follows:

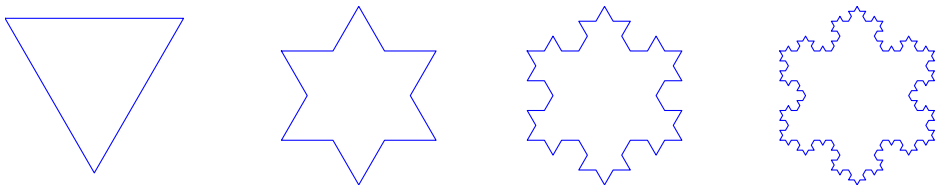


Fig. 3.4 The first, second, third and fourth iterations of an L-system fractal that constructs the Koch snowflake

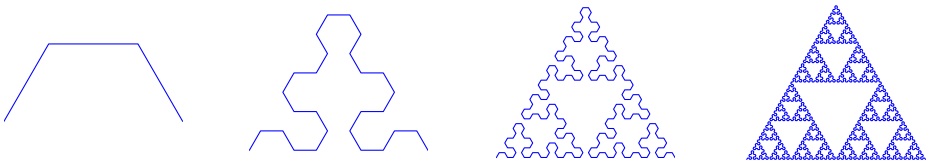


Fig. 3.5 The first, third, fifth and seventh iterations of an L-system fractal that constructs the the Sierpinski spearhead curved. This curve converges to the Sierpinski gasket.

Symbol	Instruction
+	Turn 60° right
-	Turn 60° left
F	Move forward 1 unit

These instructions belong to the language of 'turtle graphics', where shapes are drawn relative to a moving cursor on a Cartesian plane, analogous to a turtle moving along sand and leaving a trace (Françon, 1997). From these instructions, simple shapes can be drawn from a starting point. The pattern $F++F++F$ for example would signify an equilateral triangle, as the turtle moves forward one unit, then turns 120° right, then moves forward one unit again, turns again and closes the triangle. Furthermore, simple iterated rules can be used to build fractals. Using this language the Koch snowflake can be drawn starting from an equilateral triangle defined by $F++F++F$ and then applying the replacement rule $F \rightarrow F-F++F-F$. The first four iterations are drawn in Figure 3.4.

More complex shapes can be realised by expanding the alphabet used to define shapes. For example, an L-system can be defined for the Sierpinski spearhead curve, which converges towards the Sierpinski gasket (or triangle). In addition to the simple instructions for turning and moving, two instructions, $X \rightarrow YF+XF+Y$ and $Y \rightarrow XF-YF-X$ are needed. Starting from the seed $YF+XF+Y$, the fractal shape becomes increasingly evident with each iteration of the algorithm (Figure 3.5).

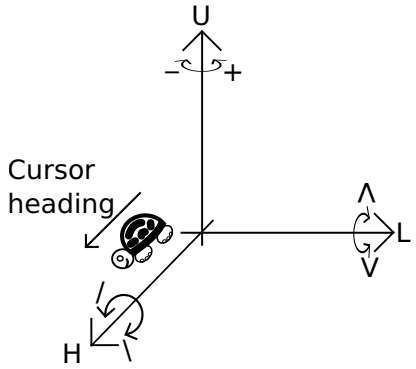


Fig. 3.6 title

L-systems can also be defined in three dimensions. Under this paradigm, the alphabet used is expanded to permit the cursor to roll, yaw and pitch, using a co-ordinate system analogous to that of an airplane, where the heading axis, local up axis, and local left axis are tracked (Figure 3.6). Using this, the instruction set for a right-angled system would have:

- **+**, yaw by $+90^\circ$
- **-**, yaw by -90°
- $\vee$, pitch by $+90^\circ$
- $\wedge$^, pitch by -90°
- ****, roll by $+90^\circ$
- **/**, roll by -90°
- |, turn around (yaw by 180∘)

A complete mathematical treatment of such a system is given by Françon (1997). Based upon the instructions above though, a Hilbert curve can be generated following the substitution rule

$$X = \wedge /XF \wedge /XFX - F \wedge XFX \vee F + XFX - FX - . \tag{3.1}$$

Such a Hilbert curve is shown in Figure 3.7. To be able to place a 3-D fractal like this in our simulation though, it needs to be converted into a series of turned and straight repeating units. We refer to this process as 'voxelising' a fractal, as we convert it into discrete, spatially located 3-D units akin to the voxels found in computer graphics (Figure 3.8). From this, a text document listing a series of turned and straight volumes to be placed can be generated, with each figure having a placement and a rotation.

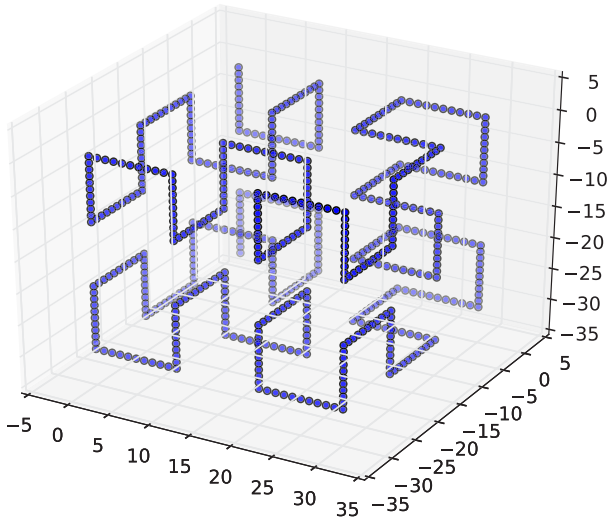


Fig. 3.7 A Hilbert curve defined by the 3D L-system presented in Equation 3.1.

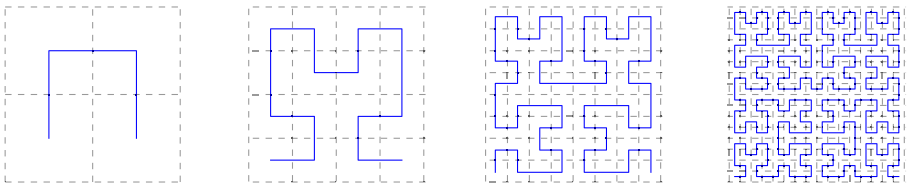


Fig. 3.8 A space-filling curve can be broken into a series of square regions containing either a straight or curved section of DNA. This structure of repeating turned and straight segments is well suited to a Geant4 simulation geometry.

Voxelising a fractal

The algorithms to generate the voxelised fractal have been written in Python[3]. They permit an L-system defined fractal to be converted into a sequence of turned and straight segments. In particular, this is done by keeping each placement box keeping track of its internal axis. The rotation necessary to convert this internal axis to the global axis defines how a turned or straight DNA segment box must be rotated to be placed in its correct orientation in the simulation. In accordance with the formats specified in 3.1, the Euler angles to be read are produced also for each voxelised element.

3.2.2 Generating DNA Volumes

The curved and straight sections in the voxelised fractal representation of DNA now need to be generated. In each placement, we built cubic placement regions containing four straight or four turned DNA regions[4]. Two stages of modelling were required for this, the first involves modelling a base pair of DNA as consisting of a collection of molecules, rather than discrete atoms. The second involves building chains out of these base pairs.

The locations of atoms in base pairs of B-DNA are well defined, and we base our approximation of the double helix structure on measurements of the molecular positions of DNA constituent elements made by Arnott & Hukins (1972). In DNA damage modelling however, the positions of the DNA constituent molecules is of greater importance than the positions of individual atoms, so it is necessary to extract a molecular position and volume estimate from the constituent atoms. DNA can be considered as being composed of phosphate (H_3PO_4) and deoxyribose ($C_5H_{10}O_4$) molecules forming a backbone that supports the nucleotide bases guanine ($C_5H_5N_5O$), adenine ($C_5H_5N_5$), cytosine ($C_4H_5N_3O$) and thymine ($C_5H_6N_2O_2$). We approximated the phosphate and sugar molecules as spheres, whilst the bases, due to their flatter shape were interpreted as ellipsoids.

A given base pair molecule has its position approximated by its van der Waal's radius-weighted mean position, so that larger atoms are more important in determining the atoms position than smaller ones. In the case of base pairs, the ratios of the major axes were determined by the ratios of the maximum extents along the cardinal axes of the constituent atoms. The semi-major axes (and the radii in spherical molecules) were then set so that the molecule had the same volume as that of its constituent atoms (Figure 3.9). The volumes of the constituent molecules were calculated based on the position and van der Waal's radius

[3]They are available both with the application, and at the Git repository http://github.com/natl/fractaldna.

[4]These Python routines are also available at http://github.com/natl/fractaldna alongside routines to generate placement volumes containing one individual strand of DNA or eight individual strands. A chromatin-like geometry has not yet been developed.

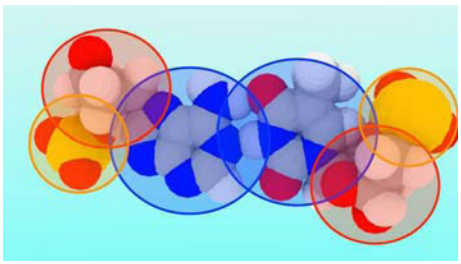


Fig. 3.9 We represent a base pair of DNA as six molecules: two phosphate molecules (yellow), two deoxyribose molecules (red) and two nucleotide bases (blue). The molecules are based on the positions and sizes of their constituent atoms (Hydrogen in white, Carbon in grey, Nitrogen in blue, Oxygen in red).

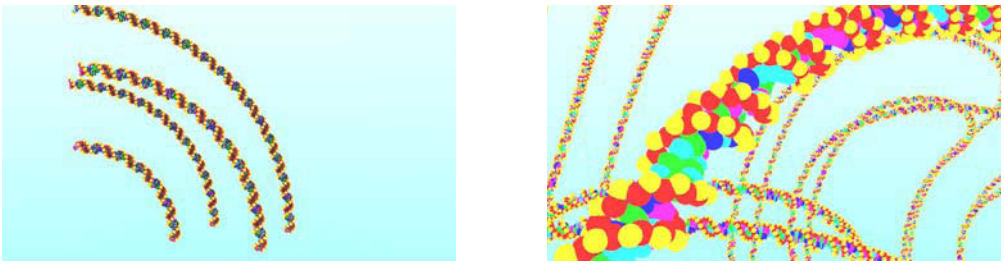


Fig. 3.10 Left: Four turned strands of DNA make one placement element. The size of the element can be varied to change the DNA density. Right: A small section of a space filling curve made of repeating curved and straight elements. When repeated enough times, this can seed a genome.

of each atom in the molecule (Bondi, 1964; Kammeyer & Whitman, 1972), taking into account double and triple overlaps between atoms (following the method of Gibson & Scherage, 1987).

From here, chains of DNA can be built based on the 3.3 Å separation between base pairs, and their 34° turn per base pair. Multiple strands were placed a set distance from the centre in the case where four or eight strands were placed in a placement volume. In these cases, a four-fold symmetry was required so that strands would be continuous between boxes. Rotational transformations were also made to bend the DNA by 90° for turned segments, and twist the DNA by 90° along its long axis to ensure strands joined continuously. Four DNA strands in a single turning placement volume are shown in Figure 3.10, rendered with Python based visualisation scripts in Blender. Also pictured is a segment of DNA in a larger fractal DNA structure.

The Python package for generating these geometries provides its own documentation on their use, as well as scripts to aid visualisation. When designing curved geometries, it is worth bearing in mind the persistence length of DNA (about 50 nm) as this determines how rapidly DNA may realistically fold back on itself without the aid of folding proteins. We

now turn to the implementation of the molecular level DNA simulation, which reads these geometries and measures radiation-induced DNA damage in them.

3.3 Implementation of the molecular level DNA simulation

Our Geant4 simulation to model DNA damage has a number of sections, linked to geometry, chemistry, scoring and analysis. Here, we describe these sections, giving an overview of the theory that motivates them. In doing so we discuss some of the technical details of their implementation, though an attempt is made to focus on methods and motivations rather than technical specifics.

Under geometry, we consider two key elements. First, we consider the chromosome interface which is used to define regions of interest and where DNA can validly be placed. In this way, a generic square fractal can be made to fill an elliptical, spherical or rod-shaped volume. Next we consider the steps necessary to read the geometry files that were described in Section 3.1, and build a geometry. We then consider the changes to the chemistry in Geant4 that were necessary to enable this simulation, and to allow the chemistry module to communicate with the simulation geometry. Finally we talk about how we model DNA damage, and how this is implemented in scoring and analysis.

A summary of the classes defined is provided in Appendix B, as sometimes these classes are referenced in text.

3.3.1 Chromosome definition and structure

Before the fractal DNA geometry is built in Geant4, a series of regions of interest can be defined in the macro by the user, which serve as a proxy for defining chromosomes. Their flexibility and definition independent of the geometry can allow the same geometry file to model several different chromosomal geometries and configurations if the user so wishes. In particular, this allows results to be recorded only in regions that the user specifies, and allows such regions to be broken down into smaller sub regions. As a result, DNA geometries are only ever placed in volume that is occupied by a chromosome (Figure 3.11).

Chromosomes are defined in the user supplied macro file based on a simple one line input structure (See Appendix A). We have implemented spherical, elliptical, cylindrical and rod shaped chromosomes thus far, though new chromosomes can be added by inheriting from the **MolecularVirtualChromosome** class, and adding an interpreter for the chromosomes specification in **MolecularChromosomeFactory**. Typically, fractal geometries are square, as spherical shapes are not very good seeds for repeating structures. Defining chromosomes

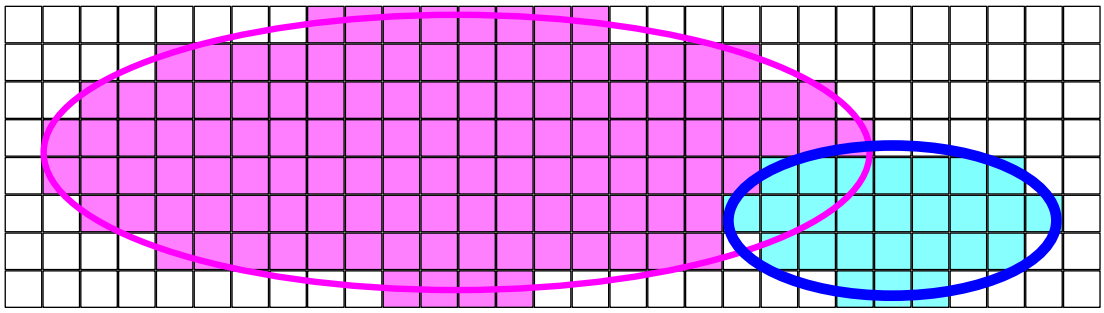


Fig. 3.11 Chromosomes are defined independently of the geometry and serve as a means of implementing geometries. Only placement boxes with their centre inside a chromosomal region are placed in the simulation and count towards damage (cyan/magenta shaded boxes). In the case of overlapping chromosomes, the first chromosome defined in the user macro file (blue in this case) takes precedence over any overlapping chromosomes for assigning damage.

facilitates the use of more traditional DNA structures, as DNA is only placed when the centre of a placement volume lies inside a chromosome. This entails some loss of continuity in the DNA strands we simulate, however the effects of this are small.

3.3.2 Reading in geometries

The geometry that we read in is based on the inputs established in Section 3.1, and requires a number of steps to be realised. First, individual placement volumes need to be constructed, which are filled with DNA. Both local rotations of the base pair need to be considered in addition to the position of the base pair in relation to other molecules near it. Also, an effective way of spatially searching DNA molecules is necessary for both the physical and chemical stages of the simulation, as physical damage models and chemical reaction models require a knowledge of nearby molecules, independent of the radius of the placement volume for a given molecule. After the placement volumes have been built, the fractal geometry can be built, and data structures are needed to let the application keep track of which DNA strands are continuous.

Use of parallel worlds The Geant4 chemistry module has difficulty dealing with complicated geometries due to dissociation processes, which can place the products of the molecular dissociation of an energetic molecule way from the dissociating molecule. To avoid having too many geometrical boundaries in our simulations, all the physical volumes are placed in a separate parallel world, using the layered geometries offered by Geant4 (Enger et al., 2012). Thus, the physically placed DNA molecules described in this section are only seen

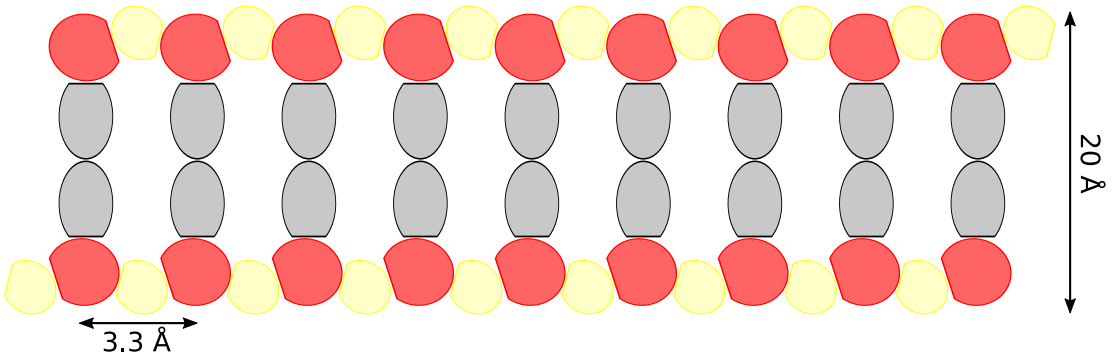


Fig. 3.12 We represent a base pair of DNA as six molecules, two deoxyriboses (red), two phosphates (yellow) and two base pairs (grey). They are modelled internally in Geant4 as ellipsoids, cut along their z-axis.

by physical processes, and their boundaries are effectively ignored by chemistry. Chemical reactions are able to look up nearby molecules using an octree data structure, eliminating any navigational problems that could arise from placing a complicated geometry in the chemistry stage of the simulation.

Reading in DNA placement volumes

When placement volumes are read into the Molecular DNA application, a few changes to them need to be made to eliminate any overlaps between the volumes. These changes are designed to satisfy the following criteria:

- Phosphate molecules are aligned to point to the next sugar molecule in the backbone. They are cut along this axis so they do not overlap the following deoxyribose molecule.
- Deoxyribose molecules are aligned to point along their respective DNA backbones to the next phosphate molecule in the chain.
- Base pairs are oriented along the axis running from their centre position to their adjoining deoxyribose molecule. They are cut along this axis so as not to overlap neither the deoxyribose molecule nor their complementary base pair. They are shrunk along the DNA's long axis so that their height along this axis never exceeds 1.7 nm, thus preventing two adjacent molecules from overlapping.
- The first and last molecules in a placement volume (which would otherwise extent beyond the boundary of the volume) are oriented to face the placement volumes wall along their z-axis. They are then cut along this axis so as not to overlap the volume boundary.

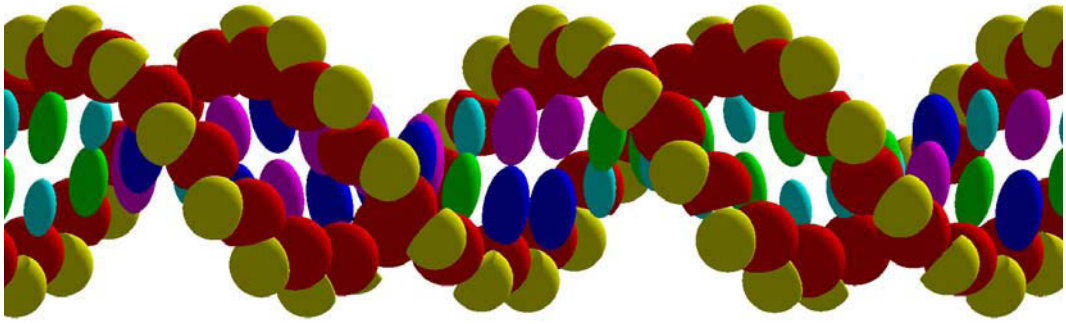


Fig. 3.13 OpenGL render of DNA arranged DNA placements inside the Geant4 viewer. Note that the major and minor curves of the DNA molecule are visible. (Yellow: Phosphate; Red: Deoxyribose; Green: Guanine; Cyan: Cytosine; Blue: Thymine; Magenta: Adenine.

These criteria are shown schematically in Figure 3.12, and their appearance in Geant4's OpenGL viewer is shown in Figure 3.13.

Octrees for Rapidly Localising DNA Molecules

When energy is deposited, we want to rapidly see if it has occurred in the vicinity of a DNA structure or not. To do this, the molecules near to any energy deposition need to be quickly found. In three dimensional space this can be easily done using octrees, which are a tree data structure that divides any given region (node) into exactly eight smaller regions (nodes). An octree is made for each different placement volume, and is stored in a map based on it's memory address. The octrees contain a position based record of all physical volumes they contain. As tracks in Geant4 know their location in a hierarchy of physical volumes, they can identify whether the current volume they are in, or one of their parent volumes, possesses an octree.

This allows energy depositions an arbitrary distance from the base pair molecules to be assigned to base pairs (Algorithm 1). This is important as not all damage models assume that strand break-causing energy depositions occur inside a molecule's van der Waal's radius. Furthermore, this is important in chemistry simulations, where even distant molecules must be locatable in order to identify possible chemical reactions.

Arranging placement volumes

Having built the placement volumes they are physically placed in the simulation according to a separate definition file, commonly built from a fractal. The Euler angles in the input file specified in Section 3.1 are converted into a Geant4 Rotation Matrix by the following operation, which takes into consideration the different specifications of Euler angles used in

Algorithm 1 Assigning energy to a molecule via an octree

```
if energy_deposited > 0 then
    octree = get_octree(energy_position)
    molecules = octree.get_nearby_molecules(en_position, rad)
    min_distance = rad
    closest_molecule = None
    for molecule ∈ molecules do
        if get_distance(molecule.position, en_position) < min_distance then
            min_distance = get_distance(molecule.position, position)
            closest_molecule = molecule
        end if
    end for
    if closest_molecule ≠ None then
        assign_energy_to_molecule(closest_molecule, energy_deposited)
    end if
end if
```

generating our geometries and Geant4 internally.

$$R_i = R_z(-\phi)R_y(-\theta)R_x(-\psi), \tag{3.2}$$

A placement volume is put into the parallel world containing the physical description of DNA, and an empty placement volume containing only water, but having the same rotations and translations as the parallel world placement, is put into the main simulation world.

Permitting the macro-structure to be reassembled A significant problem arises in these simulations when attempting to join together damage that occurs across multiple placement volumes. Each base pair needs to know its location in a larger chain, that extends beyond the placement volume it resides in. This is not difficult to implement when a single continuous chain is used in each placement volume, as a data structure can easily store the index of the base pair which starts each placement volume, but as our bacterial DNA model builds itself from up to eight chains in each placement volume, each with differing numbers of base pairs, eight indices are necessary. For each placement made, a value is assigned to each of the possible eight strands that globally may be labelled from 0 to 7, containing the index of each base pair at the point of entry into the volume. Additionally, depending upon the rotation of the placement volume, any of the four (or eight) DNA strands interior to the placement volume could correspond to any other global strand. This is illustrated for the two dimensional, two-strand case in Figure 3.14.

To identify base strands, we defined in the **MolecularDNAGeometry** class a member

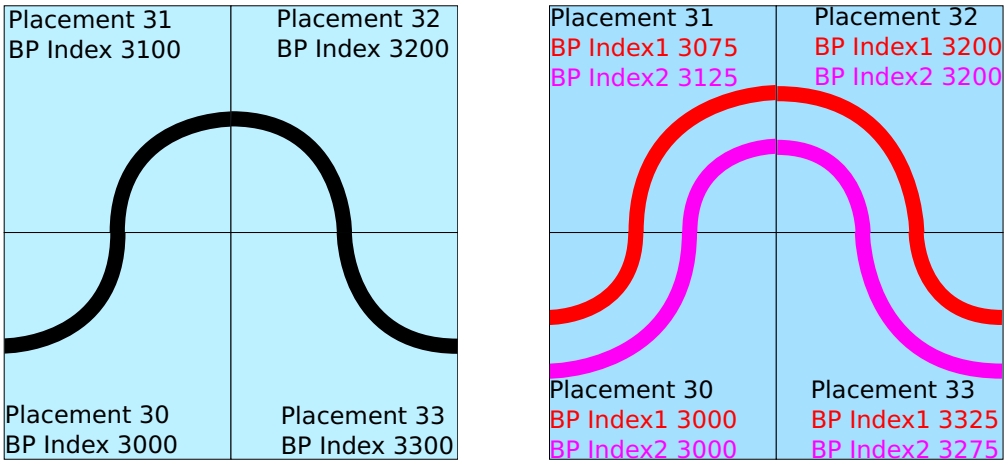


Fig. 3.14 The index of the base strands that are being counted needs to be recorded for each placement volume so that damage that occurs on the boundaries on placement volumes can be correctly identified. This is simple in the case of a single strand, but more complicated in the case of a placement volume containing multiple DNA strands, where the index has a dependency on the strand number, and additionally, where each strand locally (here, the long and short strand) need to know globally whether they belong to strand 1 or strand 2.

vector *fPlacementTransformations*, where each element corresponds to a placement volume specified in the input placement volume file. The vector is indexed by the placement index specified in the input file, and each element contains a tuple with three elements. These are 1) the global strand ID in an 8 element array, indexed by the local strand ID; 2) the indices of the first base pair of each strand in the placement volume, in an 8-element array indexed by the global ID of each strand; and 3) the indices of the last base pair of each strand in the placement volume, in an 8-element array indexed by the global ID of each strand.

To fill this data structure for a placement volume with index i, it is necessary to know the transformation between the local and global chain indices for the placement volume $i-1$, and the rotations of both these placement volumes. The other prior information necessary is that DNA placement volume segments that are straight run along the z-axis from $-z$ to $+z$, whilst segments that turn always enter the placement volume at it's $-z$ face and leave from its $+x$ face. If one considers the 3-d axis running along the DNA strand in each of these cases, two scenarios are possible. In the first case, the DNA chain is straight, and the axis moving along the DNA strand doesn't change moving up the strand. In the second case, the axis is rotated by $+\pi/2$ around the local positive y-axis.

To express this mathematically, first the rotation matrices of each volume must be found. As we are considering the position of the rotation axes within each volume, the rotation matrix R_i for the i-th element

$$R_i = R_z(\phi)R_y(\theta)R_x(\psi), \tag{3.3}$$

will contain the local axes, as has been specified for input files. This means that the first, second and third columns of this matrix contain the unit vectors for the local x, y and z-axes. The transformation T of the axis along the chain in the placement volume $i-1$ depends on the strand geometry. In the case where the DNA segment is straight $T = I$, however in the second case, we define the transformation using a rotation matrix built from a $\pi/2$ rotation about the y-axis local to placement volume $i-1$, $R_{ii-1,y}$:

$$T = R(\frac{\pi}{2}, R_{i-1,y}). \tag{3.4}$$

This leaves us with two, new, rotated sets of axes A as follows:

$$\begin{aligned} A_{i-1} &= TR_{i-1}I \\ A_i &= R_iI \end{aligned} \tag{3.5}$$

By construction, each of these two sets of axes will have the same z-axis, and we want to find the angle by which the set of axes A_i has been rotated from the axes A_{i-1} about it's local z-axis. This comes from demanding the transformation between A_{i-1} and A_i, which we call M:

$$\begin{aligned} A_i &= MA_{i-1} \\ R_iI &= MTR_{i-1}I \\ M &= R_{i-1}^{-1}T^{-1}R_i. \end{aligned} \tag{3.6}$$

As mentioned, M is a rotation about the local z-axis. The global axis corresponding to this rotation can be found then by decomposing the matrix into a vector and an angle. Geant4 provides this through the `getDelta` and `getAxis` methods of rotation matrices. An idiosyncracy of these methods however is that the final rotation $\delta \in [0, \pi]$. The proper quadrant can be identified from the rotation axis of $\vec{M}_r$, specifically we use the fact that if the rotation is positive, then

$$\vec{M}_r \cdot \vec{R_{i,z}} = 1, \tag{3.7}$$

where $\vec{R_{i,z}}$ is the third column of the rotation matrix R_i. If the quantity in Equation 3.7 is negative, the rotation δ_0 obtained from interpreting the rotation matrix M can be transformed into a positive rotation by $\delta = 2\pi - \delta_0$.

Knowing which global strand index every local strand index corresponds to in each placement volume allows the final base pair index for each global strand to be obtained, given that each local strand is of a fixed length, stored in the **MolecularPlacementVolumeInfo** class. The **MolecularDNAGeometry** class contains the method *AddNewPlacement* to facilitate the addition of new placement volumes, with the assumption that where placement volumes are meant to be joined, they are added one after the other along a continuous strand.

To add the supplementary information for each placement volume, *AddNewPlacement* is called with the pointer to each the placements logical volume, the four element array generated previously mapping each local chain to its global chain, and a boolean specifying *true* if the strand undergoes a half-twist between entering and leaving the volume. Using this information, the information for the placement volume pertaining to its base pair indices is added to *fPlacementTransformations*.

The **MolecularDNAGeometry** class then exposes three public methods to obtain the base pair information from each placement volume. These require the placement index of the physical volume of interest and either the global or local chain index, depending on the function being called. The methods are listed for clarity:

```
int GetGlobalChain(placement_idx, local_chain_idx);
long long GetStartIdx(placement_idx, global_chain_idx);
long long GetEndIdx(placement_idx, global_chain_idx);
```

In the simulation, these can be accessed within a Geant4 ‘Stepping Action’ class, in order to recover the position of each molecule on its DNA chain. This allows us to ensure that all DNA base pairs are correctly joined together, permitting damage that occurs on the edges of placement volumes to be identified and correctly classified.

Building unique identifiers for each molecule In order to track chemical reactions, and localise nearby damage on the DNA strand, all molecules placed need to be uniquely identifiable. Geant4 does not afford users many ways to attach arbitrary information to physical volumes, so we use the physical volume name to hold an index detailing the molecule type, and its position along a given DNA chain. Each larger DNA region placement volume also contains an index in its name, which is used combined with the name of each molecule to yield a unique identifier for each molecule. From this unique identifier, every positional characteristic for every DNA molecule regarding its location in the DNA chain can be reconstructed. Additionally, as this identifier is unique, it can be used to flag which molecules

Table 3.1 Chemical reactions defined between the DNA bases and deoxyribose-5-phosphate, and radicals (From Buxton et al., 1988)

Reaction	Rate (L mol^{-1} s^{-1})
$^{\bullet}OH$+ adenine	6.1×10^9
$^{\bullet}OH$+ thymine	6.4×10^9
$^{\bullet}OH$+ guanine	9.2×10^9
$^{\bullet}OH$+ cytosine	6.1×10^9
$^{\bullet}OH$+ $C_5H_{12}O_7P$	1.8×10^9
e^-_{aq}+ adenine	9.0×10^9
e^-_{aq}+ thymine	1.8×10^{10}
e^-_{aq}+ guanine	1.4×10^{10}
e^-_{aq}+ cytosine	1.3×10^{10}
e^-_{aq}+ $C_5H_{12}O_7P$	1.0×10^7
$H^{\bullet}$+ adenine	1.0×10^8
$H^{\bullet}$+ thymine	5.7×10^8
$H^{\bullet}$+ guanine	–
$H^{\bullet}$+ cytosine	9.2×10^7
$H^{\bullet}$+ $C_5H_{12}O_7P$	2.9×10^7

have already participated in chemical reactions, preventing them from participating in future reactions.

3.3.3 Implementation of Chemistry

Given the large spatial scale of a nucleus, a new implementation of Geant4 DNA chemistry was used in these simulations based on an Independent Reaction Time (IRT) model[5]. Previous chemistry implementations followed a 'step-by-step' approach, which solves the Smoluchowski diffusion equation adaptively in time, for a series of fixed minimum time steps. A Brownian Bridge in this model is used to ensure time steps do not skip chemical reactions (Karamitros et al., 2014). The IRT model identifies all possible reactions that could occur in a reasonable amount of time, and then calculates the marginal distributions of reaction times for these reactions (see Green et al., 1990). From this distribution, reactions are selected starting from the earliest reaction to the latest. In this manner, the dynamics of the simulation are greatly accelerated as they are driven by a time stepping based less on determining the paths taken by molecules, and more based on when they will react.

[5]The implementation of the IRT chemistry model in Geant4 is the exclusive work of Mathieu Karamitros.

This model was combined with the geometry via a new template class. The template class queries the geometry every time step to find which DNA molecules are nearby and available to participate in chemical reactions. This is done by querying the octree belonging to a given chemical track's physical volume. From this, the unique ID's of each molecule can be constructed, and the material of the molecule found.

The material of each placed molecule is necessary as it is used to calculate the reaction rate between the radicals simulated and the DNA molecules (Table 3.1). We consider the reaction rates for the $^{\bullet}$OH, H$^{\bullet}$ and e^{-}_{aq} radicals with the four bases and the sugar phosphate molecule. Following Buxton et al. (1988), it is possible to model either the reaction of radicals with deoxyribose as an isolated molecule (here, the reaction rate is 2.5×10^9 L mol^{-1} s^{-1}), or as deoxyribose-5-phosphate. In DNA, the bonds between deoxyribose and adjoining phosphate molecules play a role in altering its reaction chemistry, with the affect of reducing its reaction rate. We follow the approach taken by Kreipl et al. (2009), which considers one deoxyribose-5-phosphate molecule at each side of the base pair, providing a more realistic interpretation of the underlying chemistry. Other approaches that have modelled only deoxyribose have added an empiric adjustment to the chance of a reaction between radicals and deoxyribose to account for the fewer available reaction sites (Meylan, 2016).

In order to further accelerate the chemistry, we implemented a method of killing chemical tracks that are unlikely to contribute to biological damage. This has been done in the past by Nikjoo et al. (1997), who did not simulate $^{\bullet}$OH radicals more than a certain distance from DNA strands. We implement the same procedure, allowing the distance at which radicals are killed to be specified by the user. Additionally, all radicals outside a DNA placement volume are killed, as they are also far from DNA, and unlikely to cause physical damage. This also acts as a crude way of mimicking the effects of radical scavengers, however a better implementation of scavenging would allow for the medium to react with radicals, based on the scavenger concentrations found in cells.

3.3.4 DNA Damage Model

Mechanistic DNA simulations are dependent upon a DNA damage model to relate energy depositions close to DNA, and chemical reactions with DNA to actual DNA damage. Such models contain three components. The first relates energy depositions by physical process close to DNA to single strand breaks (SSBs). The second relates chemical reactions with DNA molecules to chemically induced SSBs, and the third component identifies the complexity of strand breaks, identifying double strand breaks (DSBs) from nearby SSBs.

We have attempted to allow the parameters of the damage model to be selectable by the user, and we have explored their impact upon simulation outputs in Section 3.4. This

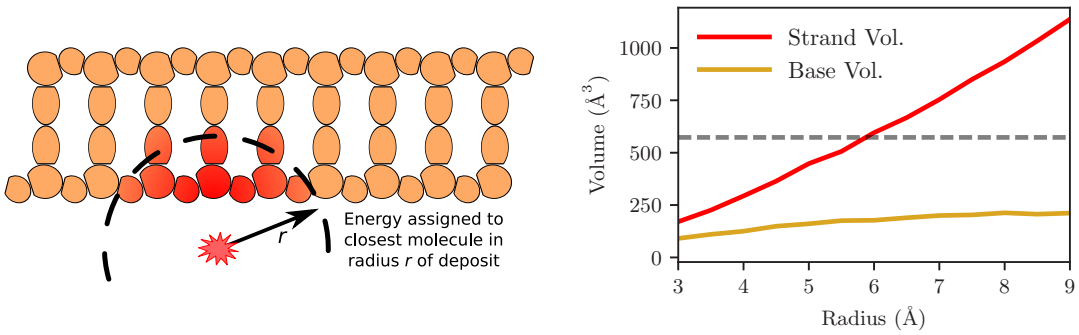


Fig. 3.15 Left: Physical energy deposits are assigned to the closest DNA physical volume within a user specified radius, *r*. Right: Variation in the sensitive volume of DNA strands (sum of phosphate and sugar molecules) and DNA bases as *r* changes. The dashed line is the volume of the semi-annulus considered as the sensitive region by Charlton & Humm (1988).

allows a comparison to existing platforms, which evaluate the processes that can lead to strand breakages from measured damage in different ways, and also allows assumptions about damage to be challenged, by seeing how different parameters can effect results. The parameters that are changeable, and their associated macro commands, are presented in Appendix A.

Scoring Physical Damage

Physical damage is scored by associating all energy deposited in a region around a DNA molecule with that molecule. When enough energy has been accumulated in this region, a strand break is said to have occurred. Fixing a strand break energy is difficult however, as numerous processes can allow low energy electrons to damage DNA, each with a different activation energy and likelihood (e.g. Barrios et al., 2002; Boudaiffa et al., 2000). Additionally, the region of space which constitutes a phosphate, deoxyribose or base molecule for the purposes of measuring DNA damage is difficult to define, and will be correlated to the energy chosen as a threshold for DNA damage. Nikjoo et al. (2016) commonly defines the region of a strand sensitive to physical breaks as a semi-annulus, within which a 17.5 eV energy deposit causes a strand break, based on a model originally used by Charlton & Humm (1988). The PARTRAC code (Friedland et al., 2011) instead determines probabilistically whether a break occurs, with the chance of a break being induced by physical processes increasing linearly with the energy deposited near the strand. Energy here needs to be deposited in the hydration shell of the DNA strand, or the strand itself, calculated on a per molecule basis as the van der Waal's radius of a given molecule either doubled or increased

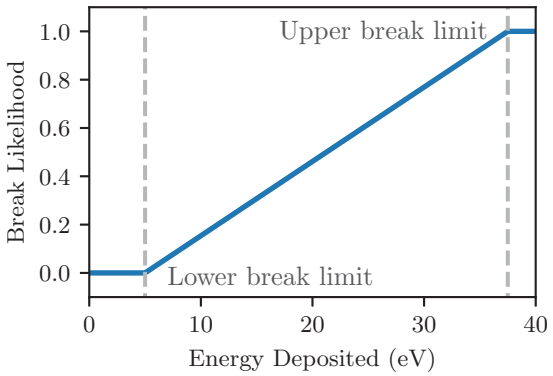


Fig. 3.16 The amount of energy needed to induce a strand break is set via two parameters, a lower and an upper break limit. The probability of a strand break varies linearly from zero to one between these limits. The values plotted here (5 eV and 37.5 eV) are taken from the PARTRAC simulation platform (Friedland et al., 2011).

by 1.6 Å.

We define the region in which energy deposits are assigned to DNA molecules based on a single distance value. The energy from a given energy deposit is always assigned to the closest sugar-phosphate moiety or base molecule, provided their is a molecule within this radius (Figure 3.15, left panel). Also, for the purposes of calculating the chance of a break, the sugar and phosphate molecules are considered together. By using the closest molecule in our criteria for energy depositions in DNA, we do introduce a little bit of complexity into the model when compared to other models, as the volume of DNA then sensitive to ionising radiation does not vary simply with the distance parameter the user can set. To allow a comparison to other platforms, in particular the semi-annulus model used by Charlton & Humm (1988), we show in Figure 3.15 (right) the variation in the volume of both the DNA strand (phosphate and deoxyribose) and base pair regions with the distance parameter, calculated by Monte Carlo integration.

Both a minimum and maximum energy are able to be specified for the probability that direct energy depositions in these regions causes a strand break. This allows a simulation of damage following a PARTRAC-like model, and also following a fixed energy limit (Figure 3.16).

Scoring Chemical Damage

Chemical damage is scored when chemical reactions take place between radicals and DNA molecules. Beyond the chemical reaction rates themselves, a number of parameters have been used to determine whether a strand break follows a chemical reactions. Measurements

(Balasubramanian et al., 1998) and simulations (Aydogan et al., 2002) show not all sites on the deoxyribose-5-phosphate component of the DNA strand react equally with $^{\bullet}$OH radicals, and that the DNA sequence can itself modulate the likelihood of SSB induction (Sy et al., 1997). Meylan (2016) choose to simulate this by assuming that certain reaction sites are blocked, reducing the chance SSBs can be formed, while other authors consider a strand break efficiency that is informed by measurements of the rate of the reaction DNA $+ \, ^{\bullet}$OH $\rightarrow$ SSB (e.g. Milligan et al., 1993; Udovicić et al., 1994). Our damage model allows the likelihood that a chemical reaction with a strand or base component of DNA proceeds to an SSB or base damage to be set by the user, as values found in the literature range from $p = 0.42$ (Meylan, 2016) to $p = 0.7$ (Kreipl et al., 2009). It is also possible that a strand break is induced by chemical damage on the base caused by $^{\bullet}$OH. In an attempt to provide a consistent interface for these secondary parameters for chemical damage, we introduce four variables that may be set for the various radicals e^-_{aq}, $^{\bullet}$OH and $H^{\bullet}$. These variables control:

- The likelihood that the radical interacting with a DNA strand molecule results in the induction of an SSB.
- The likelihood that the radical interacting with a DNA base molecule damages the base (beyond what the free reaction rate would suggest).
- The likelihood that, following an interaction between a base and a given radical, an SSB is induced, which may arise physically for example via bond breakages following resonant electron attachment to bases (Boudaiffa et al., 2000).

These variables are designed to give users the ability to control for known effects that can impact the efficiency of strand break formation. Care needs to be taken to make sure they are not used to tune a simulation to a desired outcome, but are somewhat motivated by experimental observations.

For this work, we consider that base damage never induces an SSB, and that all reactions between radicals and bases produce base damage. We also treat all radicals identically, rather than considering the different impacts of e^-_{aq}, $^{\bullet}$OH and $H^{\bullet}$ unless otherwise stated. With these two conditions, we replicate previous mechanistic studies of DNA damage, whilst also arriving at a preliminary quantification of the impact of the $H^{\bullet}$ and e^-_{aq} radicals in DNA damage. This is possible as the reaction rate between deoxyribose-5-phosphate and $^{\bullet}$OH is significantly higher than that with $H^{\bullet}$ and e^-_{aq}, and thus whether simulated or not, strand breaks caused by $^{\bullet}$OH vastly outnumber those caused by any other radical.

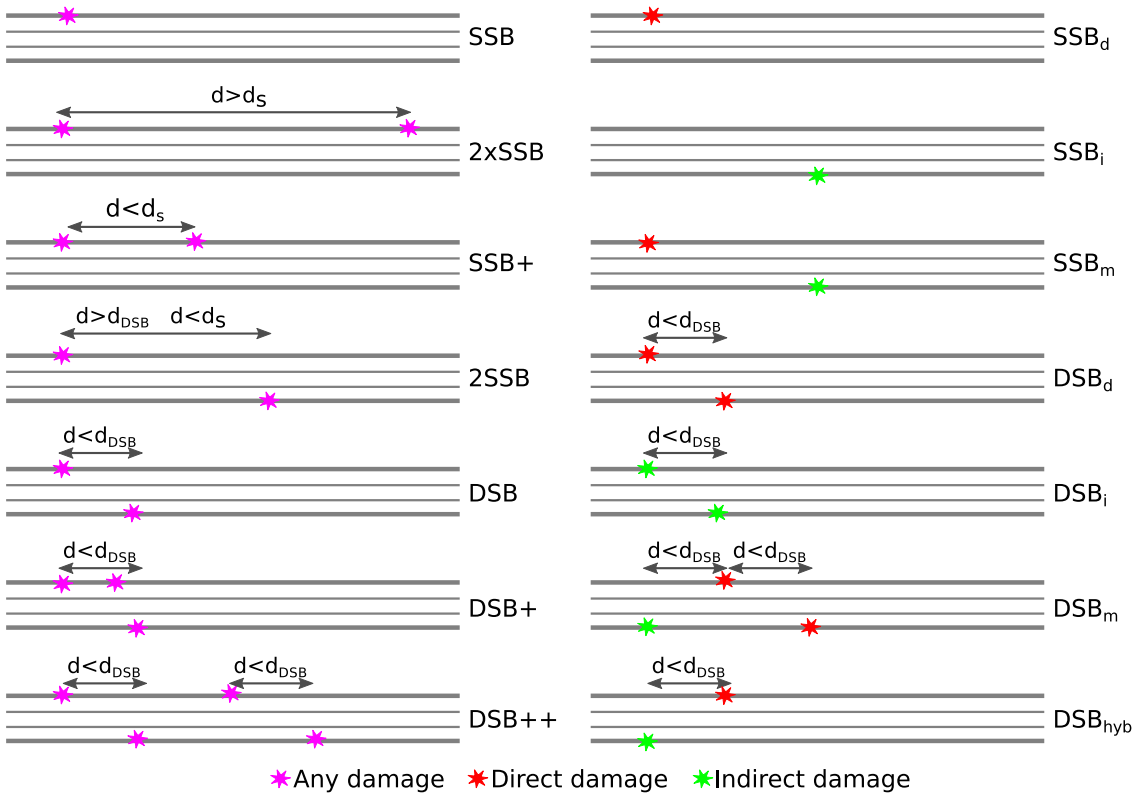


Fig. 3.17 We follow the strand breakage scheme presented by Nikjoo et al. (1997). Breaks in a DNA segment are classified both by complexity (left) and source (right). The model entails two parameters, d_{DSB} is the maximum separation between two damage sites on alternate sides of a DNA strand for us to consider that a DSB has occurred (typically $d_{\mathrm{DSB}} = 10$ bp). d_s is the distance between two damage sites for us to consider that the damage events should be considered as two separate breakages (yielding two separate segments that need classification). Whilst many of the classifications are clear, we note that a DSB+ requires a DSB and at least one additional break within a ten base pair separation, while a DSB++ requires at least two DSBs along the segment, regardless of whether they are within d_{DSB} of each other or not. For break complexity, the most complex break type is always chosen. When classifying breaks by source, we pay attention not to all damage along the strand, but to the damage which causes DSBs only. DSBs from only indirect sources are classified as $\mathrm{DSB_i}$, and those only from direct sources are classified as $\mathrm{DSB_d}$. $\mathrm{DSB_{hyb}}$ is distinguished from $\mathrm{DSB_m}$, as $\mathrm{DSB_{hyb}}$ requires that the DSB not occur in the absence of indirect damage. Otherwise, a break caused by indirect and direct sources is classified as $\mathrm{DSB_m}$. Where a segment contains both indirect and direct DSBs, it is classified as $\mathrm{DSB_m}$. Similarly, when a segment contains a DSB classified as $\mathrm{DSB_{hyb}}$ in conjunction with a direct DSB or mixed DSB, it takes the $\mathrm{DSB_m}$ classification, otherwise it keeps the classification $\mathrm{DSB_{hyb}}$.

Identifying SSBs and DSBs

The complexity and source of strand breaks can be classified in a variety of ways. We follow the classification scheme of Nikjoo et al. (1997), which classifies strand breaks by both complexity and source as shown in Figure 3.17. Typically, a DSB is considered to have occurred when the distance between SSBs on opposite sides of the DNA strand is within $d_{\mathrm{DSB}} \leq 10$ bp. We introduce a new parameter, d_s, which allows the distance between two damaged sites to be increased or decreases before the damaged sites are counted as two separate damage events.

Classification by complexity is important for determining the severity of strand breaks and their implications for biological damage, particularly as DSBs are a major driver of cell death. By classifying breaks by source, it is possible to see the relative impacts of direct and indirect damage induction on the DNA. In particular, the classification $\mathrm{DSB}_{\mathrm{hyb}}$ is noteworthy, as it indicates DSB's that would not occur unless indirect damage is induced.

3.4 Model Parameter Studies

A key challenge of this kind of simulation is gathering predictive data based on a true physical understanding of DNA damage mechanisms, whilst avoiding tuning a model to match an experiment, and accepting those results as truth. In saying this, one is reminded of an expression attributed to John von Neumann, *with four parameters I can fit an elephant, with five I can make him wiggle his trunk* (Dyson, 2004). Many of the parameters chosen in these simulations to date are chosen based on fits of a model to an experiment. This is unavoidable to some extent, and remains regarded as state-of-the-art[6]

In this section, we try and understand the impacts of the different parameters in our model, in order to better understand their impact on our work. We draw inspiration from Nikjoo et al. (1997), who present a past parameter study on a simple geometry. We explore in particular how the parameters that define physical damage are related, with certain parameters showing a clear inverse proportionality. The impact of different physics models is also explored with clear differences in biological damage existing between models. We then explore the impact of chemical damage, and how different assumptions about strand break induction can change the results of the simulation. Here we also address the relative importance of the radical species $^{\bullet}\mathrm{OH}$, $\mathrm{e}^{-}_{\mathrm{aq}}$ and $\mathrm{H}^{\bullet}$ for strand and base damage.

[6] A recent review (Nikjoo et al., 2016) states that *the quantity 17.5 eV energy deposition in the S-P volume was first suggested by Charlton & Humm (1988) in a simulation of an experiment by Martin & Haseltine (1981) remains the only value to date for the energetics of DNA damage based on biological experiment and evidences*, however we note that in the original work by Charlton & Humm, this value is the best fit of a simulation model to experimental data.

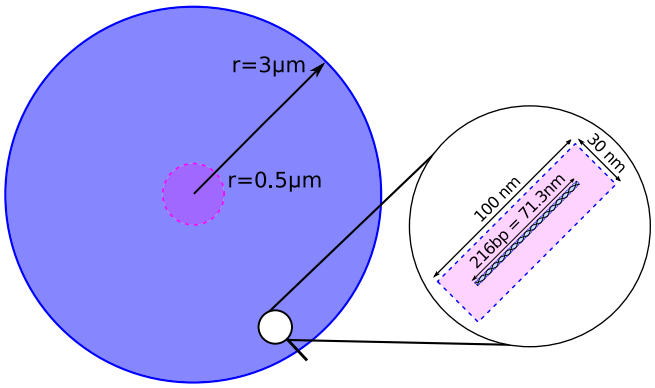


Fig. 3.18 Our geometry for parameter sweeps consists of a 3 μm sphere filled with 200,000 individual 216 bp long straight DNA segments in a $100 \times 30 \times 30$ nm placement volume. Primary electrons are generated randomly, with a random direction in a smaller 500 nm sphere in the centre of the test region. As we use primaries with energies no greater than 4.5 keV, no primaries can escape the larger spherical region, and all primaries see an equivalently random region.

3.4.1 Geometry for Parameter Studies

We have defined a simulation geometry based upon a previous study of direct and indirect DNA damage yields in straight DNA fibres (Nikjoo et al., 1997). In order to see how different model parameters affected DNA damage yields, we replicated a geometry similar to this, studying DNA damage in randomly placed 216 bp long straight DNA fibres, in a $100 \times 30 \times 30$ nm placement volume. These volumes were placed randomly in a 3 μm sphere, with random orientations. The placement algorithm, written in Python, made sure that the spheres didn't overlap (based on the method of separating axes). We placed 200,000 such volumes in the sphere, filling 20% of the sphere with DNA regions, approximately the maximum density possible without resorting to a packing algorithm (Figure 3.18).

In the majority of tests, we explored how the strand break yield varied for primary electrons with energies of 300, 500, 1,000, 3,000 and 4,500 eV. We chose these energies to follow the work of Nikjoo et al., and also because they represent a range of energies significant in the radiolysis of water, with 300 eV representing a few 'spurs', or a 'blob', and 4.5 keV being equivalent to the amount of energy typically found in a short track (c.f. Mozumder & Magee, 1966).

3.4.2 Results

We first tested the physical damage parameters, not considering chemical damage, to understand how this model behaved when the parameters that define physical damage were varied.

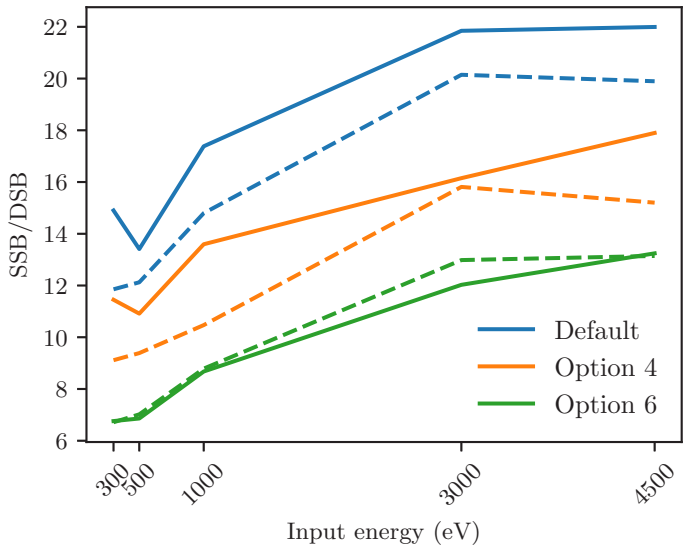


Fig. 3.19 For a physical damage radius of 6 Å, the ratio of SSBs to DSBs from physical processes alone is shown for a fixed break energy of 17.5 eV (solid lines) and a break energy that varies from 5 37.5 eV following the model used by PARTRAC (dotted lines). These are in turn shown for three physics lists.

Next, we explored how the model responded when radicals more than a set difference from the DNA strand were killed, rather than simulating them until the simulation ended - this is of interest as the majority of indirect damage comes from radicals created close to DNA. Finally, we explored how varying the likelihood that chemical reactions between radicals and DNA strands impacts the amount of DSBs.

Direct Damage We explored how the direct damage of DNA changed in response to changes in the radius for scoring physical damage, the threshold energy for physical damage, and the physics list used. Tables of break damage for selected sweeps are presented in Appendix C for reference. For each input electron energy, we simulated 1 GeV worth of events (ie. 10^6 events for 1 keV) so that the total yield of breaks for different events is roughly comparable.

We consider first the relationship between the input energy and the ratio of SSBs to DSBs. In general, higher input electron energies increase the ratio of SSBs to DSBs (Figure 3.19), regardless of physics list or the energy required to induce a strand break. This is consistent with the LET of low energy electrons increasing as their energy decreases, thus inducing breakages more easily. An exception to this is noted at 300 eV, where high threshold energies for induced breaks can reduce the number of DSBs relative to SSBs due to the electrons having less energy with which to cause breaks initially. The PARTRAC damage model, which

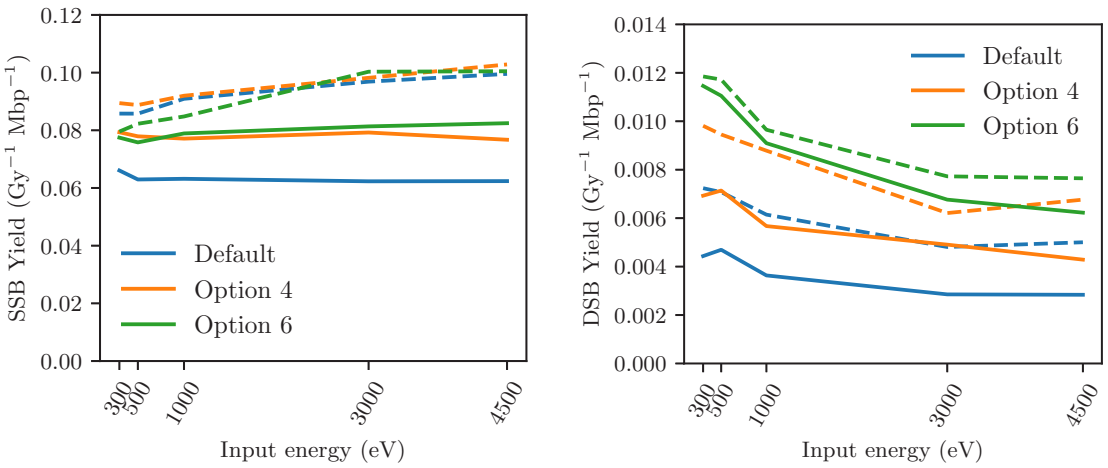


Fig. 3.20 The absolute yields of SSBs and DSBs for the Nikjoo (solid lines) and PARTRAC (dotted line) damage conditions outlined in Figure 3.19 can also be quite different, depending on the physics model.

allows DNA strand breaks to occur with increasing probability from energy deposits as small as 5 eV predicts more damage to be located in DSBs than SSBs in general than a flat 17.5 eV cut off for strand breaks. The Geant4-DNA Option 4 physics model (Kyriakou et al., 2015), which provides a more realistic implementation of the dielectric response of liquid water, shows a notable (20%) deviation from the predictions of the default Geant4-DNA model for liquid water, indicating that the results obtained for DNA damage are sensitive also to the physics model used, to at least a similar extent as they are to the choice of damage model. The Geant4-DNA Option 6 models, which is a recent implementation of the CPA-100 model set within Geant4 (Bordage et al., 2016), shows an even stronger deviation than the default model from the default strand break yields, though for this model the 17.5 eV break threshold and PARTRAC models give very similar values for SSB/DSB.

The absolute yields of physical damage in this test geometry are also quite sensitive to the physics model and the physical damage model. For example, while the Geant4-DNA option 6 model shows the SSB/DSB ratio to be near invariant when passing from a 17.5 eV SSB induction threshold and a variable threshold, this is not due invariance in the underlying absolute numbers of SSBs and DSBs (Figure 3.20). Here, it can be seen that for a constant damage induction threshold, the yield of SSBs remains constant for electron energies above 1 keV, whilst some variability is seen when lower electron energies are considered in the (variable) damage model. DSBs are signifcantly more sensitive to spatial clustering in models, and show a stronger sensitivity to both the physics models used, and the criteria for break induction. At worst, changing the break induction and physics models

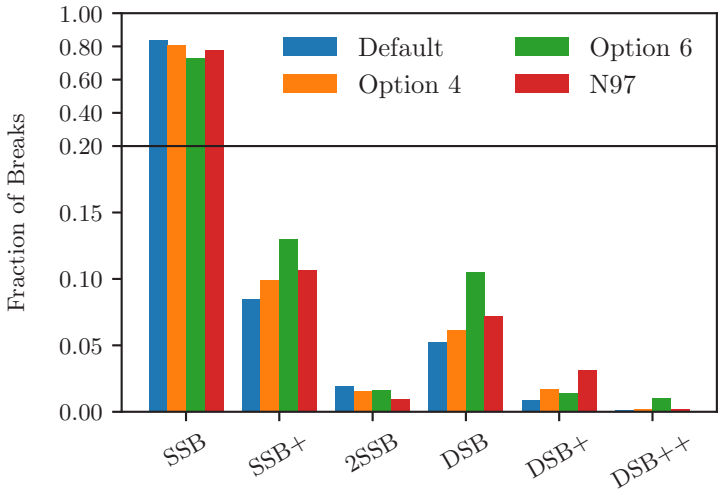


Fig. 3.21 The fractions of breakages caused by 300 eV electrons with a 17.5 eV simulated in this work using the Geant4-DNA physics lists and Nikjoo et al. (1997) compare reasonably well. Geant4-DNA produces a larger fraction of SSBs than Nikjoo et al., but fewer DSBs, while the CPA-100 models in Geant4 produce relatively more DSBs than SSBs.

can cause a two-fold difference in DSB yield.

For 300 eV electrons, the classifications of strand breaks measured in Geant4-DNA option 4 compare favourably to those simulated by (Nikjoo et al., 1997). Using a 17.5 eV break threshold, broadly the fractions of SSBs and DSBs are the same (Figure 3.21). A slightly higher fraction of direct damage in Geant4-DNA is located in SSBs however, and a correspondingly smaller fraction of damage is located in DSBs. This is due to the physics models used by Nikjoo et al. simulating a higher LET for very low energy electrons than Geant4-DNA option 4. In particular, Nikjoo et al. consider an older version of the CPA-100 models (i.e. option 6), though recent improvements to the model appear to have increased the ratio of DSBs to SSBs.

The impact of the energy threshold for breaks is shown in detail in Figure 3.22, where the fraction of breaks for the different classifications is given. As the energy required for an SSB decreases, the fraction of complex breaks increases, across all input electron energies. It's also clear that low energy electrons are more likely to cause complex breaks than high energy electrons. This behaviour changes for the highly complex DSB++ breaks when the energy threshold for breaks increases. Partially, this could be due to the fact that for 300 eV electrons, at least 70 eV is required to cause a DSB++ classified fracture, which is a significant fraction of the input electron energy. It could also be linked to the LET modelled in Geant4-DNA for low energy electrons being such that the 17.5 eV energy threshold has a significant impact on the total yield of breaks compared to other energy differences. This is

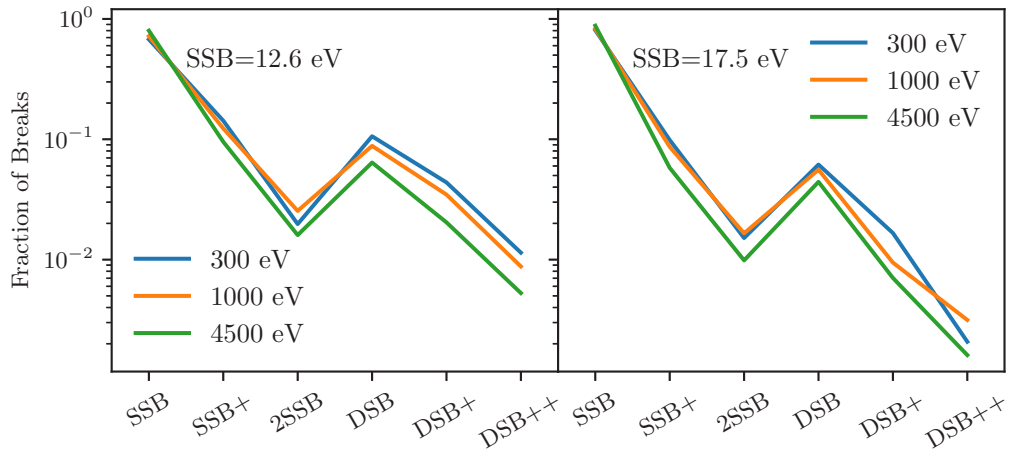


Fig. 3.22 Distribution of strand breaks by classification for a 12.6 eV (left) and 17.5 eV damage threshold, considering a damage threshold radius of 6 Å and using the Option 4 Geant4-DNA physics constructor.

supported by Figure 3.23, where we consider the impact of both radius and break energy on the ratio of SSBs to DSBs, noting that the small change in the break yield energy from 15 eV to 17.5 eV causes a large change in the SSB:DSB ratio.

The impact of the damage radius chosen is quite important in determining the ratio of single to double strand breaks as it is correlated with the break energy. The PARTRAC damage breakage model however appears to be consistently in between the 15 eV and 17.5 eV constant damage threshold models across most radii considered, which is somewhat unexpected as the model predicts a $p = 0.5$ chance of a strand break only at 21.25 eV. This likely reflects the significant preference in energy deposits in DNA towards lower energies, particularly below 17.5 eV, as the PARTRAC model allows these events to be sometimes counted as damage. Notable also in Figure 3.23 is the consistent prediction of more DSBs per SSB from the option 4 model set than the default model set. In particular, option 4 shows significantly better agreement with the strand break ratio found by Nikjoo et al. (1997) in a similar geometry (where we calculate the radius for the Nikjoo et al. study as the radius that gives the same volume as their geometry, c.f. Figure 3.15), especially at 15 eV. A large difference exists however between what we simulate at 17.5 eV and what Nikjoo et al. simulate. We are inclined to have more confidence in the Geant4-DNA option 4 model here than the overplotted data points, as we see in option 4, for each increase in the energy required for a strand break, a roughly equivalent increase in the ratio of DSBs to SSBs. This is not the case in the verification data points, where the total SSBs to DSBs changes by varied intervals, up to values as high as 90 and 54 (not plotted) for break energies of 21.1 eV and

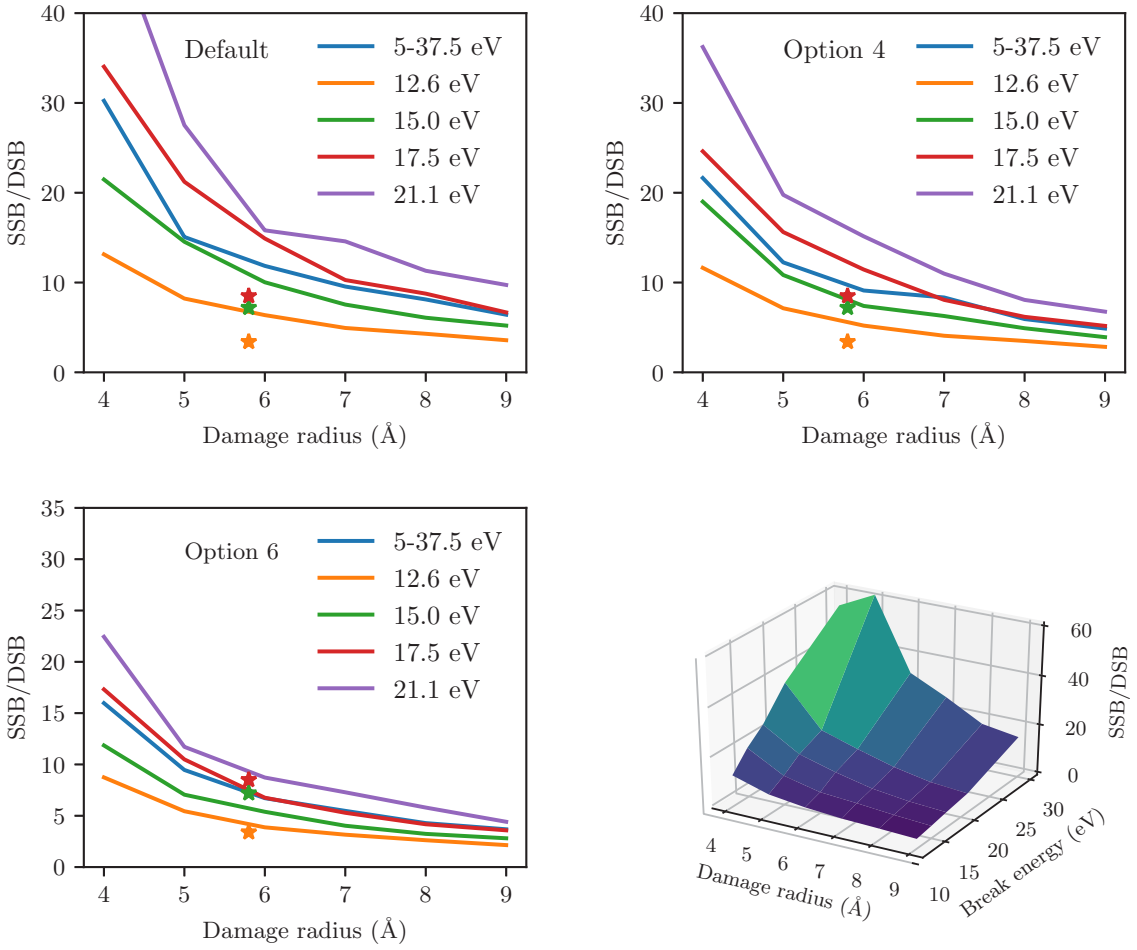


Fig. 3.23 Variation of the SSB to DSB ratio with the damage radius for a variety of break energies for the default (top left), option 4 (top right) and option 6 (bottom left) physics models. The surface plot in the bottom right hand panel highlights that the SSB to DSB ratio is degenerate for a range of break energy and damage radius combinations. Stars in each of the first three panels indicate the results found by Nikjoo et al. (1997) in a similar configuration.

30 eV respectively.

Looking at and above $r = 6$ Å in particular, it is interesting that all the Geant4-DNA models are quite linear for most energy deposit thresholds when $r > 5$ Å. This occurs because for large radii, changes in the radius cause a fractionally smaller change in the volume of the sensitive area, flattening the curve.

Impact of Radical Diffusion Chemical damage is of particular interest in DNA damage because radical species can diffuse, and thus it is significantly less localised than physical damage. Here we try and ask at what extent does this de-localisation of damage affect the amount of SSBs and DSBs. It is conceivable that radicals that diffuse towards DNA from afar and then react with it are significantly less likely to cause complex DNA damage, rather they are likely to have isolated chemical reactions with DNA which would be easily repaired by cellular systems. This question is important for code optimisation also, as if far away radicals do not cause significant DSB formation, their tracks can be killed without greatly altering the number of DSBs simulated.

Radical tracks are able to be killed when the radical is located more than a certain distance from the DNA strand. We varied this distance in our cylindrical test geometry, from 0 nm (chemical reactions can only occur if particles are created co-incident with DNA molecules) to 9 nm. 10^5 events were simulated, considering a single 4.5 keV primary particle in each event (the Geant4-DNA physics option 4 constructor was used in the physical stage). For radical damage, the likelihood of a chemical reaction with a base proceeding to base damage was set to one, the probability that base damage could induce a strand break damage was set to zero, and 0.65 was chosen as the likelihood that a chemical reaction with a strand leads to strand damage (and a strand break), following Nikjoo et al. (1997). Physical damage required an energy deposit within 7 Å of the DNA molecule of at least 17.5 eV.

In this section, we allow the simulation of chemistry to proceed up to 1 μs. The aim is to investigate the impact of diffusion on breakages. Due to this however we see that, across all radii at which to kill radicals considered, that the number of strand breaks seemingly grows without bounds. This is in contrast to what Nikjoo et al. (1997) indicate, that $^{\bullet}$OH attacks plateau for radii above 4-6 Å(this corresponds to 3-5 Å in our schema), where the chemical simulation is stopped at 1 ns. The increase is seen across all breakage complexities in Figure 3.24, but is particular pronounced in the most simple breaks (SSB and DSB). As expected, as more diffusion is permitted, SSB+ and 2SSB increase at a similar rate. There are always however less 2SSB than SSB+, as when spatial correlations are high, damage events resembling 2SSB will be classified as DSBs, whilst when two radical attacks events occurs randomly across a segment (i.e. far apart), SSB+ and 2SSB occur with almost equal

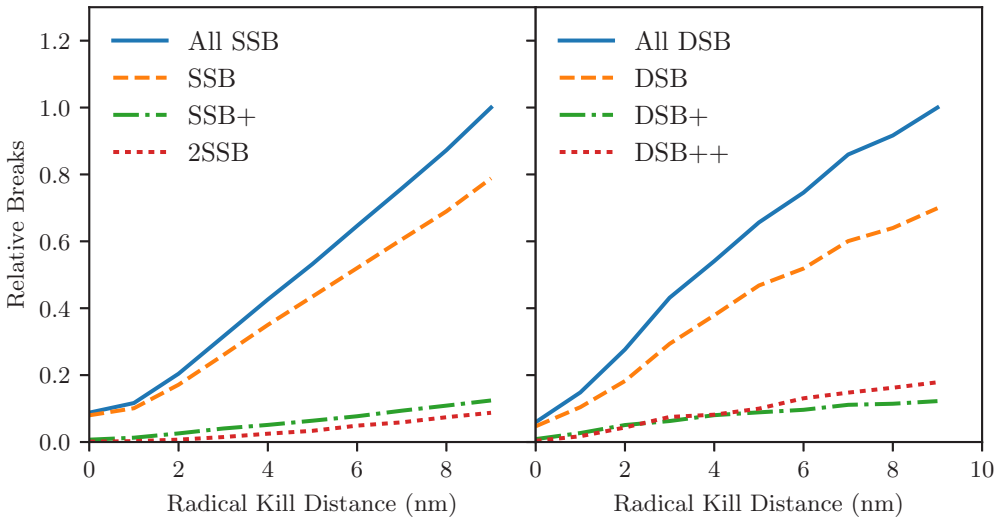


Fig. 3.24 The relative number of strand breaks caused by direct and indirect effects as the distance from the DNA strand at which radicals were killed is varied, broken down by complexity. SSBs are shown on the left, normalised to 1 (from 1.17 SSB Gy ^{1} Mbp ^{1}), and total DSBs are shown on the right, normalised to 1 (from 0.12 DSB Gy ^{1} Mbp ^{1}). Due to diffusion, the number of strand breaks never ceases to grow across the range considered.

probability.

Figure 3.25 shows the sources of the breaks recorded in these simulations. As the radius from molecules at which chemical species are killed increases, a growth in indirect SSBs and DSBs is seen. As the amount of physical damage in the simulation is unchanged with increases in the distance from DNA at which radical tracks are killed, the amount of DSB_d damage doesn't increase, but rather, direct SSBs and DSBs are converted into DSB_{hyb} and DSB_m, which eventually plateau.

A similar behaviour is seen when one considers the energy depositions in DNA which cause strand breaks (Table 3.2). Regardless of the radical kill distance, the same energies are always deposited in the DNA strand (as expected), and are in close agreement with other work. We find however that compared to other works we frequently see events where radicals react with DNA strands without the primary track depositing any energy in the DNA molecule, and as a corollary of this, significantly more DSBs are caused by when energy deposits in the DNA molecule are on the scale of 0 60 eV rather than 60 150 eV as has been previously suggested.

[7]The frequency is given relative to segments that have $E_{dep} > 0$ eV. This ignores breaks that are solely caused by indirect effects. The frequency of such breaks relative to those we consider is shown in the 0 eV column.

[8]Nikjoo et al. (1997), results for a 4.5 keV electron interacting with 216 bp long DNA segments.

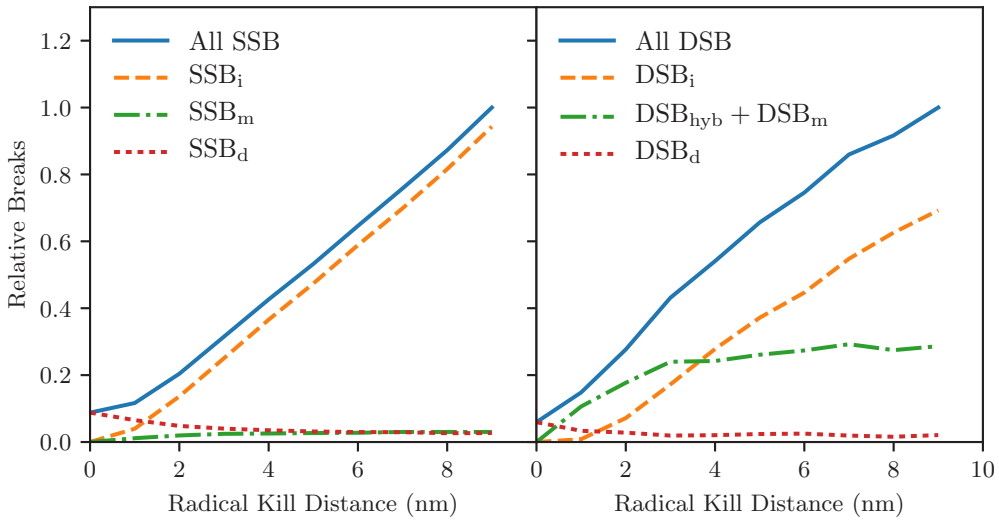


Fig. 3.25 The change in SSB (left) and DSB (right) abundance, broken down by source, as the distance at which radicals are killed changes. Normalisations for each curve are the same as in Figure 3.24. As more radicals are included in the simulation, all the growth in damage is attributable to growing indirect damage. Typically, DSB_{hyb} dominates DSB_m. The percentage of tracks in each category is given in Table C.11.

One reason for this discrepancy is that other works end their chemistry simulation after a few nanoseconds, or provide some measure of radical scavenging. This decreases the amount of reactions with radicals that can occur and is investigated in the following section. On the same note, while our selection of $p_{SSB} = 0.65$ for the chance a chemical reaction between a strand and a base yields an SSB follows that used by Nikjoo et al. (1997), our reaction rates lead to $^{\bullet}$OH radicals interacting with strands and bases in a 32:68 ratio, rather than the 20:80 ratio counselled. As a result, when $^{\bullet}$OH interacts with DNA, the chance a strand break is induced is 24%, rather than 13% as used by Nikjoo et al..

Radical Diffusion and Chemical Reactions The number of reactions that occur as a function of the radius at which radical tracks are killed is shown in Figure 3.26. Chemical reactions are dominated by the base attack reactions from e_{aq} and $^{\bullet}$OH, while attacks from H$^{\bullet}$ are negligible by comparison. As expected, reactions between $^{\bullet}$OH and the sugar-phosphate moiety occur less frequent than base damage, and in the model used for this section 65% of these reactions go on to cause strand breaks. We note that across the range of data points studied, $^{\bullet}$OH base attacks represent $67.6 \pm 0.3\%$ of all $^{\bullet}$OH damage, disagreeing with the 80% fraction of $^{\bullet}$OH attacks measured to occur in bases (Scholes et al., 1969). This hints that the reaction rate chosen between $^{\bullet}$OH and deoxyrobose-5-phosphate is too high, or doesn't

Table 3.2 Fraction of energy deposits and DSB yields for different energy deposits in a 216 bp straight DNA segment by 4.5 keV electrons for different radical simulation radii.

Radius	E_{dep} frequency[7] (%)				DSB frequency (%)			
(nm)	0 eV	$0^+ - 60$ eV	$60 - 150$ eV	> 150 eV	0 eV	$0^+ - 60$ eV	$60 - 150$ eV	> 150 eV
0	0.0	91.6	7.9	0.5	0.0	19.7	67.5	12.8
1	56.6	91.8	7.9	0.3	0.4	44.5	51.0	4.1
2	199.9	91.7	8.0	0.2	9.9	52.6	34.9	2.5
3	348.8	91.7	8.0	0.3	17.3	55.1	26.4	1.3
4	473.7	92.1	7.6	0.3	27.2	49.8	21.7	1.3
N97[8]	-	92	7	1	-	30	64	6

take into account the interactions that may arise to block or preferentially favour reaction sites when this molecule is part of the DNA backbone (Balasubramanian et al., 1998). Future simulations may try and consider how the DNA molecule can geometrically block reaction sites, though this is computationally difficult.

Near 7 nm, we remark that $^{\bullet}$OH reacts more frequently with bases than e^-_{aq}. This is possibly tied to the slightly higher yields of $^{\bullet}$OH than e^-_{aq} in water. The impact of diffusion is shown in the right hand panel of Figure 3.26, where the simulation was stopped after 1 ns. In this scenario, the number of $^{\bullet}$OH reactions reaches a plateau around 4-5 Å, while the number of e^-_{aq} reactions continues to grow due to the electrons higher diffusivity. The consequences of this can be seen in the numbers of SSBs and DSBs recorded when diffusion is limited to only 1 ns (Figure 3.27). The number of indirect damage events plateaus for DSBs around 4 nm, and for SSBs around 6 nm.

Figure 3.28 shows the impact of cutting the simulation at 1 ns for a range of chemical distances considered. Both as the time permitted for diffusion, and the distance radicals can diffuse are increased, the amount of strand breaks recorded increases. Cutting the simulation at 1 ns, we place ourselves closer to the work of Nikjoo et al. (1997), though we find that in these conditions, when simulating radicals within 4 nm of DNA, the yield of SSBs (DSBs) is 0.41 SSB (0.058 DSB) Gy^{-1} Mbp^{-1}, whilst Nikjoo et al. predicts the yield of SSBs (DSBs) to be 0.29 SSB (0.018 DSB) Gy^{-1} Mbp^{-1}.

Efficiency of SSB formation in strands The likelihood that a chemical reaction between a base and a sugar-phosphate moiety induces a single strand break is one of the most poorly known parameters in our simulation. Up to now, we have considered that the efficiency with which such events induce breaks (here referred to as p_{SSB}) is 65%, following other authors who combine an estimate of the break efficiency of reactions between $^{\bullet}$OH and DNA of 12% (Milligan et al., 1993), and the observation that only about 20% of reactions

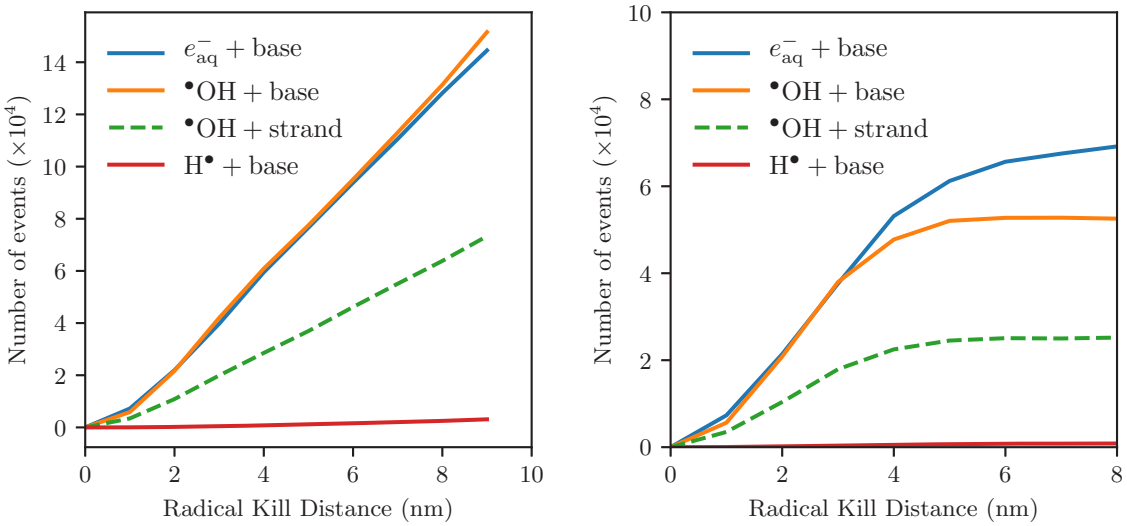


Fig. 3.26 The amount of chemical reactions that occur as the distance from the DNA at which radicals are killed grows up to at least distances of 10 nm (left), though when the simulation is stopped after 1 ns, diffusion causes the number of reactions to plateau (right). The reaction between the $H^\bullet$ radical and a base is the least common of the reactions plotted (at 9 nm, $\approx 3 \times 10^3$ reactions). The reactions between e_{aq} and $H^\bullet$ with strands are at least a factor of five less likely again.

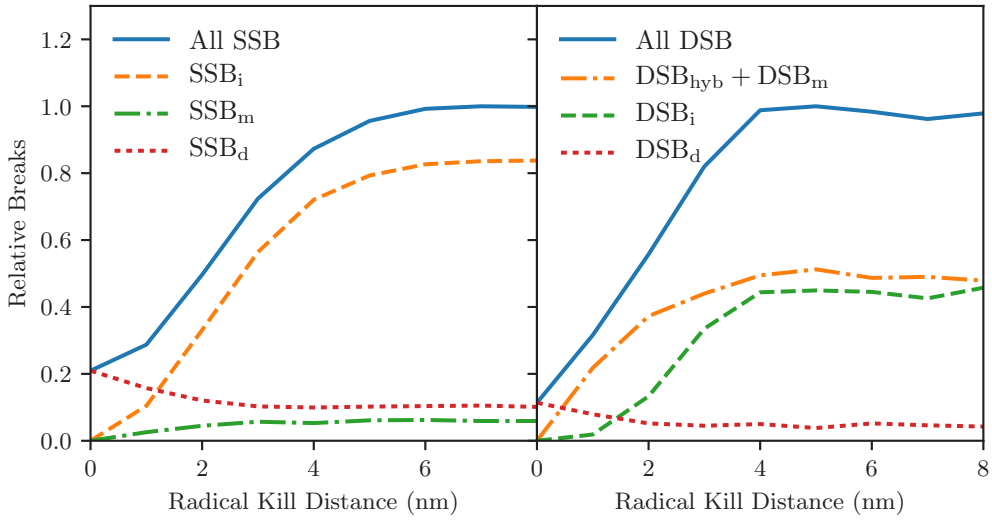


Fig. 3.27 The change in SSB (left) and DSB (right) abundance, broken down by source, as the distance at which radicals are killed changes. The percentage of tracks in each category is given in Table C.11, and SSBs (DSBs) are normalised by 0.48 SSB (0.058 DSB) Gy^{-1} Mbp^{-1}.

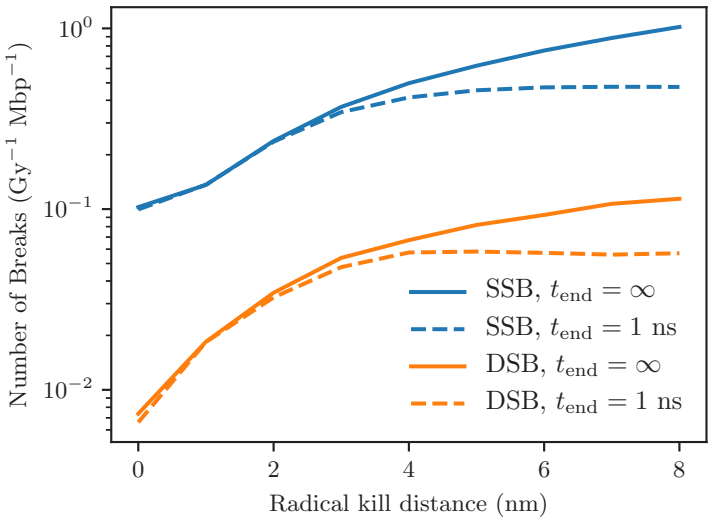


Fig. 3.28 When considering only radicals within less than 4 nm of DNA for the chemical simulation, restricting diffusion to 1 ns has little effect on damage yields. Above this, however further DNA damage is stopped.

between •OH and DNA occur in the sugar phosphate moiety (Scholes et al., 1969)[9]. Other estimates of the efficiency of strand break induction range up to 29% for when •OH reacts with DNA (Udovicić et al., 1994), it is suggested that these variations are related to scavenger concentration.

We investigated the importance of the strand break induction parameter, simulating 3.33×10^5 primary electrons with input energies of 4.5 keV in our test geometry, for values of $p_{\text{SSB}} = 0,\ 0.2,\ 0.3,\ ...,\ 0.9,\ 1.0$. A radical kill distance of 4 nm was used, and simulations proceeded until no more radicals were left. From Figure 3.29, it can be seen that choosing $p_{\text{SSB}} = 0.65$, corresponds rather in our simulations to near the middle of these two averages. This is because our simulated reaction rates overestimate the number of chemical reactions between radicals and strands compared to bases, relative to Scholes et al. (32% rather than 20%). In order to conduct simulations in line with Nikjoo et al. (1997), a value of $p_{\text{SSB}} = 0.4$ should be favoured.

As two SSBs are required to form a DSB, decreasing p_{SSB} impacts the ratio of single to double strand breaks (Figure 3.30, left panel). From the rate corresponding to physical damage alone ($p_{\text{SSB}} = 0$), the ratio of SSBs to DSBs drops near linearly as the efficiency of converting chemical reactions to strand breaks increases. The different individual responses of SSBs and DSBs to changes in p_{SSB} can be seen in how the fraction of indirect breaks

[9]This would tend to suggest $p_{\text{SSB}} = 0.65$ is a slight overestimate of what should be $12\%/20\% = 0.6$. Nevertheless we use the 65% adopted by Nikjoo et al. (1997).

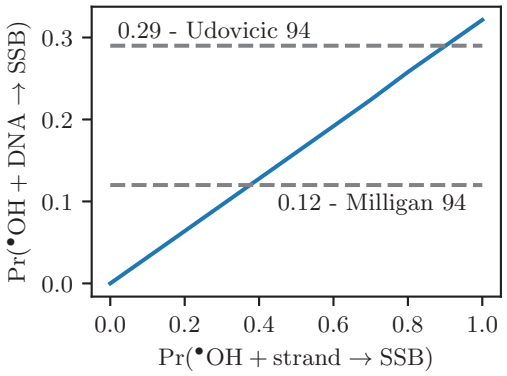


Fig. 3.29 The likelihood a chemical reaction with a base increases, impacts the likelihood that a reaction between •OH and DNA induces an SSB. This can be related to measured efficiencies of SSB formation for •OH reacting with DNA.

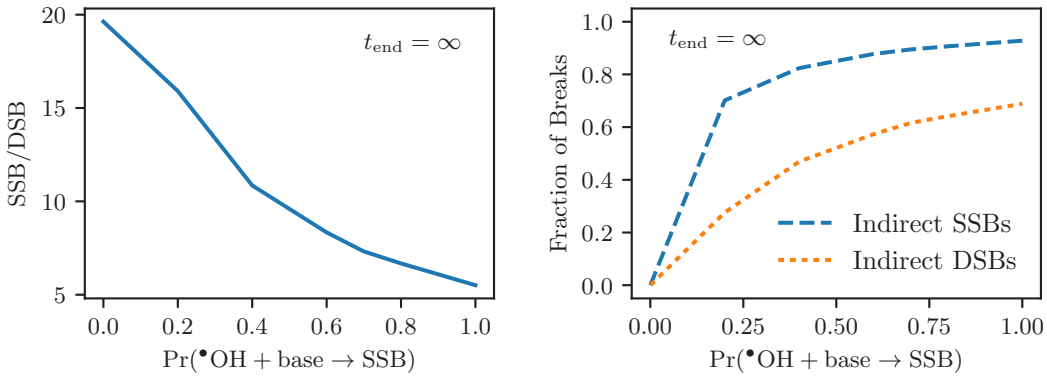


Fig. 3.30 As the likelihood of a strand break from a radical reaction increases, the relative number of SSBs per DSBs decreases (left). Equally, the damage becomes entirely dominated by indirect strand breaks (right). Both these figures consider a simulation that ends when all radicals are consumed.

as a fraction of total breaks changes as the efficiency of strand breaks increases (Figure 3.30, right panel, here we consider only SSB_i and DSB_i, as a fraction of total breaks). The sensitivity of the results to these parameters is striking, especially given that SSB/DSB is a key measurable in these simulations, and can vary by almost a factor of 3 across a range of reasonable break induction probabilities. The fraction of purely indirect and direct SSBs plateaus as the probability of inducing a break rises. Whilst it approaches 1 for SSBs, this is not the case for DSBs as a non-negligible proportion of DSBs, at least for the simulation radius considered, belong to mixed damage classifications (DSB_m and DSB_{hyb}).

3.4.3 Discussion of Parameter Sweep

The differences in physical damage recorded between the Geant4 models are a consequence of the underlying physics being simulated. The default, option 4, and CPA-100 physics models all model electron scattering cross sections differently below 10 keV. The default option is based on modelling the dielectric response function of liquid water, and the option 4 model is based on a refinement of this approach, redistributing the imaginary part of the dielectric function to better model electrons close to their binding energies. Option 6 is based on the CPA-100 model set, which models ionisation cross sections via the binary encounter Bethe model. The result of this is that option 6 has in general higher interaction cross sections below 10 keV, while the option 4 models have higher total cross section below 1 keV than the default models. Thus options 4 and 6 model energy depositions more densely than the default option, contributing to a higher number of both SSBs and DSBs, driving down the ratio of the two quantities. Nikjoo et al. used an older version of the CPA-100 code than we model here, however this explains the good agreement between the physical damage modelled here in option 6 and their work.

Our simulations of chemistry in the test geometry considered three parameters, the distance from DNA at which radical tracks are killed, the end time of the simulation (which impacts diffusion), and the likelihood of inducing a DNA break from a recorded chemical reaction between a sugar-phosphate moiety and a radical. Even though we applied the same treatment to $H^\bullet$ and e^-_{aq} that we applied to $^\bullet OH$, we discuss these changes only in the context of $^\bullet OH$ as e^-_{aq} and $H^\bullet$ radicals rarely react with deoxyribose-5-phosphate.

Both the end time of the simulation and the radius of radicals simulated are highly linked to the concepts of diffusion and radical scavenging. These two ideas are linked by considering scavenger abundancies and diffusion distances in cells. From the diffusion constant, D_c, of the $^\bullet OH$ radical ($D_c = 2.8 \times 10^{-9}$ m^2 s^{-1}, from Schwarz, 1969), the radical's mean

displacement $\bar{x}$ in terms of time can be found in a three dimensional system as

$$\begin{aligned} \bar{x} &= \sqrt{6D_c t}, \\ &= 4.09\sqrt{t} \text{ nm}. \end{aligned} \tag{3.8}$$

Numerous authors conclude their chemical simulations of $^{\bullet}$OH after only a few nanoseconds (e.g. Meylan, 2016; Nikjoo et al., 1997), implying that this provides a simulation of radical scavenging effects. Limiting the time of the simulation makes sense in many ways, as the majority of $^{\bullet}$OH-induced damage comes from radicals that are non-scavengeable, and are created close to, or in the hydration shell of DNA (Daly, 2012; Ward, 1988). In cellular media, the average diffusion distance of the $^{\bullet}$OH radical is 6 nm (Roots & Okada, 1975), corresponding to an average lifetime of 3.7 ns, and a scavenging efficiency of around 3×10^8 s^{-1}.

The need for this to be modelled in some way can be seen by the limitless growth of indirect SSBs and DSBs as the distance at which we model radicals increases, far beyond what is seen in living cells (≈ 0.01 DSB Gy^{-1} Mbp^{-1}). That said, yields of DSBs as high as 0.23 DSB Gy^{-1} Mbp^{-1} are observable in plasmids when scavenger abundances are low (Souici et al., 2016), and $^{\bullet}$OH molecules are likely to diffuse tens of nanometres. While simulating this is outside of the domain of this study, it would be feasible to expand the radius at which radicals are simulated and conduct a study of DNA damage in the complete absence of scavenging. The growth in indirect strand breaks as the chemistry simulation radius increases is also an encouraging 'sanity check' for our classification routines. The classifications of complexity all behave as expected (Figures 3.24 and 3.25), with the SSB+ and 2SSB classifications showing similar though offset behaviour (as at small radii, the two radicals that react on opposite strands to cause a 2SSB classification are likely to be so close they cause a DSB). Similarly, the DSB++ classification grows faster than the DSB+ classification, as increasing the amount of radicals simulated pushes DSBs to become DSB+, and breaks classified as DSB+ to a final state of DSB++. When the source classifications are studied, hybrid and mixed double strand breaks plateau, as the limited amount of direct breaks is consumed by a growing amount of indirect damage.

In real systems, some level of scavenging should be considered. Following other authors, we investigated ending the simulation at 1 ns, in order to see the impact of limiting radical diffusion on our results. It's worth bearing in mind though that, from the mean lifetime τ of

a radical, the probability p_{scav} that it will have been scavenged within a time period t is

$$\begin{aligned} p_{\text{scav}}(t) &= \frac{\int_0^t e^{-t/\tau}\,dt}{\int_0^\infty e^{-t/\tau}\,dt}, \\ &= 1 - e^{-t/\tau}. \end{aligned}$$

Substituting the values for the $^\bullet$OH radical, we find that after 1 ns, the probability that a radical has been scavenged is

$$\begin{aligned} p_{\text{scav}}(t) &= 1 - e^{-t/3.7\text{ ns}}, \\ p_{\text{scav}}(1\text{ ns}) &= 0.24. \end{aligned}$$

This does question the results of simulations that cut the simulation time as a means of considering scavenging, as a significant fraction of radicals may still be present in the environment. How many of these will react with DNA however within the next few nanoseconds has not been investigated. It is arguably better practice instead to limit the radius of the chemical simulation to a region where scavenging is not relevant, and better yet is to simulate scavenging as is done by Friedland et al. (2011).

When simulations are limited to 1 ns, we reproduce the chemical reaction behaviour of Nikjoo et al. (1997), notably seeing that $^\bullet$OH radicals further than 4 nm from the DNA molecule do not greatly contribute to damage (Figure 3.26). A similar threshold applies to the production of indirect and mixed SSBs and DSBs, a consequence of a reduction in base damage (Figure 3.27). Based on the measurements here, we find that SSB (DSB) damage plateaus at 0.48 SSB Gy^{-1} Mbp^{-1} (0.058 DSB Gy^{-1} Mbp^{-1}), higher in both cases than the value measured by Nikjoo et al. (0.29 SSB Gy^{-1} Mbp^{-1} and 0.018 DSB Gy^{-1} Mbp^{-1}). In part, this is caused by an elevated value for p_{SSB}, the chance that a reaction between $^\bullet$OH and a strand leads to an SSB (Figures 3.29 and 3.30). We compare our work to a simulation that bases their DNA breakage model on the work of (Milligan et al., 1993), where the efficiency of a reaction between DNA and $^\bullet$OH leading to a strand break is 12%. In our work, we cannot choose this parameter, it is rather a consequence of the reaction rates for reactions between strands and bases, as well as p_{SSB}. By setting $p_{\text{SSB}} = 0.4$, we can contrive a simulation that matches that of Nikjoo et al., however this may be a way of avoiding a better calibration of the reaction rate between $^\bullet$OH and deoxyribose-5-phosphate. In the DNA chain, each sugar-phosphate moiety is bonded to two other moieties and a base pair, which may in turn decrease its reactivity. In their measurement of the efficiency of SSB formation following $^\bullet$OH-DNA reactions, Udovicić et al. (1994) find a reaction efficiency of 29%, significantly higher than the values we considered here, though their reaction rate for

the reaction between $^{\bullet}$OH and DNA was $2.5 \pm 0.5 \times 10^8$ L mol^{-1} s^{-1}, an order of magnitude lower than the reaction rates we consider here for $^{\bullet}$OH.

Much as the radius for physical energy deposition and the energy required for a break are strongly correlated, the reaction rates for $^{\bullet}$OH and the efficiency of strand breaks are correlated, and are thus difficult to precisely define. In this section, we have attempted to illustrate the general tends that occur when parameters change. Based on these, we have been able to identify some parameters that resemble the work of Nikjoo et al. that this parameter study has tried to replicate. Based upon these, we can conduct a set of simulations that illustrates the similarities and differences between the Geant4-DNA platform, and that used by Nikjoo et al..

3.4.4 Reproducing Nikjoo et al., 1997

Based upon the physical damage parameters investigated above, and the value of p_{SSB} determined to have the same break efficiency as that used by Nikjoo et al. (1997), we are able to reproduce this simulation in order to illustrate the differences between our Geant4-DNA and other mechanistic DNA damage simulation toolkits. The parameters we used for this simulation are given in Table 3.3, while the geometry remains the same as in the previous section (2×10^5 strands of DNA, each 216 bp long). We expect already that we should record less physical damage than Nikjoo et al., as we record fewer physical DSBs per SSB than they do based on a 17.5 eV break threshold. Similarly, we expect to have up to twice as much indirect damage, as the Geant4-DNA models for chemistry have about 1.5 times the amount of $^{\bullet}$OH radicals as those used in the emulated study (c.f. Karamitros, 2013; Terrissol & Beaudré, 1990).

Figure 3.31 presents a comparison between our break yields and those of Nikjoo et al. for 300 eV and 4500 eV primary electrons. Notably, Geant4-DNA finds significantly more damage in this geometry than Nikjoo et al., across most energies studied, overestimating the SSB yield by 1.5, and the DSB yield from 1.5 to 3 times (Table 3.4, see also Tables C.12 and C.13). This is attributable largely to indirect damage, which is significantly more influential in our work than in that of Nikjoo et al. across the range of energies considered. Across these energies, approximately 20% of DSBs occur in the absence of any energy deposition in the DNA. Indirect damage has pronounced effects on break complexity. When considering only direct damage, we noticed a distribution of strand breaks largely similar to Nikjoo et al., however as indirect breaks become more dominant, more complex DSBs are noticed, such that the number of DSB+ and DSB++ classifications approach each other.

The large level of indirect damage recorded in the Geant4-DNA model is not necessarily an error, but is worth investigating. When scavenging is weak, 80% (Daly, 2012) up to 96%

Table 3.3 Parameters required in order to reproduce the work of Nikjoo et al. (1997).

Lower limit for physical damage	17.5 eV
Upper limit for physical damage	17.5 eV
Radius for direct damage	6 Å
Distance from strand to kill radicals	4 nm
Simulation end time	1 ns
Simulation maximum time step	500 ps
$p_{\mathrm{SSB}}(\mathrm{radical}+\mathrm{strand})$	0.4
$p_{\mathrm{BD}}(\mathrm{radical}+\mathrm{base})$	1.00
$p_{\mathrm{SSB}}(\mathrm{radical}+\mathrm{base})$	0.00
Physics List	Option 4
Distance between SSBs to induce a DSB (d_{DSB})	10 bp
Distance between two damaged segments (d_{s})	100 bp

Table 3.4 SSB and DSB yields when reproducing Nikjoo et al. (1997)

Energy	Geant4-DNA		Nikjoo et al. (1997)	
(eV)	SSB $\mathrm{Gy}^{-1}\,\mathrm{Mbp}^{-1}$	DSB $\mathrm{Gy}^{-1}\,\mathrm{Mbp}^{-1}$	SSB $\mathrm{Gy}^{-1}\,\mathrm{Mbp}^{-1}$	DSB $\mathrm{Gy}^{-1}\,\mathrm{Mbp}^{-1}$
300	0.265	0.039	0.163	0.015
500	0.246	0.038	0.163	0.013
1000	0.257	0.035	0.156	0.013
3000	0.297	0.031	-	-
4500	0.305	0.028	0.286	0.018

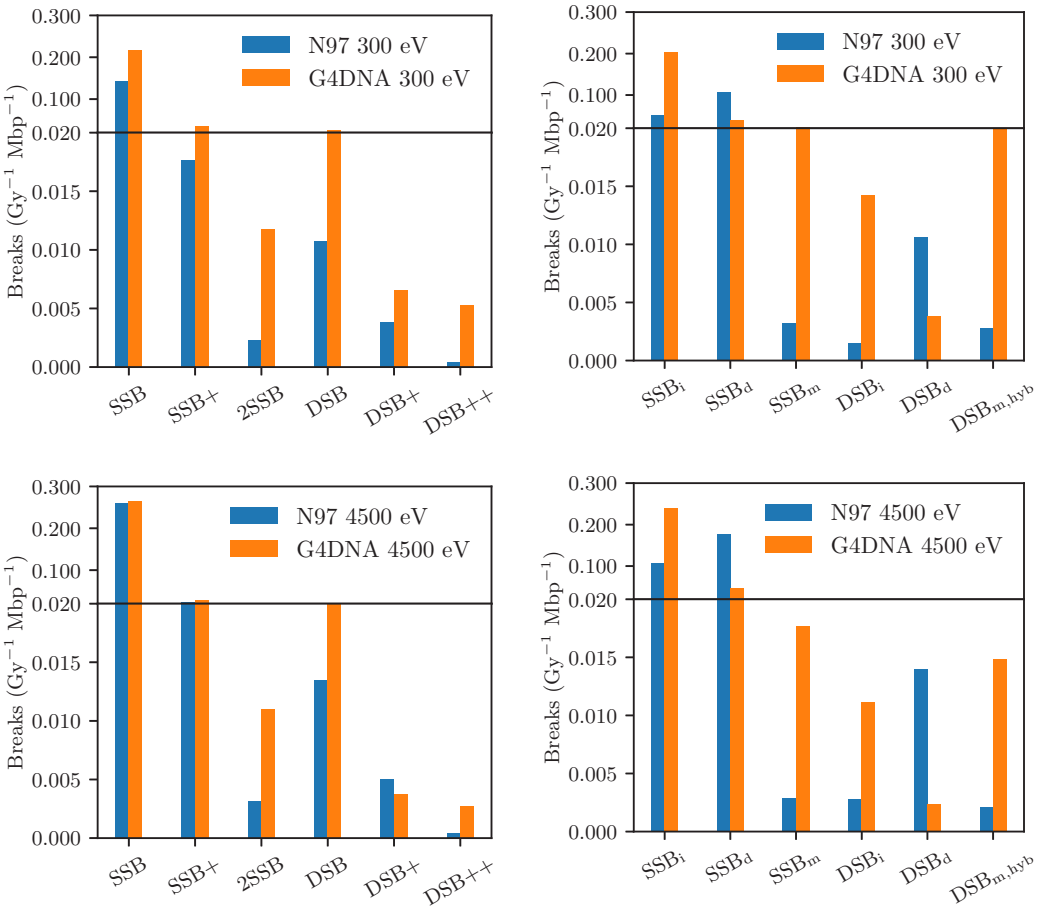


Fig. 3.31 Using the parameters defined in Table 3.3, we compare the yields of breaks by source and complexity simulated in Geant4-DNA and by (Nikjoo et al., 1997) for 300 eV and 4500 eV electrons. This information is presented for all energies in Tables C.12 and C.13. The DSB_{hyb} type dominates the DSB_m type, though we consider them together as their origins are similar.

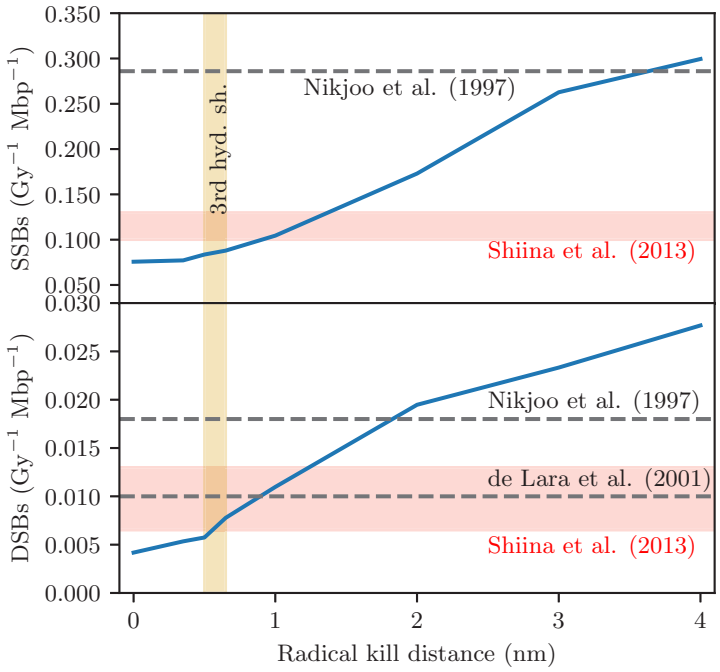


Fig. 3.32 As more and more radicals are simulated under the conditions postulated by Nikjoo et al. (1997), SSBs and DSBs increase (the above graph is for 4.5 keV primary electrons). A fair comparison with biological data where scavenging is high would only simulate radicals out to the third or so hydration shell of DNA, at 6.5 Å (Nakano et al., 2016). In this region we come close to matching the break yields from cellular irradiation by 4.5 keV X-rays de Lara et al. (2001), plasmid irradiation with very high scavenger concentration with soft (4 keV µm ^{1}) X-rays (Shiina et al., 2013).

(Souici et al., 2016) of DNA damage is caused by indirect effects. The simulation described here is in many ways not a situation of 'strong' scavenging, as radicals outside of the DNA hydration shells are simulated and tracked. This can be examined further by comparing our results to biological data, particular at varying simulation radii to consider how our simulations in a high scavenging situation, where only the first few hydration shells of DNA are simulated, compare to experimental results. This provides a decent comparison to real cellular systems as in cells, structural proteins contribute more significantly to reducing radical abundances than scavenger molecules (Ljungman et al., 1991), and the majority of indirect damage comes from non-scavengeable radicals created very close to DNA (Daly, 2012).

The results of this comparison for 4.5 keV primary electrons is shown in Figure 3.32, in comparison to experimental data where scavenging should limit most radical damage caused far from DNA. The simulations conducted by Nikjoo et al. overestimate the damage to DNA,

however we see here that when we consider only radicals created very close to the DNA molecule, good agreement with experimental data is found, particularly towards 1 nm. Of course, while based upon sound reasoning, these numbers are also somewhat tuned, being tied to a physical damage model. At 1 nm also, the ratio of SSBs to DSBs is 9.5, a little lower than expected (Shiina et al., 2013, find that $SSB/DSB \approx 12$).

The changes in break complexity and source as the number of radicals simulated is increased is shown in Figure 3.33. In these plots, we paid attention to simulation radii for radicals corresponding to the first three hydration shells of the B-DNA molecule, at 3.5, 5.0 and 6.5 Å. Water molecules within these radii are important in maintaining the shape of the B-DNA molecule, and interact with it via hydrogen bonding (Nakano et al., 2016). As expected, we are never able to simulate the same amount of physical damage as Nikjoo et al. achieve, which is a consequence of the physics models used. Between 1 and 2 nm, our distributions of break complexity best match those of Nikjoo et al..

Discussion of results While our comparison with Nikjoo et al.'s simulation is far from perfect, it highlights the importance of considering how radicals are modelled and in particular how their scavenging is considered. The fairest comparison that can be made with biological data is in the high scavenging case, where indirect damage is almost exclusively caused by radicals created in the hydration layers of DNA. Alternatively, experiments exist measuring plasmid damage in the near-total absence of scavengers, where indirect effects represent up to 96% of DNA damage (Souici et al., 2016). Here, SSB yields can be 100 times their cellular level, and DSB yields 50 times their cellular level (Butterworth et al., 2008), representing a vastly different regime to what is often simulated, but providing a point of comparison that is independent of any speculated modelling of scavengers.

When we only consider radicals created close to DNA, we find a good comparison to experimentally measured strand break yields (c.f. Figure 3.32). Tuning of the simulation could produce a better agreement with biological data, however part of the strength of our comparison here is that we use values that are already published in the literature, rather than attempting to find a set of ideal values and justifying them *a posteriori*. Part of what improves our comparison though when we only consider radicals close to the DNA is the exclusion of radicals at either end of each 216 bp long DNA segment. These are not a large issue for Nikjoo et al. who simulate long chords of DNA, however here they can contribute to an excess of indirect breaks. Reducing the radius at which radicals interact with strands has a disproportionate effect on reducing the effects of indirect damage on the ends of the modelled base pairs. This is less of an issue when we consider an entire *E. coli* bacterium in the following section, as the mostly continuous DNA model we adopt has comparatively

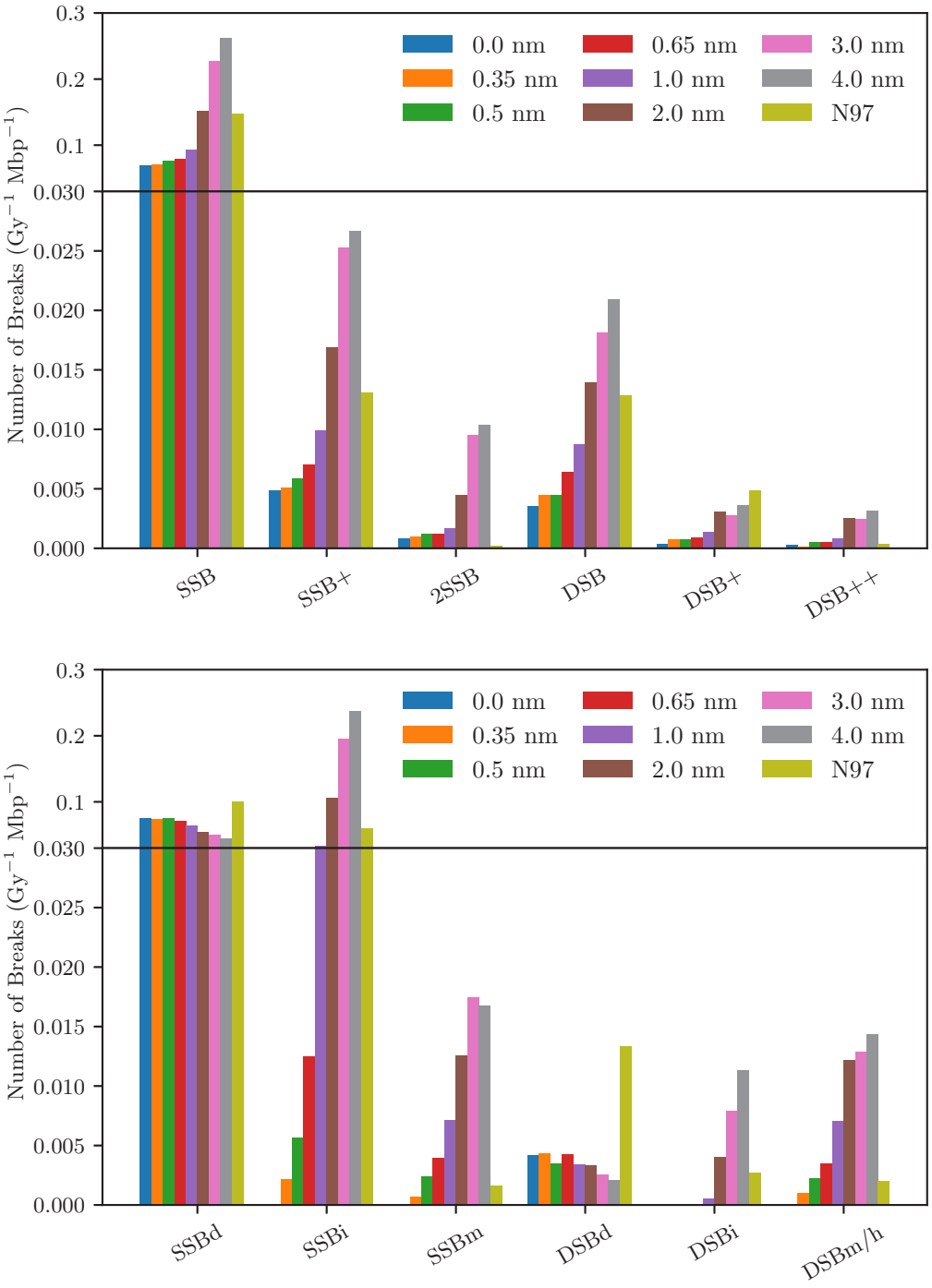


Fig. 3.33 The yields of each type of break differ significantly from that predicted in Nikjoo et al. (1997) as the radius at which radicals are considered changes.

few loose ends.

We have consistently shown that we model less direct damage than Nikjoo et al. in their work. We do not believe this to be an error in our approach, as other works have found that indirect damage dominates direct damage in both Monte Carlo simulation (Friedland et al., 2003; Meylan, 2016), contributing at least two thirds of all damage, more in line with what we show, when we consider radicals out to 4 nm from DNA. This highlights again the importance of the modelling of scavenging to the interpretation of results. These two works offer again a different conception of scavenging, which considers a 2.5 µs-long chemical simulation, which leads them to consider radicals that diffuse further than Nikjoo et al. and our work permit. Evidently, a clearer picture needs to emerge of how scavenging is best studied, however such a model requires a clear experimental picture of how many strand breaks come from indirect causes compared to direct causes.

3.5 Simulation of an *E. coli* bacterium.

The final step in our simulation studies is to determine the number of strand breaks experienced by an *E. coli* bacterium in the natural radiation background. From this, we can approach, almost as far as physically possible, the number of mutations caused by the radiation background in our evolution experiments. We adopt the same set of parameters as we used when comparing our simulations to Nikjoo et al. (1997) in the above section (Table 3.3), however the physics list is changed to Geant4-DNA Option 7. The option 7 list uses the option 4 processes for electrons below 10 keV, but passes to the default list above 10 keV, as option 4 is not defined above 10 keV for electrons. We conduct a second set of simulations considering only breaks induced by radicals created within 1 nm of the bacterial DNA, to better model the strong scavenging present in bacterial cells, induced by both scavenger molecules and folding proteins (c.f. Daly, 2012).

Simulation Geometry We simulate the genome of an *E. coli* bacterium using four, side-by-side Hilbert curve fractals with four levels of recursion (see Section 3.2). This creates 16383 placement volumes that we assign to be cubic boxes with a side length of 50 nm. We only placed placement volumes that fell inside an ellipsoid with a semi-major axis of 950 µm and two equal semi-minor axes of 400 µm, creating an elliptical geometry that corresponded roughly to the dimensions of an *E. coli* bacterium. The final geometry (Figure 3.34) contained 4.63 Mbp, similar again to the length of an *E. coli* genome. This was composed of 3,600 straight segments, and 5,652 turned segments of DNA.

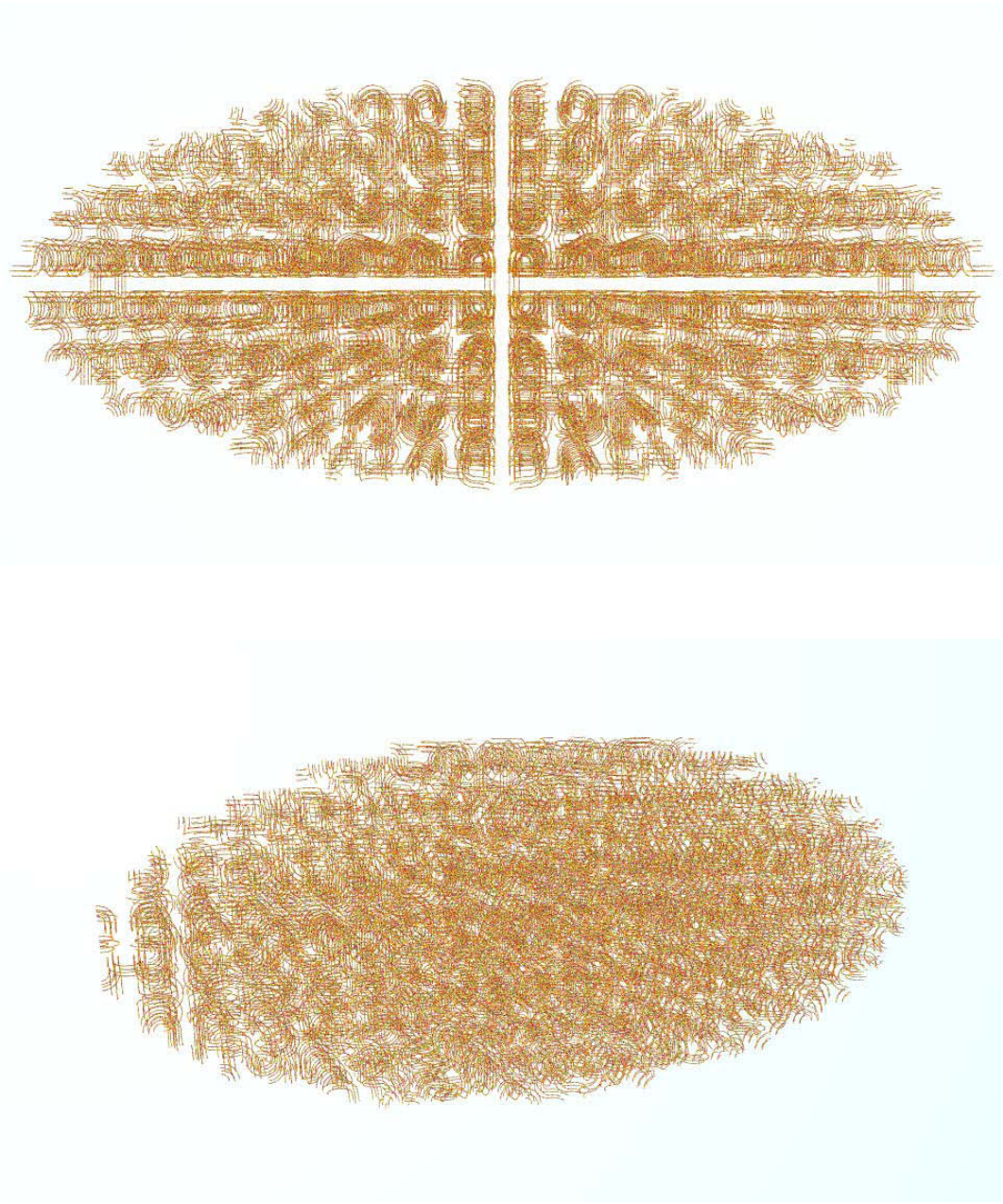


Fig. 3.34 Top: A simulated *E. coli* bacterium viewed from the front, lines can be seen delineating the main parts of the fractal. Bottom: The bacterium viewed from the side, indicating its ellipsoidal shape. The long axis of the ellipsoidal geometry is 950 μm and the two short axis are 400 μm. In total, 4.6 Mbp are shown.

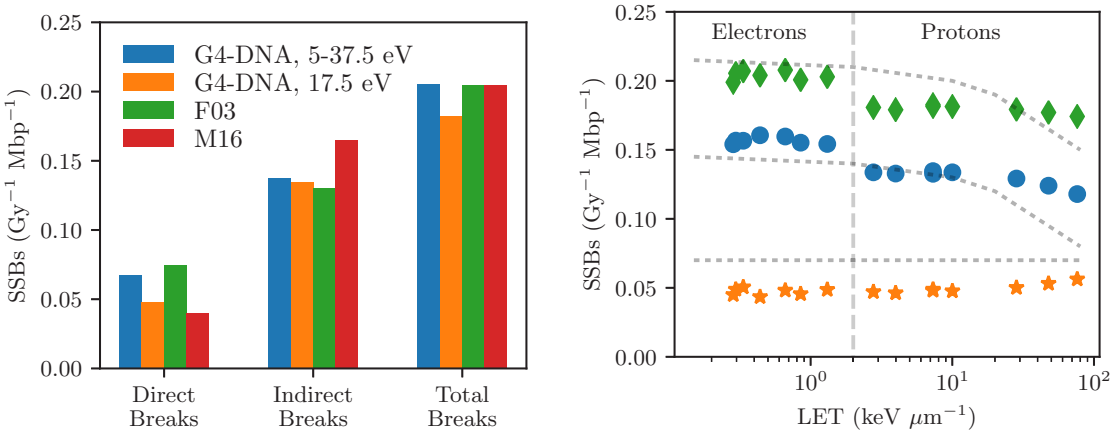


Fig. 3.35 Left: the total number of breaks from indirect and direct causes for 10 MeV protons is shown compared to proton damage yields measured by Friedland et al. (2003) and Meylan (2016). Both a 17.5 eV break induction threshold and a variable break induction threshold are considered for physical damage. Right: direct (stars), indirect (circles) and total (diamonds) break yields measured for protons and electrons measured with our simulation application and a 17.5 eV threshold for physical strand break induction. Horizontal grey lines mark the results from Friedland et al. (2003).

Electrons are simulated coming from an ellipse enclosing the bacterial cell (of the same dimensions as the cell) with energies between 1 and 990 keV, distributed approximately logarithmically. The angular distribution of electron trajectories coming from the cell surface follows a cosine law, which simulates an isotropic radiation environment. For each energy, enough events were run to deposit at least 20 MeV in the target volume, which would typically cause at least 200 DSBs. The typical statistical variation between equivalent simulation runs was 0.001 DSB Gy ^{1} Mbp ^{1} with this threshold for the number of events. Under these conditions, a limited number of simulations were also run with protons having energies up to 30 MeV in order to provide a point of comparison to other simulation work.

In order to better understand the modelling of scavenging, we again considered only radicals created very close to the DNA chain. Simulations were made at 1, 10, 100 and 990 keV, only simulating radicals within 1 nm of DNA, to provide a comparison of how distant (scavengeable) radicals impacted break yields. Again, simulations were run to ensure at least 20 MeV cumulatively was deposited in the chromosomal region.

3.5.1 Results

Before we consider double strand break yields in relation to experimental data, we consider our yields of direct and indirect strand breaks in relation to the simulations of Meylan (2016)

and Friedland et al. (2003). These two simulation platforms show, like us, that the majority of strand breaks result from indirect damage, in contrast to Nikjoo et al. (1997), whose work served as a point of comparison for our parameter sweeps. In Figure 3.35 (left panel), we show that our yields of direct and indirect strand breaks compare favourably to these two platforms for 10 MeV protons. As the PARTRAC platform, presented by Friedland et al., considers physical strand breaks to be induced by a linearly varying probability, from energy deposits starting at 5 eV up to a maximum of 37.5 eV, we have ran our simulations under this condition, as well as the 17.5 eV fixed break induction threshold (also used by Meylan). The change in the threshold for physical damage causes a significant change in the yields of direct breaks, as was indicated in our parameter testing. The 17.5 eV threshold better approximates the physical damage of Meylan, and the variable threshold better matches the PARTRAC results. Differences in the interpretation of chemical damage between each platform are responsible for the variations in indirect damage between our work and the other platforms. When compared across a range of particle energies (Figure 3.35 right), we consistently underestimate the direct damage yield from PARTRAC, though this underestimation is consistent with that which arises when using a 17.5 eV break induction threshold rather than a variable threshold. Indirect damage in general compares well to that observed by PARTRAC, though a disagreement is seen at high LETs where PARTRAC would suggest indirect damage effects start to decrease.

The yields of strand breaks for the different energies considered are shown in Figure 3.36. For low energy electrons, DSB yields are noticeably higher than for high energy electrons, where the break yield plateaus around 0.010 DSB Gy^{-1} Mbp^{-1}. These yields are lower than we found in our parameter sweeps and comparisons to (Nikjoo et al., 1997), likely due to simulating a fully continuous geometry. The yields of strand breaks compare well to those measured previously in bacteria. Bonura et al. (1975) measure using low LET 50 kVp X-rays a break yield of 0.01 DSB Gy^{-1} Mbp^{-1}, consistent with what we measure across the range $10 - 100$ keV.

As has been mentioned previously, the modelling of scavenging can play a large role in the results obtained. The results we have shown represent a modelling approach which considers damage from radicals across only 1 ns, corresponding to an $^{\bullet}$OH diffusion distance of 4 nm, and a high scavenging efficiency similar to that found in cells due to scavenger proteins. DNA folding and binding proteins also pay a large role in scavenging radicals beyond scavenger proteins, and it is worthwhile to consider in simulation only the non-scavengeable damage as a point of comparison. Thus, also shown in Figure 3.36 are yields where a 1 nm chemistry simulation radius has been considered. This reduces significantly the total yield of SSBs and DSBs (Table 3.5), producing DSB yields consistently around

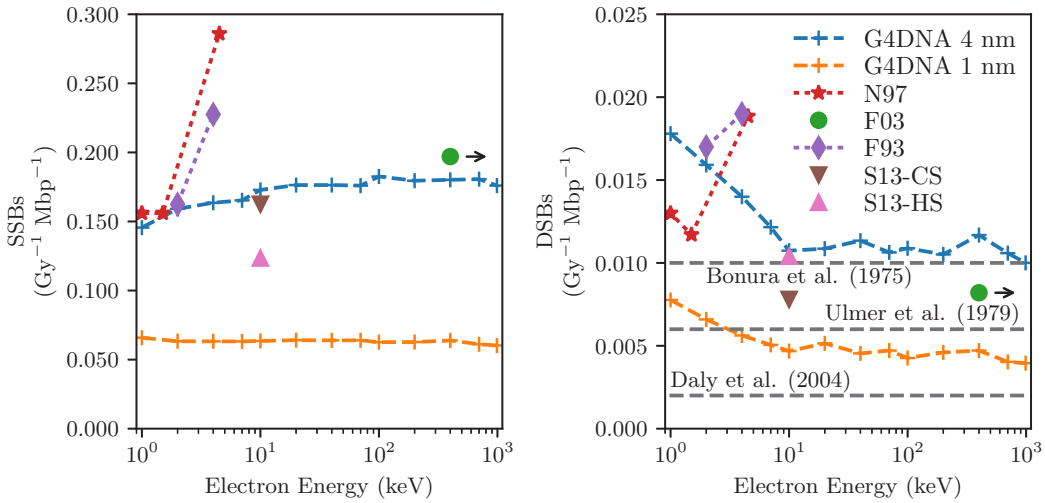


Fig. 3.36 SSB yields (left) and DSB yields (right) in *E. coli* from electron irradiation, considering chemistry within both 4 nm and 1 nm of the DNA molecule. Bonura et al. (1975), Ulmer et al. (1979) measure damage in *E. coli* from a photon source. Folkard et al. (1993) measures electron damage in plasmids, and Shiina et al. (2013) measures X-ray damage in plasmids with high scavenger abundances (the electron energy was chosen to have a similar LET). Yields are also compared with simulation data for low energy electrons (Nikjoo et al., 1997) and 30 MeV electrons (Friedland et al., 2003), as well as some plasmid measurements from electrons (Folkard et al., 1993) and X-rays at cellular (S13-CS) and high (S13-HS) scavenger concentrations (Shiina et al., 2013).

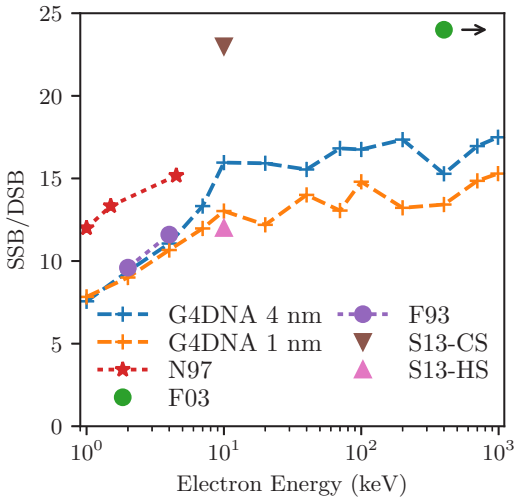


Fig. 3.37 SSB/DSB ratios for the data sets considered in Figure 3.36. Plasmid data (F93, S13-CS, S13-HS) is generally well reproduced by Geant4-DNA, however we underestimate the SSB/DSB ratios produced by other simulations (N97, F03).

Table 3.5 Impact of considering only radicals created close to DNA, mimicking unscavengeable damage.

	$r_{\text{chem}} = 1$ nm			$r_{\text{chem}} = 4$ nm		
Energy (keV)	SSBs (Gy^{-1} Mbp^{-1})	DSBs (Gy^{-1} Mbp^{-1})	SSB/DSB -	SSBs (Gy^{-1} Mbp^{-1})	DSBs (Gy^{-1} Mbp^{-1})	SSB/DSB -
1	0.054	0.007	7.8	0.118	0.015	7.6
10	0.057	0.006	13.0	0.155	0.010	16.0
100	0.057	0.006	14.8	0.166	0.010	16.8
990	0.061	0.002	24.5	0.161	0.009	17.5

0.006 DSB Gy^{-1} Mbp^{-1}. This value is consistent with break yields from ^{60}Co γ-irradiation of *E. coli* measured by sedimentation analysis (0.006 DSB Gy^{-1} Mbp^{-1}, Ulmer et al., 1979), but underestimates values found (also for ^{60}Co γ-irradiation) via pulsed field gel electrophoresis (0.002 DSB Gy^{-1} Mbp^{-1}, Daly et al., 2004). This last value seems a little low, especially given gel electrophoresis is known to sometimes underestimate DSB yields (Prise et al., 2001), however we still manage to produce it in simulation for very low LET electrons.

Averaged across the whole cell simulated, 10 keV electrons have an LET of 4.8 keV μm^{-1}. A comparison can be made then between the $r_{\text{chem}} = 1$ nm simulations and plasmids in very high scavenging environments. In particular, Shiina et al. (2013) consider X-rays with a similar LET (4 keV μm^{-1}), and find the yield of DSBs to be 0.010 ± 0.03 DSB Gy^{-1} Mbp^{-1} and the SSB yield to be 0.124 ± 0.013 SSB Gy^{-1} Mbp^{-1}. These are both higher than our yields, though our DSB yield at 10 keV is at the lower bound of the measured uncertainty.

For both values of r_{chem} considered, the ratio of single to double strand breaks induced by electrons measured by Folkard et al. (1993) compares quite well to our simulations. However, when considering the yields of SSBs and DSBs on their own, Folkard et al. produce yields significantly in excess of ours, more resemblant of the yields we record when we consider our test geometry. By assuming that high scavenger densities can be modelled by only simulating radicals less than 1 nm from DNA, we were able to reproduce experimental SSB/DSB ratios made with high scavenger abundancies (Figure 3.37, see the comparison to S13-HS). We consistently however underestimate the SSB/DSB ratios produced by other simulation works. Furthermore, we find that simulations conducted with radicals out to 4 nm better reproduce SSB and DSB yields seen in other simulations and in plasmid experiments, however we underline that, especially when comparing raw yields to plasmid yields, the differing geometries could be responsible for part of the difference.

The structure and source of strand breaks has some dependence on the energy of the irradiating electrons, particularly for low energy electrons (Figure 3.17). Notably, very low

energy electrons are more likely to cause DSBs than high energy electrons. The distributions seen here echo largely the findings seen in our parameter sweeps, as expected as break classification is not highly dependent upon the global geometry.

3.5.2 Discussion

These results represent the first complete simulation of mechanistic DNA yields in a bacterium. We achieved this by developing a versatile simulation platform built as part of the Geant4-DNA framework, that can easily accommodate a wide variety of geometries. This work establishes the adequacy of this application for mechanistic DNA damage simulations, in particular we have shown good agreement with experimental results with a minimum of parameter tuning, having rather focused our parameter set on parameters that correspond well to those identified in previous simulation studies.

An interesting observation from this study is the significantly lower damage yields seen in the *E. coli* model compared to both plasmid damage yields, and our test geometry. Naïvely, one would expect that as damage yields are normalised by the number of base pairs in the geometry, damage yields are universal, however the large scale geometry of DNA can frustrate this assumption, particularly when the DNA being considered is non-continuous. Damage yields grow as the number of independent DNA segments that are hit increases. This is proportional to the density of DNA, but is also a function of the geometry. In cellular geometries, the overall continuity of the DNA imposes some restrictions on how many DNA segments can coincide with a track, whilst in a random geometry, one track can cover many more segments of DNA for a given DNA density. The same is true for plasmids, which in terms of geometrical order sit somewhere between cellular DNA, and randomly placed DNA segments. This was briefly tested by running bacterial DNA simulations with both one quarter and twice the typical bacterial DNA density, and for the cellular geometry, we found that yields were independent of the DNA density because the large scale packing order was preserved.

Our measured strand break yields are in good agreement with experimental data measuring radiation induced strand breaks in *E. coli*. Such measurements for photon sources suggest a DSB yield between 0.002 and 0.010 DSB Gy^{-1} Mbp^{-1}, which matches our predictions, albeit across a range of different considerations for scavenging conditions. For electron induced damage, data exists only for plasmids, thus largely due to the different geometry of plasmid DNA compared to bacterial DNA, absolute damage yields are not comparable, however the ratio of SSBs to DSBs may translate. Our results, regardless of whether radicals far from DNA are killed or not are in excellent agreement with low energy electron damage

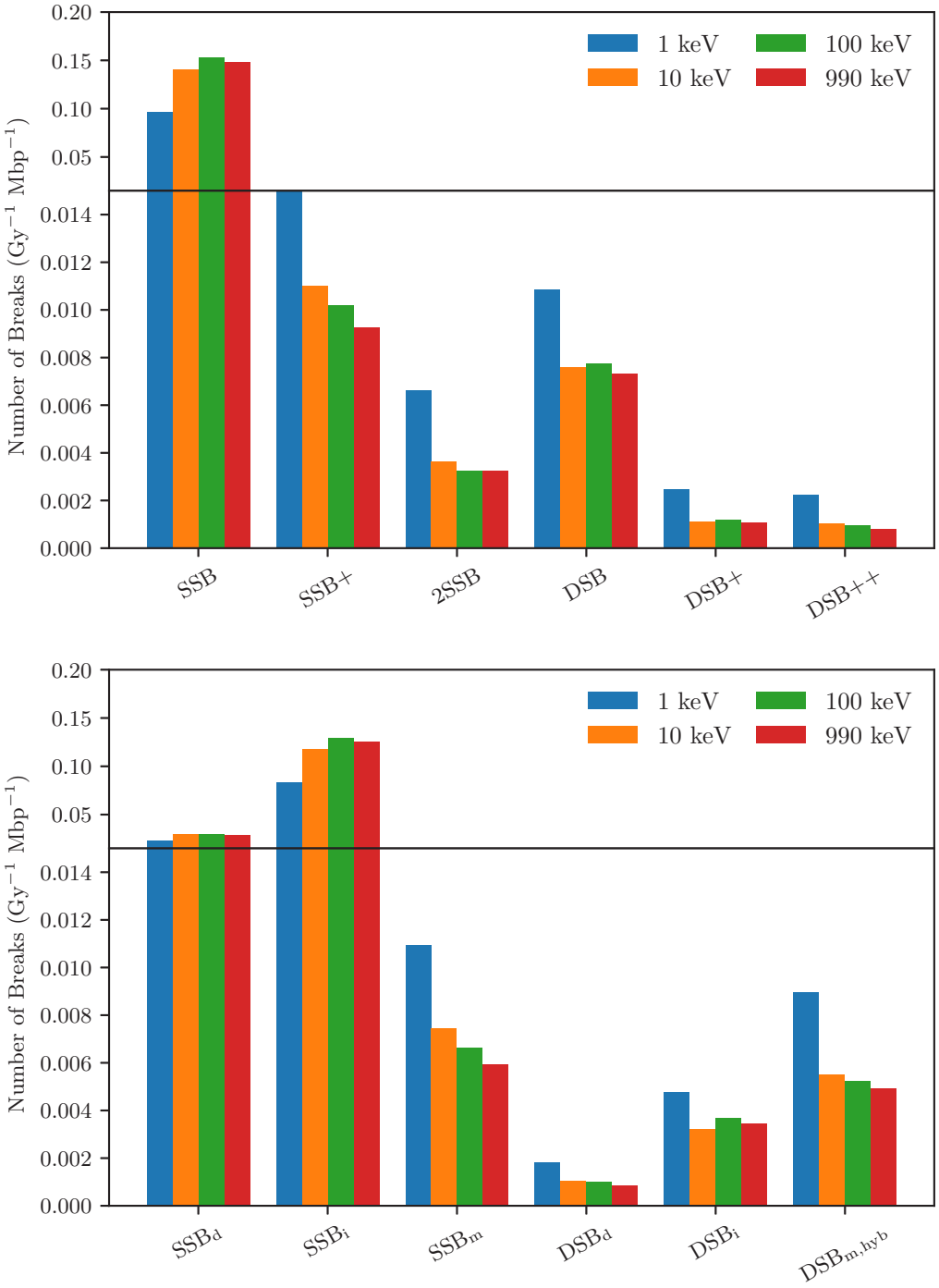


Fig. 3.38 The distribution of strand breaks by type and complexity in *E. coli* for different energies. Hybrid and mixed DSBs are considered together.

yields measured by Folkard et al. (1993), and correspond well to damage from X-rays of similar LET to 10 keV electrons measured by Shiina et al. (2013).

The breakdown of damage yields by source is again, in disagreement with that found by Nikjoo et al. (1997), but in quite good agreement with that observed in the PARTRAC simulation platform (Friedland et al., 2011, 2003) and the work of (Meylan, 2016). At low LETs, the majority of damage should come from indirect sources (e.g. Hirayama et al., 2009), a conclusion that Nikjoo et al. do not match. Meylan ran simulations using the Geant4-DNA toolkit, and thus differences between our work and his come from differences in the sensitive volume considered for direct damage, the physics list used (default or option 4), and the treatment of chemistry. In particular, Meylan overestimates the reaction rate between $^{\bullet}$OH radicals and the DNA backbone, by considering reactions with deoxyribose rather than deoxyribose-5-phosphate (which better approximates DNA structure), and chooses a poorly motivated SSB induction efficiency. The incredibly tight agreement between the total break yields measured by Meylan and the PARTRAC work seems suspiciously tuned in this regard. It's worth highlighting that our work was built around assumptions based on establishing the parameter set most equivalent to that used by Nikjoo et al., and the only change we need make in order to produce an agreement with both direct and indirect break yields in PARTRAC is to the physical damage induction model, a factor which we know from parameter sweeps can cause large variations in physical damage yields (c.f. Figure 3.20). At high LETs, we do not observe as significant a drop in DSB yield as is seen by PARTRAC, which is possibly due to errors modelling chemistry at high LET. When radicals are extremely abundant, modelling radiolysis via a particle based approximation can introduce errors, and more accurate simulations can be realised by considering radical concentrations and concentration gradients.

Simulation of only radicals created in very close proximity to constituent DNA molecules, in an attempt to model only unscavengeable radical damage lowered radical yields from a lower limit of ≈ 0.010 DSB Gy^{-1} Mbp^{-1} to ≈ 0.006 DSB Gy^{-1} Mbp^{-1}. Across a wide range of organisms, DNA damage yields are consistently around $0.005-0.006$ DSB Gy^{-1} Mbp^{-1} (Daly, 2012), suggesting this is something of a natural limit to how much organisms can naturally reduce the impact of ionising radiation on DNA. If this does reflect a natural limit, it should almost be expected to be found when only non-scavengeable damage is considered, as natural selection would tend to favour the development of cellular systems that scavenge all radicals that can be scavenged.

This work was conducted in the context of investigating the impact of the natural radiation background on bacterial cells, in order to limit their mutation rate. In Chapter 2, we showed that the natural radiation background strikes a given cell in a 24 hour period with a probability

of 6×10^{-5}. We also showed that this deposits, a median energy of 140 eV in a cell, though Auger electrons can cause 600 eV and 1.2 keV energy deposits with slightly elevated frequencies. For the cell considered in Chapter 2, a median dose absorbed of 10 mGy is found, suggesting every encounter of the cell with ionising radiation induces $\approx 5 \times 10^{-4}$ DSBs and $\approx 6 \times 10^{-3}$ SSBs. Thus the frequency with which DNA is actually damaged by the radiation background in bacteria is extremely low. In this context, it is really only rare events that deposit significant amounts of energy which are going to be responsible for DNA damage. For a 140 eV energy deposition, the amount of energy required to induce two strand breaks near each other as a fraction of total energy deposited is a significant fraction of the energy deposited, and given such deposits typically come from low LET events that do not have tightly correlated spatial energy depositions.

It has been speculated that Auger electrons emitted following the electron capture (γ emission) decay of ^{40}K could be highly mutagenic (Moore & Sastry, 1982). These electrons have energies from $2.5 - 3.2$ keV, the majority of which would be deposited in the cell (Bé et al., 2010, 1999). At a maximum though, we estimate that these electrons would cause around ≈ 0.05 DSB decay^{-1}, based on simulations of randomly oriented 4 keV electrons, which while significant, is significantly reduced in impact by the infrequence of electron capture decays occurring inside a bacterium $\approx 2 \times 10^{-9}$ day^{-1}.

Based on this analysis, we can conclude that mutations that have their origin in the radiation background are exceedingly rare. It is difficult to correlate SSBs and DSBs with mutation events, and nearly impossible to correlate them with beneficial mutation events, but we can limit the frequency with which radiation causes DNA to break (via SSBs or DSBs) in bacteria to $\approx 4 \times 10^{-7}$SSB day^{-1} cell^{-1}. Because of the uncertainty in our measurements of DNA damage yields, which depend on a number of inexact parameters, we stress the approximateness of this number. Nevertheless, as an order of magnitude estimate, we can see that DNA damage events are exceedingly rare where the radiation background is considered. Given the rarity of DNA damage events, it is qualitatively likely that changes in the radiation level will manifest themselves in the regulation of cellular processes, measurable through gene regulation and protein abundances, rather than through genetic damage, or evolutionary changes. This finds implicit support in environmental measurements in high radiation environments such as Chernobyl (Galván et al., 2014), and underground biological experiments (Castillo et al., 2015).

3.6 Conclusions and Perspectives

The work conducted in this chapter supports our conclusions from Chapter 2, namely that the ionising radiation background contributes negligibly to the mutation rate, are further supported in this investigation. SSBs and DSBs are significantly rarer than interactions between cells and background radiation sources. Because of this rarity, it is probable that if small changes in the radiation background are to influence living systems, these changes will manifest themselves in regulatory systems, rather than genetic changes. Biological systems are complicated though, and these claims require experimental examination, which is the target of Chapter 4.

This chapter has also delivered a Geant4 user application that combines physics, chemistry and user-defined cell geometries in one flexible simulation package. This is a considerable milestone in the Geant4-DNA project, and permits many experimental systems to be explored in simulation. By adapting the geometries we have defined here, chromatin structures may be defined as well as plasmids. Clever use of the underlying file definition structure could even allow users to define complex DNA loop structures, and full chromosome domains for human cells. As work proceeds to better predict the mechanical causes of DNA damage, this has the potential to be quite useful in bridging the gap between heuristic models of cell death, such as the local effect model, and the underlying causes of cell death. A longer term extension of this would be towards using Geant4-DNA to measure radiation induced damage to other cell structures, such as proteins and membranes. This may be of particular importance to the study of how gold nano-particles can increase the effective dose in a region, for example.

The DNA damage measured in these simulations is based on parameters that were chosen for their compatibility with existing simulation platforms. However, these parameters are essentially based on fits to data, and a certain level of *a posteriori* logic. In many ways this is necessary, in order to develop mechanistic DNA damage simulations, and is a consequence of any one group lacking the resources to thoroughly explore the vast array of reaction rates and damage models that are possible in these simulations. It is hoped that by sharing the application developed here with the community, a wide parameter space will be explored. Tightly controlled experiments may be devised to assess where the assumptions behind many of the parameters we chose break down, and how the underlying physics models can alter results. Assumptions about scavenging and DNA damage can also be tested in a wide range of conditions, allowing, in general, a far better understanding of how radiation induces biological damage.

Chapter 4

Long Term Evolution Experiments in Different Radiation Environments

Radiation can undoubtedly trigger mutations in DNA, but its effect on evolutionary time scales is difficult to quantify. Many coincidences can be pulled together to make the case that radiation is a driving force in evolution. Early multi-cellular life is thought to have appeared in Gabon (El Albani et al., 2010, 2014) in close proximity to the Oklo natural fission reactor (Gauthier-Lafaye et al., 1996), and Potassium-40 has been hypothesised to have played a role as a primordial gene irradiator (Moore & Sastry, 1982).

More broadly though, the radiation background is an environmental factor to which life adapts. Genetic and epigenetic changes have been seen repeatedly in the Chernobyl environment, and are hinted at by experiments conducted in underground laboratories, which effectively suppress the base level of radiation on earth to which most life has adapted. Extremophiles provide an even more intriguing picture of how ionising radiation can shape life. The bacteria *Candidatus Desulforudis audaxviator*, which lives in single species colonies deep underground, metabolises molecules in minerals produced by radiolysis from Uranium, Thorium and Potassium decay (Atri, 2016; Chivian et al., 2008). *Deinococcus radiodurans* has adapted extreme radio-resistance due to exposure to extremely oxidative conditions (Battista, 1997), as has the multicellular tardigrade, a consequence of its ability to withstand incredibly dry environments (Beltrán-Pardo et al., 2013).

The impact of different radiation environments on living systems is measurable *in vivo* by studying different ecosystems with different radiation levels. On earth, the radiation level varies from that found in protected underground spaces to the very high levels that are found in areas where artesian groundwater rich in Radon leeches high radon into the air[1].

[1]The highest reported background radiation level is from Ramsar, Iran, where radiation-rich groundwater contributes in some regions to a background absorbed dose of 30 μGy hr^{-1} (Ghiassi-nejad et al., 2002)

Significant work has been carried out in the Chernobyl environment to quantify how the large release of radiation in this disaster has impacted the local environment. Interestingly, long term census data reveals that the density of large mammals in the Chernobyl zone is not suppressed due to the higher natural radiation levels present, nor is mammal abundance negatively correlated with radiation dosage (Deryabina et al., 2015). Meanwhile, many organisms studied, ranging from bacteria and plants, to arachnids and birds have shown adaptations to the environment. Across a variety of avian species at Chernobyl, amongst them barn swallows and great tits, enhanced adaptation to oxidative stresses has been noticed in populations with chronic radiation exposure (Galván et al., 2014). Despite this positive adaptation, a significant (23%) fraction of adult barn swallows in the Chernobyl area were non-reproducing. This is a significantly increased fraction of sterile adults compared to that in similar unpolluted environments (Møller et al., 2005). The same study finds that annual adult survival, hatching success and brood size were also reduced amongst barn swallows in the Chernobyl region. Examining plants, Boubriak et al. (2008) showed that haploid (birch pollen) and embryo (seeds from evening primrose) cells in γ/β-emitter contaminated environments adapted to their new radiation environments, possibly by improving their DNA repair capabilities. Intermediate levels of radiation (450 nGy hr^{-1}) near Chernobyl caused bacteria found in the feathers of barn swallows to show heightened radioresistance compared to a control population (Ruiz-Gonzalez et al., 2016). *E. coli* collected from near the Chernobyl power station have been shown to be significantly more resistant to X-rays, UVC radiation and the mutagen 4-Nitroquinoline 1-oxide compared to samples from a control site (Zavilgelsky et al., 1998).

Work done in underground laboratories supports the idea that changes in the radiation environment can cause adaptive changes in cells. In Chapter 1 recent research in underground laboratories has been summarised. Amongst the key findings however have been lower growth rates for simple organisms such as *D. radiodurans* and *Shewanella oneidensis* in low background environments (Castillo et al., 2015), as well as *Paramecium tetraurelia* and *Synechococcus lividus* (Planel et al., 1987). Reduced tolerance to radiation exposure and reduced antioxidant abundancies have been noticed in long term experiments in mammalian cells grown in the Gran Sasso National Laboratory (Carbone et al., 2009; Fratini et al., 2015; Satta et al., 1995, 2002).

The mechanisms of the changes that occur when cells are introduced to low background environments remain unexplained. It is quite possible that these responses, taken together with the ensemble of unorthodox responses of cells to low doses, such as bystander effects (Morgan, 2003a,b), genomic instability (Dubrova, 2003; Limoli et al., 1999) and radiation hormesis (Calabrese & Baldwin, 2003; Feinendegen, 2005; Kudryasheva & Rozhko, 2015)

share a common origin in protecting cells from the rare but catastrophic consequences of radiation exposure. A key difficulty in assessing this is the lack of a viable model organism that can serve as an experimental control when radiation backgrounds are changed. It is possible that such a model species does not exist in a practical sense, and thus biological processes that are well understood are needed to enable the exploration of the radiation-response of life at low doses. In this thesis, we examine the extent to which evolution can be used as a tool to probe the low background radiation response of life on long time scales. In working with *E. coli*, we are also able to begin to assess to what extent it is suitable as a control organism in different low radiation environments.

Examining the role that evolutionary studies can play in assessing the impact of radiation on life serves two purposes. While in the preceding chapters we have shown that ionising radiation is unlikely to be a significant contributor to the mutation rate, which is dominated by biological processes, it can be an agent to which cells adapt. This theoretical assertion however requires experimental validation. Long term evolution experiments in different radiation environments can assess this. In relation to low background studies, the first hypothesis we seek to test then is that reducing the radiation background does not significantly slow evolution, or more precisely, the rate at which beneficial mutations become fixed in a clonal population. Answering this question elucidates furthermore, at a crude level, whether background radiation environment on earth has been a significant driver of evolution.

If it is the case that evolutionary dynamics play out similarly in reduced and normal radiation background environments, then the mechanisms behind changes that occur in cellular systems in response to low radiation environments may be able to be explored via evolutionary experiments. It is not known if the changes that occur in these systems appear in a binary way, manifesting only below a certain threshold, or gradually. Nor is it clear if they always appear instantaneously. Resolving the apparition changes in response to a low background radiation changes in time allows them to be better studied. Combining this with the common practice of long term evolutionary studies of preserving organisms at various time points allows genomes and phenotypes before and after the apparition of behavioural changes to be explored.

In this chapter, we present a brief discussion of fitness trajectories and mutation rates, as these underpin the observations that we make in our experiments. This discussion draws on both mathematical predictions and controlled evolution experiments that study changes in mutation rates. Our specific hypotheses for a long term evolution experiment conducted in standard and low background radiation environments are then specified, framed by these works. Next, our experimental methods are presented, before our results. We finish with a discussion of the results and their implications for both the understanding of evolution at a

general level and their impact on future long term evolution experiments in varied radiation environments.

4.1 Fitness Trajectories and Mutation Rates

Fitness is a measure of the competitive advantage of a population in an environment. Naturally it is a relative measure, detailing the advantage of one population relative to another. An absolute fitness can be defined however as the advantage of one population relative to its ancestor at a set time[2]. The idea of fitness landscapes, and the step-like increases in fitness that are observed in long term evolutionary experiments have already been presented in Chapter 1. Here we talk about them specifically in relation to how they relate to an LTEE run in different radiation environments as an experimental end-point.

In Chapter 2 we quantified the impact of changing the radiation background level on the time frequency with which radiation tracks hit cells. Between the LPC and LSM, there is a 7-fold decrease in the frequency with which cells are impacted by ionising radiation. A 7-fold decrease in the mutation rate ought to be experimentally verifiable, however at the ambient radiation level, endemic biological mutations outstrip any estimated upper limit on mutations caused by ionising radiation by a factor of at least 100 (Barrick et al., 2009; Drake, 1991; Lampe et al., 2016). Based on these numbers alone, removing the radiation background shouldn't cause a change in the evolutionary comportment of bacterial systems. While unlikely, it is possible that radiation, by virtue of its ability to significantly damage DNA can have a much stronger ability to cause unique mutations that confer evolutionary benefits. A controlled evolution experiment in different radiation environments can show this. Using the LSM and LPC as environments, this requires that the experiment chosen be capable of discriminating an approximately 7-fold change in the mutation rate.

LTEE's have the capacity to distinguish changes in the beneficial mutation rate based on changes in competitive fitness that occur as a population develops in a fixed environment. These ideas are introduced in Section 1.4. Specifically, a well designed LTEE that hopes to observe different fitness behaviours in above and below-ground environments should show different fitness curves (c.f. Figure 1.9). The dynamics of fitness curves can be understood mathematically. In particular, the selection rate of beneficial mutations is dependent upon not just the mutation rate, but also the population size. As the supply of beneficial mutations increases, either by increasing the mutation rate or the population size, the changes in the fitness curve become less and less pronounced, as the population becomes overwhelmed by clones 'interfering' with each other.

[2]In a sense though, this is just fixing one relative fitness as the fitness by which all others are measured.

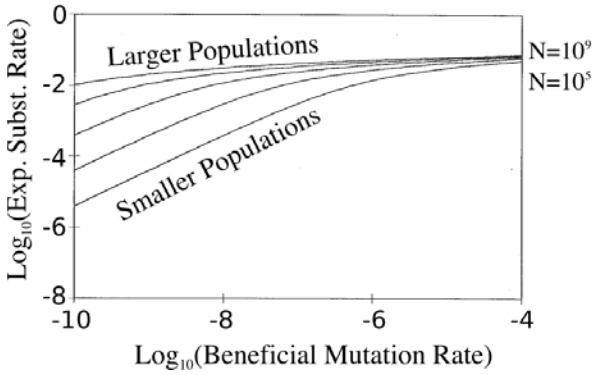


Fig. 4.1 Population size and the beneficial mutation rate both impact the rate at which mutations can become fixed in a population. For large population sizes, even large changes in the beneficial mutation rate will not necessarily cause a change in the selection (fixation) rate. (Adapted from Gerrish & Lenski, 1998).

4.1.1 Mathematical Predictions

The dynamics of evolving *E. coli* populations require an understanding of clonal interference. In a sexual population, mutations are shared between members during reproduction. In a clonal population, the mechanisms that permit this are fewer, though horizontal gene transfer notably can occur in *E. coli* (we use a strain where this is negligible). As a result, fixation of a mutation in a population requires that an ancestral bacterium undergoes a mutation, and that its progeny subsequently come to dominate the population. If one beneficial mutation arises in one cell, and another in a different cell, the two cells cannot share their mutations with each other, and for fixation to occur, one cell's children must ultimately out-compete the other's.

A consequence of this is that as the size of a clonal population increases, the rate at which beneficial mutations are selected decreases. This has been well explored by Gerrish & Lenski (1998), and its consequences have been widely discussed (Lenski et al., 1998; Sniegowski et al., 2000). In particular, Gerrish & Lenski built a mathematical model of a competing asexual population which demonstrates how clonal interference impacts the selection and fixation of beneficial mutations when the population size, mutation rate, and comparative advantage of mutations are changed. In Figure 4.1, the impact of the beneficial mutation rate and population size on the selection rate are shown. The selection rate corresponds to the rate at which beneficial mutations become fixed in our LTEE, and is thus the determining parameter for when steps are likely to occur in a fitness trajectory.

We are trying to determine whether radiation-induced mutations are responsible for the evolutionary behaviour of *E. coli*. Gerrish & Lenski estimate that in the original LTEE we mimic (Lenski et al., 1991), the beneficial mutation rate, μ_B is $\mu_B = 2 \times 10^{-9}$ beneficial

mutations per replication. Given our two radiation environments likely induce a 7-fold change in the mutation rate, it is important in designing our experiment that the effective population size we use is not larger than $N = 10^7$, or else any changes in the beneficial mutation rate caused by the radiation environment may go unnoticed due to clonal interference.

Since Gerrish & Lenski's analysis of clonal interference, other mathematical models have been developed to explain the behaviour of fitness curve. These are summarised well by Sniegowski & Gerrish (2010) who highlight that the original work by Gerrish & Lenski is built around a framework where selection of beneficial mutations is comparatively strong, while their frequency of occurrence is relatively weak. This contention has become contested over the last decade, as experimental work has indicated that beneficial mutations may not be as rare as once thought Perfeito et al. (2007). Experiments in yeast (Joseph & Hall, 2004; Wloch et al., 2001) indicate that between 2% and 5% of mutations are beneficial, and Sniegowski & Gerrish (2010) were able to model the Lenski LTEE with a beneficial mutation rate of $\mu_b = 5.7 \times 10^{-5}$.

These more recent developments complicate the estimates spoken about earlier which may guide the development of our controlled evolution experiment. Fortunately, experimental work has been done within the cadre of the Lenski LTEE that can guide us in reproducing his work at low backgrounds. A key parameter here becomes the supply of mutations S, the product of the mutation rate μ and the population size:

$$S = \mu N \tag{4.1}$$

At very low mutation supply rates, the speed with which a population adapts to an environment will rapidly grow as S increases, whilst at higher mutation supply rates, clonal interference ought to cap the speed at which populations can adapt to a medium. Ideally, we would like S to be such that, in our experiment, a 7-fold change in the mutation rate causes a significant change in the fitness curve. We can turn to the experimental work of de Visser et al. (1999), who have investigated how varying mutation supply rates impact the rate of change of fitness.

4.1.2 Experimental Predictions

As part of the Lenski LTEE, de Visser et al. (1999) explored how different mutator genes spliced into *E. coli* as well as different population sizes changed the development of fitness over 1000 generations in a series of LTEE's with different population sizes and mutation rates. Mutator genes can disrupt the fidelity of gene transcription or DNA repair to alter the mutation rate (e.g. Sniegowski et al., 1997). In de Visser et al.'s work, two mutator genes

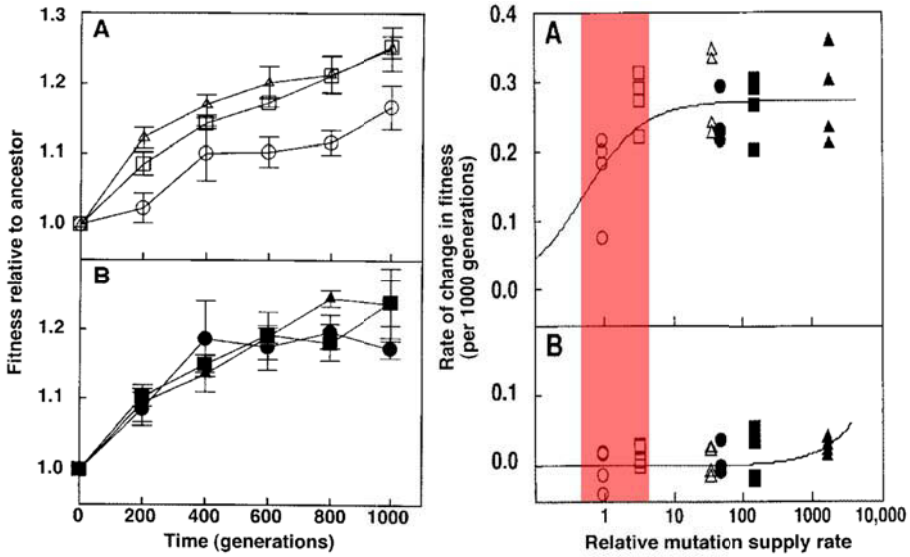


Fig. 4.2 When the mutation supply rate is changed, the rate of change of fitness also can change. The left hand panel shows the fitness relative to the ancestor for (A) small effective population sizes (hollow diamond $s_r = 1$, hollow squares $s_r = 3.3$, hollow triangles $s_r = 34.9$) and (B) large effective population sizes (filled diamond $s_r = 50$, filled squares $s_r = 165$, filled triangles $s_r = 1745$). The right hand panel shows the rate of change of fitness with relative mutation supply rate in a population that was unadapted to the growth medium (A) and one that had already spent 10,000 generations adapting to the medium (B). The region highlighted in red shows the range of relative mutation rates probed by our LTEE if radiation is a significant driver of beneficial mutations. Relative mutation supply rates S_r are given relative to that of a population with an effective size of 3.3×10^5 cells with no introduced mutator genes. Adapted from de Visser et al. (1999).

were used to vary the mutation rate, *mutY*, which increases the mutation rate by 3.3, and *mutS*, which can increase the mutation rate by a factor of 35. Additionally, one set of cell lines was studied that had already become well adapted to the growth cycle used in the LTEE, alongside one set of ancestral REL606/REL607 cells.

Figure 4.2 shows the results of de Visser et al.'s study. The rate of change of fitness, a proxy for the selection rate of beneficial mutations becomes flat for high effective population sizes and high mutation rates. The rate of change of fitness was also quite close to zero for experiments conducted using the populations that were already well adapted to the experimental conditions.

Influenced partially by these results, we have conducted our evolution experiments with an effective population size of $N = 2.5 \times 10^6$, and our baseline mutation rate is the same as that in the Lenski LTEE. In Figure 4.2, the span of mutation supply rates that apply to our experiment is highlighted in red. The upper limit on this marks the relative supply rate of the

LPC population, where the effective population size is 3.8 times higher than that used by de Visser et al. for their 'small' population. Compared to de Visser et al., the relative mutation supply rate in the LPC then is $s_r = 3.8$, and in the LSM may be as low as $s_r = 0.5$ if radiation is significantly responsible for mutations (that is, $\mu_{B,\,\mathrm{LPC}} = 7.3\mu_{B,\,\mathrm{LSM}}$). As this falls in the region where rate of change of fitness is significantly dependent upon mutation supply, it is highly likely that our LTEE will show significant differences between the two environments if ionising radiation plays an important role in the selection of beneficial mutations.

4.1.3 Summary of hypotheses

Based upon these works, we can now write the hypotheses that our experiments hope to show. From our work in Chapter 2, mutations occur far more frequently than ionising radiation interacts with cells, thus it is likely that both populations will evolve the same way and their fitnesses at the end of a 500 generation fitness experiment should be indistinguishable. We have built our experiment though to detect a difference in the mutation rate between both environments through the fitness curve. Thus, the hypotheses of our experiment can be summarised in the following points:

1. A significant difference in the fitness of the two populations after 500 generations means that radiation environments affect the selection of beneficial mutations.

2. If the fitnesses of the two populations are non-distinguishable after 500 generations, radiation may affect our populations, but in ways that are not measurable with fitness.

3. The null hypothesis is that the two populations have the same fitness after 500 generations.

4. Based on our work in Chapter 2, we do not expect to be able to reject the null hypothesis, as this work suggests that the mutation rate is dominated by biological factors, and we have no strong *a priori* reason to believe the radiation damage strongly favours beneficial mutations.

The last point above is well supported by other LTEE work. Drawing again from de Visser et al. (1999) we see that changes in the selection of beneficial mutations are achievable by just changing the mutation rate, independently of the level of radiation. Nevertheless, low dose radiobiology experiments have shown unorthodox results in the past, so we believe that despite evidence to suggest that the null hypothesis cannot be rejected, it remains a possibility.

606-1				606-5				606-9			
		607-1				607-5				607-9	
606-2				606-6				606-10			
		607-2				607-6				607-10	
606-3				606-7				606-11			
		607-3				607-7				607-11	
606-4				606-8				606-12			C-1
		607-4				607-8				607-12	C-0

Fig. 4.3 Layout of independent bacterial lines on a 96-well microplate throughout the experiment. The well labelled C-0 contains 1.5 mL of pure DM250 as a control. C-1 also contains 1.5 mL of DM250 as well as a 5 µL aliquot of the previous days control.

4.2 Method and Experimental Protocols

At both the LPC and LSM, 800 generations of *E. coli* were grown following a method similar to that used by Lenski et al. (1991). In each environment, 12 independent lines of REL606, and 12 independent lines of REL607 were grown. The bacteria were grown in a 96-well microplate (Greiner Bio-One 780271), in 1.5 mL of Davis Medium (Carlton & Brown, 1981) containing 250 mg mL ^{1} glucose (DM250). DM250 typically sustains a maximum population of 5×10^8 cells mL ^{1}. Daily transfers to a new medium placed 5 µL of bacteria from a maximally dense solution into fresh DM250, giving an effective population size of 2.5×10^6 cells. Between the start of each daily cycle and the bacteria reaching their maximum density, 8.23 generations pass. Regular fitness assays were made comparing the evolved cells to the ancestral strains.

REL606 and REL607 are two near identical *E. coli* strains derived from *E. coli* B (Jeong et al., 2009). Notably, REL606 has the genotype Ara , preventing it from digesting the sugar arabinose. REL607 is a spontaneous mutant of REL606 that is Ara$^+$ and can metabolise arabinose. When grown on tetrazolium-arabinose (TA) agar plates, the Ara$^+$ strain excretes acetic acid as a consequence of arabinose metabolisation, which acidifies the area surrounding the colony. This changes the otherwise red tetrazolium indicator to white. In this way, REL606 and REL607 are distinguishable when grown on TA plates, as REL606 appear a dark red, whilst REL607 grown a light red or pink. Importantly, this one mutation does not have any impact on the behaviour of either REL606 or REL607 when grown in DM250 (the mutation is fitness-neutral).

In the sections below, the overall principal of our LTEE is explained, before detailing the protocols used within these experiments to grow bacteria daily, pause and restart the experi-

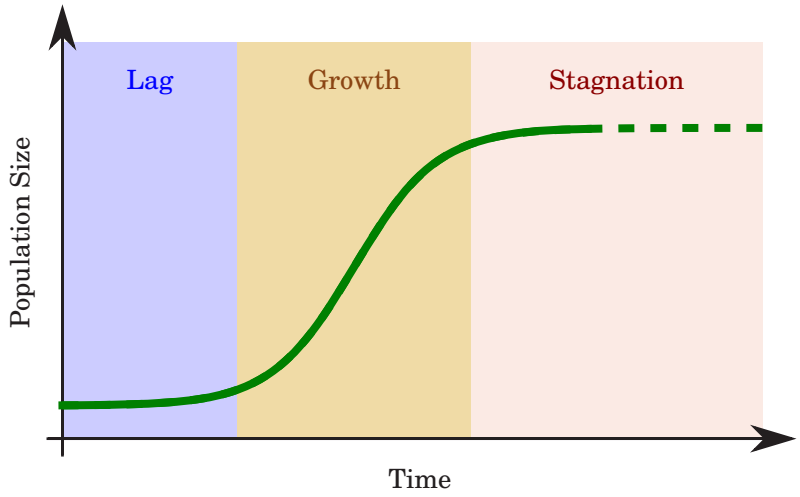


Fig. 4.4 Bacteria follow a cyclic growth pattern every day, beginning in the lag phase where they repair damage sustained in the preceding stagnation phase, and prepare to enter the growth phase. In the growth phase, cells grow exponentially until the nutrients present in the culture medium are depleted, whereupon the cells enter the stagnation phase. Here growth is halted, and cells typically enter a stressed state (reproduction of Figure 1.6).

ment using frozen plates, and conduct fitness assays. Additionally, every 100 generations, the contents of each well was diluted and spread onto a TA agar plate to ensure that there was no cross-contamination between the REL606 and REL607 lines. All growth experiments are conducted on a 96-well microplate with 24 independent lines arranged as shown in Figure 4.3.

4.2.1 Experimental Principle

In our LTEE, bacteria undergo repeated 24 hour cycles of acclimatisation to a fresh culture medium, growth, and then stagnation. This cycle (Figure 4.4) provides a cyclic stress which acts as a selective pressure on the cells. Every day, a 5 μL aliquot of each bacterial line is taken from the previous days solution and placed in fresh culture medium. In this way, cells experience a lag phase, adapting to a new medium, followed by a growth phase before finally a stagnation phase which occurs when the glucose in the solution is depleted. Following the daily transfer, cells recommence in the lag phase.

Over time, spontaneous mutations occur in the cell population, some of which are beneficial. These confer an advantage upon the cell, and there is a chance that a sub-population of cells with this mutation can dominate the population. When a single, new mutation is present in the vast majority of cells in a population, it is said to be fixed. This

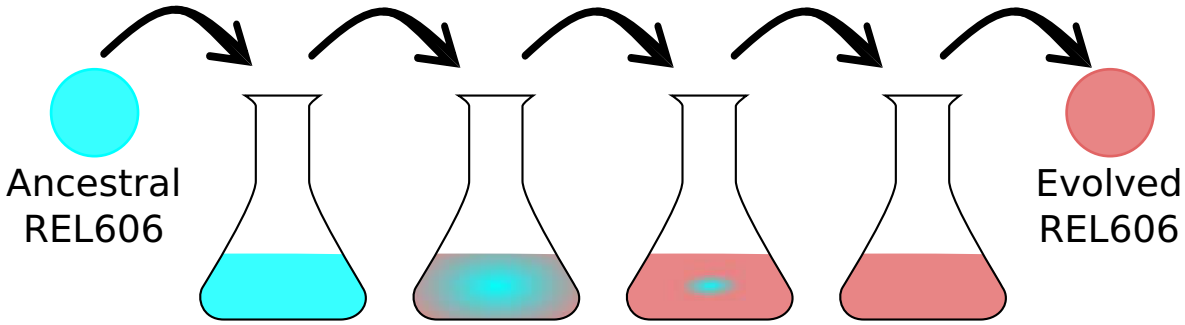


Fig. 4.5 In the LTEE we consider, daily transfer are made from a solution saturated with *E. coli* to a fresh growth medium each day. Spontaneous beneficial mutations can appear that over time and repeated transfers come to dominate the population. Here, a mutation fixing itself in the population is shown using colours, as the fraction of the flask which is red increases until the red, evolved strain of REL606 dominates entirely the original population of ancestral (blue) REL606.

scenario is illustrated graphically in Figure 4.5, where a mutation in the ancestral strain of REL606 comes to dominate the population.

The fixation of a mutation in a population can be measured by studying the fitness of the bacterial population relative to its ancestor. The mutations that become dominant in the population typically cause an increase in the fitness of the population, as beneficial mutations are significantly more likely to reach fixation than non-beneficial mutations. In order to measure fitness, regular assays are performed every 100 generations which compare how much better an evolved population grows in DM250 than the ancestral population. As ancestral REL606 and REL607 have the same behaviour when grown in DM250, the fitness of evolved REL606 strains is measured against ancestral REL607 strains and vice versa. This allows the bacteria to be placed in competition and also distinguished when grown on TA plates. The number of REL606 and REL607 colony forming units (CFUs) is measured by plating the bacteria at the beginning and end of each growth cycle (Figure 4.6).

The fitness is calculated by considering the relative growth rates of the evolved strain and the ancestral strain. Growth rates are typically considered in log-space, so considering the number of evolved N_E and ancestral N_A CFUs at the beginning and end of the 24-hour competitive growth cycle, the fitness, F can be calculated as

$$F = \frac{\log(N_{E,\,24\text{h}}/N_{E,\,0\text{h}})}{\log(N_{A,\,24\text{h}}/N_{A,\,0\text{h}})}. \tag{4.2}$$

Fitnesses of $F = 1$ indicate that the growth rate of the ancestor and evolved strains are equal. If $F < 1$ the evolved strain is less adapted to the environmental conditions than the ancestor, and if $F > 1$ mutations have occurred that increased the growth rate in the environment

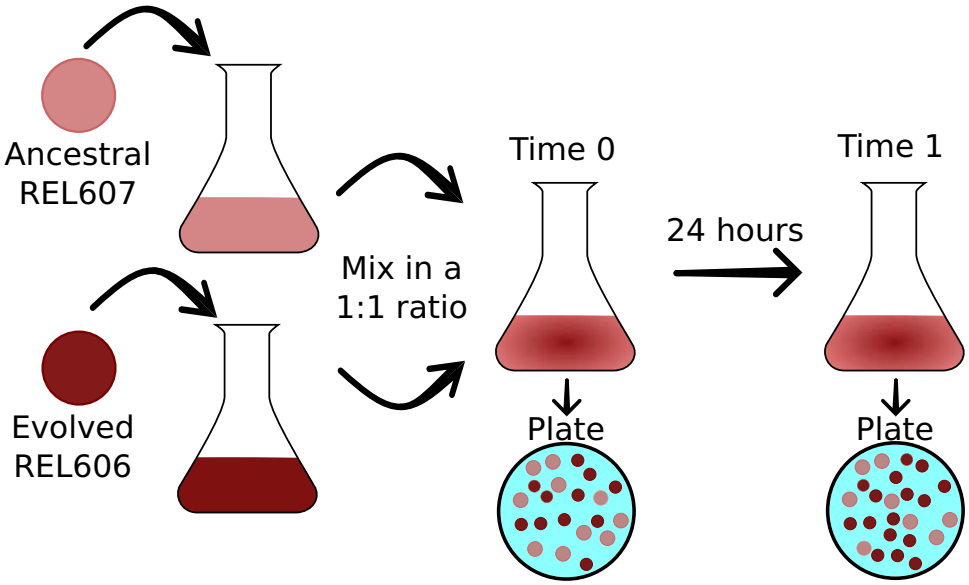


Fig. 4.6 Fitness experiments are conducted by mixing an ancestral bacterial strain with an evolved strain in a 1:1 ratio. At the start and end of a 24 hr growth cycle, the number of colonies in the mixed sample are measured by spreading them on a TA plate. From this, the growth rates for two strains can be found, and the fitness computed.

tested.

4.2.2 Preparation of solutions

DM250 Davis Medium was prepared in 1 L batches using pre-mixed Davis minimal broth without dextrose (Sigma-Aldrich 15758). Davis Medium can also be prepared from its constituent chemicals following the recipe in Carlton & Brown (1981). 10.6 g of Davis broth powder was dissolved in 1 L of water and then autoclaved for 20 min at 121°C. Next, under sterile conditions, 1 mL of 0.2% thiamine (Sigma-Aldrich T1270) solution was added, alongside 2.5 mL of 10% glucose solution (D-glucose, Sigma-Aldrich G7021). The final solution can be kept refrigerated until needed for use.

TA Plates To prepare approximately 50 TA agar plates, 1 L of agar solution was required. In a 1 L bottle, 16 g of agar, 5 g of sodium chloride, 10 g of tryptone and 1 g of yeast extract were combined. The bottle was filled with 500 mL of water, and 1 mL of antifoam was added, before mixing the solution well. 10 g of L-arabinose (Sigma-Aldrich A3256) was added to a second 500 mL bottle, before filling the bottle with 500 mL of water. The two bottles were autoclaved for 20 min at 121°C.

Immediately after the autoclave cycle had finished, the 500 mL arabinose solution was

poured into the agar solution, taking appropriate care given the mixtures were very hot. Under sterile conditions, a workspace was prepared to pour the agar into Petri dishes. When the agar solution reached 60°C, 1 mL of a 5% tetrazolium chloride (Sigma-Aldrich T8877) solution was added to the agar in sterile conditions. It is prudent to keep the agar in a heat bath to ensure it is at the correct temperature before adding the tetrazolium indicator, as the indicator is not thermostable and may change colour if it is added to a solution that is too hot. After adding the tetrazolium to the agar, the solution was poured into Petri dishes. After the Petri dishes were cooled, they were stored in a refrigerator until needed for use.

It is recommended that when Petri dishes are used in the LSM, they are placed in foil before being placed in an incubator. The lower humidity inside the LSM has a tendency to dry out the agar in each plate. This is prevented by wrapping the dishes in aluminium foil.

4.2.3 Starting the experiment from bacterial beads

The independent REL606 and REL607 lines were started from bacterial beads. 2 conical flasks (25 mL each) were filled with 10 mL of DM2000 (identical to DM250, but with 20 mL of 10% glucose solution added). Two frozen beads containing REL606 were added to one flask, and two beads were added to the other. The conical flasks were then incubated at 37°C for at least 16 hr, agitated at 170 RPM.

At the end of the incubation period, for each line 100 μL of bacterial culture was mixed with 900 μL DM250 solution and the optical density of this solution was measured. Based on the optical density measured, a 25 mL bacterial solution was prepared (diluting the bacterial culture with DM250) with a final optical density of 0.06. In each well of the 96-well experimental microplate, the two bacterial lines were distributed according to the scheme in Figure 4.3. 250 μL of diluted bacterial culture was placed in each well with 1.25 mL of DM250. The microplate was covered with sterile adhesive film and incubated at 37°C and 170 RPM for 24 hr, at the end of which the daily growth protocol was followed.

4.2.4 Using lines from a frozen microplate

When recommencing the experiment from lines stored on a frozen microplate, or using frozen lines in a fitness assay, the frozen microplate was first removed from the freezer and allowed to thaw for 15 min. Under sterile conditions, 1 μL inoculating loops were used to transfer the frozen bacteria to the required wells on the new microplate. When this was done to continue the daily evolution of bacteria, the scheme in Figure 4.3 was always followed on the target plate.

4.2.5 Daily Growth

For the daily experimental transfer of bacterial lines, each weekday, 5 μL of bacterial culture was transferred in a sterile environment from each well on the microplate from the previous day to a new microplate which had been prepared containing 24 wells filled with 1.5 mL of DM250. Transfers were made using an 8-channel multipipette, transferring one column of four wells at a time, so that transfer errors would manifest themselves as contamination between REL606 and REL607 (these can be distinguished when grown on a TA agar plate). Two control wells were used containing pure DM250 to identify daily whether an external contaminant had spread to the plate. Following the transfusion, the plates were agitated for 24 hr at 170 RPM and 37°C.

At the LPC, the agitation occurred in an incubator, whilst in the LSM cells were incubated inside lead shielding consisting of a 10 cm lead exterior and 5 cm copper interior, in order to shield the bacteria from terrestrial radiation in the laboratory.

Every 50 generations, all lineages were grown on TA plates to ensure that they contained only pure REL606 or REL607. Every 100 generations, bacteria were frozen under glycerol and stored below −25°C.

At the end of each week, microplates were placed in a refrigerator at 4°C until Monday, when the daily growth was restarted. At the LSM, refrigerated microplates were stored in copper shielding to provide some protection from ambient radiation. The shielding was a minimum of 5 cm thick all around the stored microplate. The shielding was used to reduce the likelihood of radiation-induced damage while the cell lines were under refrigeration, though shielding during the refrigeration phase isn't as necessary as it is during the growth phase, as in this phase, DNA repair mechanisms are not active, making it significantly less likely that cells that have been damaged while refrigerated can be viable for growth when the experiment recommences.

4.2.6 Freezing bacterial microplates for storage

Bacteria were stored, frozen, in 96-well microplates following the scheme in Figure 4.3. Storage in frozen plates was always performed after the transfer of lines to a new daily growth plate when both storage for freezing, and daily transfers were performed on the same day. When storing bacterial lines, each of the 24 wells to be filled with culture were first filled with 300 μL of sterile 50% glycerol (Sigma-Aldrich G5516) solution, and 450 μL of bacterial culture, in sterile conditions. After all the wells were filled, an adhesive filter was placed on the microplate and the plate was wrapped in foil before being frozen. In the LSM, the most recent frozen microplate was kept in a small radiation shield consisting of a

minimum thickness 5 cm of copper all around the plate.

4.2.7 Fitness Assays

Fitness assays were made following the protocol defined by Lenski et al. (1991). Frozen ancestral and evolved strains were recovered and grown in 1.5 mL of DM250 for 24 hr at 170 RPM and 37°C. 6 replicates were made for each line defrosted, each one seeded using a 1 μL inoculating loop. The next day, a transfer was made to a fresh well containing DM250, allowing the bacteria another 24 hr to grow. This reduces any effects that may bias the experiment from defrosting the bacteria.

Following this growth, 2.5 μL of one evolved and 2.5 μL of one ancestral strain of bacteria were mixed within 1.5 mL of DM250. The evolved and ancestral strains were always chosen to place REL606 in competition with REL607. 6 replicates were made for each pairing. Specifically, the first evolved replicate was paired with the first ancestral replicate, then the second evolved replicate was paired with the second ancestral replicate, and so on. Before incubating these competing lines, a sample was taken from them and diluted, before spreading 100 μL of diluted culture on TA agar plates. We conducted this dilution in microplates in two steps to achieve a final dilution factor near 1:1000 (we used 1:1080). The plates were labelled as T0 and allowed to grow for 24 hr at 37°C before being photographed.

After incubating the mixed lines for 24 hr, they were again diluted using DM250 and plated on TA plates, labelled T1. The dilution here was carried out in microplates in three stages. The final dilution factor should be near $1{:}5 \times 10^5$ (we used $1{:}7 \times 10^5$). The T1 plates were grown for 24 hr at 37°C before being photographed.

The colonies on the plates at T0 and T1 were counted using OpenCFU (Geissmann et al., 2013), for which we had developed an extension that performed automatic colour recognition on colonies. REL606 is identifiable as dark red colonies, and REL607 as light red colonies. The automated colour counts made in OpenCFU were supervised at all times by a trained researcher, who corrected for any misidentified colonies ($\lesssim 5\%$ of the total number of colonies).

4.3 Colony Counting With OpenCFU

In order to reduce errors when counting colonies and aid the reproducibility of counts, we used the open source program OpenCFU (Geissmann et al., 2013). OpenCFU is a C++ based tool that uses the OpenCV library (Itseez, 2015) to process images of Petri dishes and identify bacterial colonies. To aid in counting our colonies, an extension to OpenCFU was written

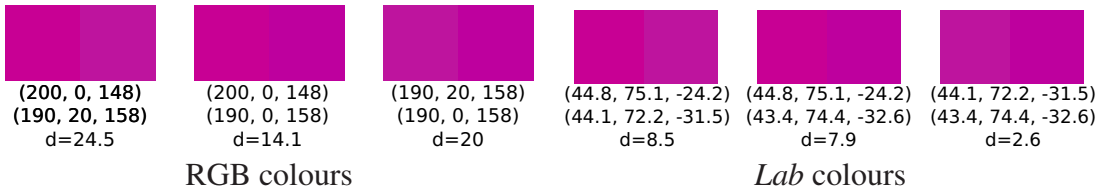


Fig. 4.7 For three different pairs of colours, the RGB and *Lab* co-ordinates of each colour pair are shown, along with the distance measure between the two sets of co-ordinates. The metric in *Lab* provides a better measure of perceptive colour difference.

that automatically counts the numbers of different coloured colonies[3]. The extension written uses the DBSCAN algorithm (Ester et al., 1996) to identify similarly coloured clusters in the *Lab* colour space[4].

To identify similarly coloured clusters, the *Lab* colour space is used as it is designed to have a metric that increases with increasing perceptive difference between colours (Figure 4.7). Most modern cameras save files using the sRGB gamut, which is a standardised RGB gamut that has been corrected for camera behaviour. OpenCFU measures the mean colour of each colony found, based on averaging the sRGB encoded pixels that belong to each colony. The built-in OpenCV methods to convert sRGB to *Lab* were used to generate a mean *Lab* colour for each colony. All the colony colours identified can now be considered as existing in a three-dimensional space specified by the three colour co-ordinates from their *Lab* colours. The DBSCAN algorithm identifies clusters of colonies based on their position in this three-dimensional colour space.

DBSCAN requires two parameters to run, the minimum number of points per cluster, and a maximum distance between points for them to be considered clustered. Using the *Lab* gamut, a Euclidean distance between two points of around 2 corresponds to a barely perceptible colour difference. When clustering colonies by colour, a minimum distance in the *Lab* space of around 5 typically returns good clustering, but this can be increased by the user for noisy data. Similarly, the minimum number of points per colony is variable, but should be at least four (corresponding to one more than the number of dimensions being clustered).

An example of the colour clustering algorithm is shown in Figure 4.8. Generally, the algorithm can recognise up to 500 colonies of different colours, at which point the Petri dishes begin to become too crowded. This limit is not intrinsic to the algorithm, but rather represents the fact that colonies have become sufficiently dense on a standard Petri dish that their separation cannot be guaranteed. About 5% of colonies are missed, often due to

[3]The OpenCFU project, with the colour counting extension can be found at http://github.com/qgeissman/OpenCFU

[4]Technically we use the $L^*a^*b^*$ colour space, also referred to as CIELAB

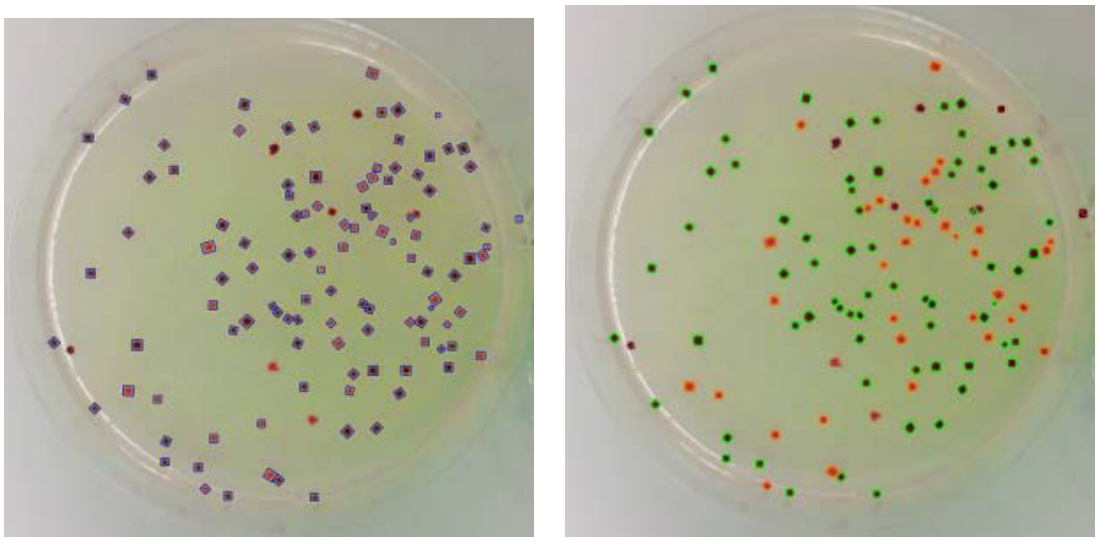


Fig. 4.8 A plate showing REL606 (dark red) and REL607 (light red) is loaded in OpenCFU and analysed (left). When colour clustering is enabled (right), the algorithm identifies the REL606 and REL607 colonies separately based on their colour (orange and green squares).

their morphology being unusual. For this reason, we supervised all our counts in OpenCFU, correcting the small errors manually made by the classification algorithm.

4.4 Results

At the LPC and LSM, contamination-free lines were selected for fitness measurements. At the LPC, measurements were made for various lines between generations 0 and 800, while at the LSM measurements were only made at 0, 200 and 500 generations. Occasionally, fitness measurements were rejected from our analysis because a very large (> 500) number of colonies had grown on the Petri dishes, and when this occurs colony counts can become biased as it is difficult to discriminate between two adjacent colonies or one large colony that has divided (see Appendix D). Our measurements were motivated by previous experimental results which suggest that by 500 generations the fitness between the two environments considered should be different under the hypothesis that radiation significantly affects the selection of beneficial mutations. Figure 4.9 justifies the selection of measurements at 0, 200 and 500 generations in particular, as we see when compared to the Ara-1 lineage studied by Lenski & Travisano (1994), the behaviour of fitness at the LPC is roughly similar. Only a minor disagreement exists at 300 generations, where only two independent lines are present in the LPC measurement.

Before analysing our data statistically, we present the ensemble of data collected graph-

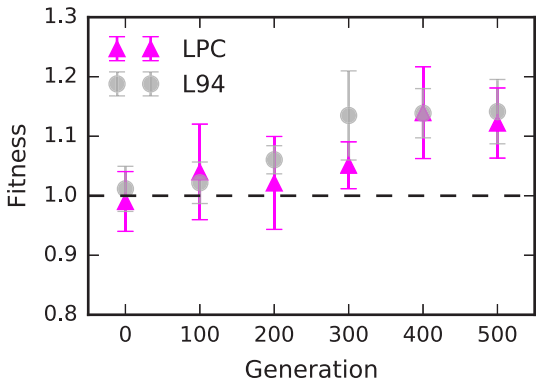


Fig. 4.9 We compare the change of fitness averaged across a number of LPC lines and one single line from Lenski & Travisano (1994). Uncertainties are larger in our lines as each data point represents multiple independent measurements (6, 3, 8, 2, 4 and 6 independent lines at 0, 100, 200, 300, 400 and 500 generations respectively).

ically. In Figure 4.10, all our fitness measurements are plotted at both the LPC and LSM. Uncertainties represent the standard deviation of the six replicates used at each measurement. Particularly large uncertainties are noticeable in the LPC REL606 line at 100 generations. On this day, this sample showed a greatly elevated colony count compared to other lines, and was excluded from later analysis on account of this abnormality.

All the lines from the LPC show a similar behaviour, while at the LSM, certain REL607 lines show a slightly depressed increase in fitness (or no increase in fitness for lineage 1), however these isolated points are not significant enough to say that the LSM lines have a slower growth in fitness. A much clearer picture comes from considering the mean fitnesses in each environment (Figure 4.11).

Mean fitnesses and their associated errors were calculated on the assumption that all the lines measured were independent. Figure 4.11 maintains a separation between REL606 and REL607 in each environment, while the mean fitnesses considered in Figure 4.12 does not. Considering the mean fitnesses by strain in the LPC (Figure 4.11a), the mean fitness follows that measured by Lenski & Travisano (1994). The Lenski & Travisano measurements however refer to only one lineage, while our measurements here are the average of multiple lines. Notably, an increase in fitness is noticed between 300-400 generations. In the LSM (Figure 4.11b), the measurements we have made again follow those of Lenski & Travisano. At 500 generations, the fitness is slightly smaller than that seen in the LPC, but the two data sets agree within uncertainties.

A statistical analysis was conducted for the LPC and LSM populations at 200 and 500 generations under the assumption that errors are normally distributed. Here, the strains

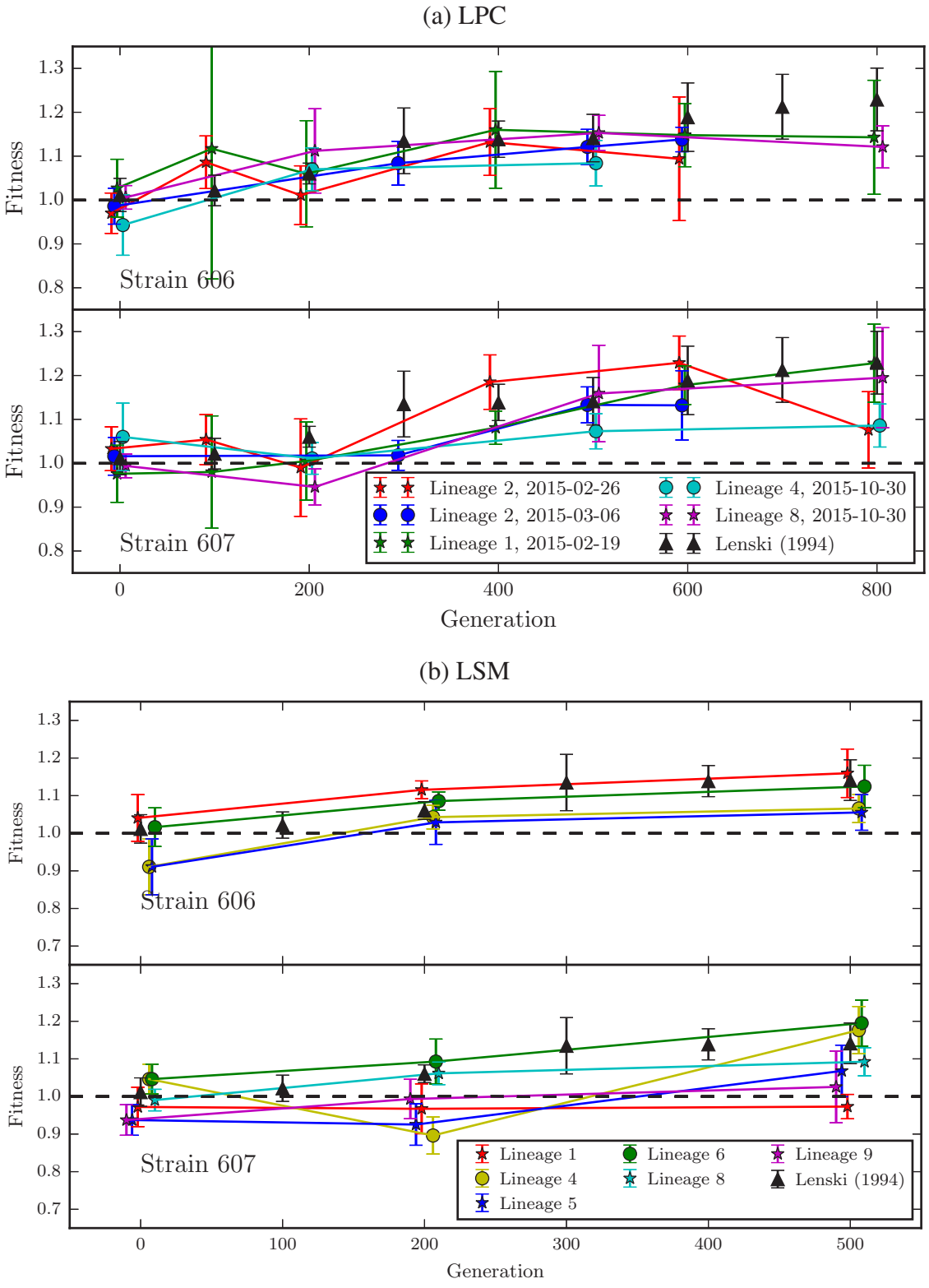


Fig. 4.10 Fitness measurements for various lineages made in the LPC and LSM.

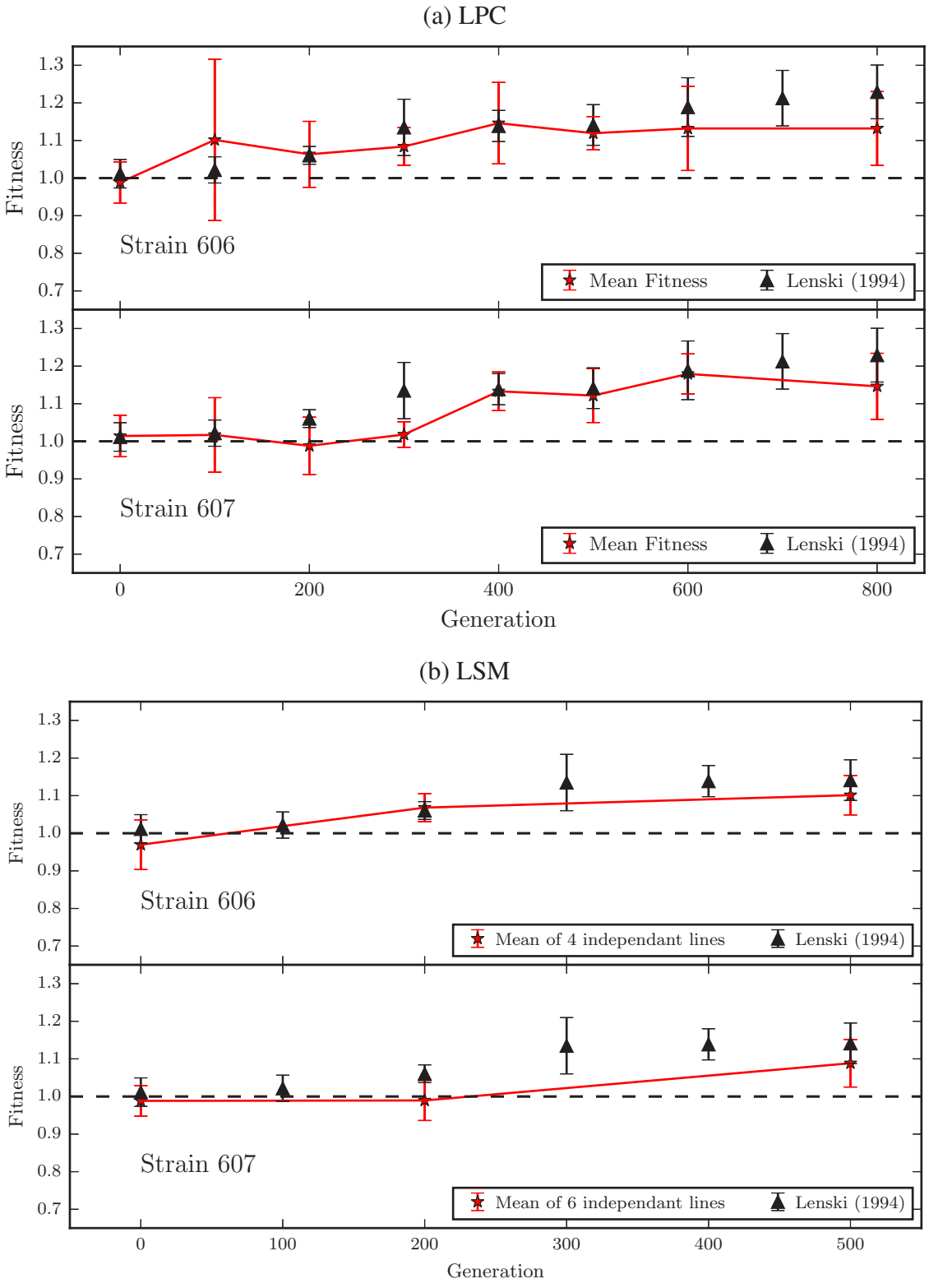


Fig. 4.11 Averaged fitness measurements from the LPC and LSM.

REL606 and REL607 were combined as their behaviour should be identical, and thus combining them increases the number of independent replicas that we study. Table 4.1 presents a summary of the likelihood that the mean fitness has changed from that at 0 generations at 200 and 500 generations. An unequal variance (Welch's) t-test was used to calculate this based on the number of independent lines at each point and their variance. At 200 generations, in both environments, the fitness is not significantly different to that at zero generations, however by 500 generations, the distribution of fitnesses of the independent lines we measured show that in both cases the fitness has significantly increased ($> 3\sigma$).

Table 4.1 Mean fitness, for n independent lines, with the likelihood that the distribution of fitness has changed since the ancestor in each environment.

Environment	Gen.	n	Fitness	$p(F > F(0))$
LPC	0	6	0.99 ± 0.05	-
	200	8	1.02 ± 0.08	0.19
	500	6	1.12 ± 0.06	1.2×10^{-3}
LSM	0	6	0.99 ± 0.05	-
	200	9	1.02 ± 0.04	0.14
	500	9	1.09 ± 0.05	8.7×10^{-4}

This is shown graphically in Figure 4.12, where the mean fitnesses are shown in each environment. To better illustrate the distribution of fitnesses that are summarised by the errorbars, the collection of fitnesses used at 0, 200 and 500 generations are also shown. Here, we can also ask whether the data points for the LPC and LSM are likely to be the same or not. A two-tailed unequal variance t-test, for 0 (200, 500) generations shows that the points have the same mean at the $p = 0.96$ $(0.87,\ 0.37)$ level. In more conventional terms, this is a test of a null hypothesis that the two points are the same. If $p < 0.05$, we could reject this null hypothesis at the 95% level, in support of the idea that the fitness trajectories in each environment are different. With the data we have recorded, we cannot do this at any of the three observation points considered.

It is worth bearing in mind that the LSM data shows a slightly larger spread than the LPC data, with some lines not greatly increasing in fitness across 500 generations. This is lost somewhat in the above analysis due to the inherent assumption of a normal distribution for our data. Implicit in this assumption is that each measurement of a lineage we take is one attempt to measure the underlying mean fitness of a controlled evolution experiment at a certain time point, and its variation. In particular, it is possible that the significance of the separations between distributions at $t = 0$ and later generations are overestimated, as we may not have measured enough lineages for the central limit theorem to apply well.

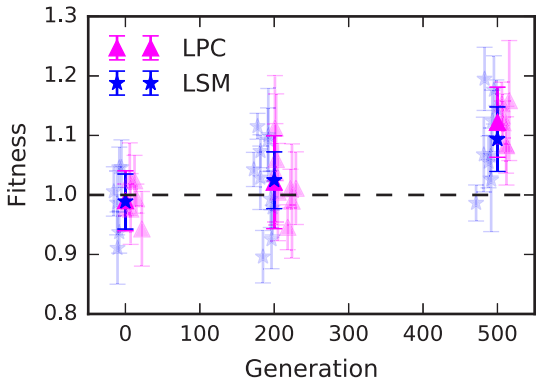


Fig. 4.12 Competitive fitness averaged across multiple populations at the LPC (red triangles) and LSM (blue stars) after 0, 200 and 500 generations of growth in each environment. The fitnesses of each individual lineage sampled are plotted transparently behind the averaged points.

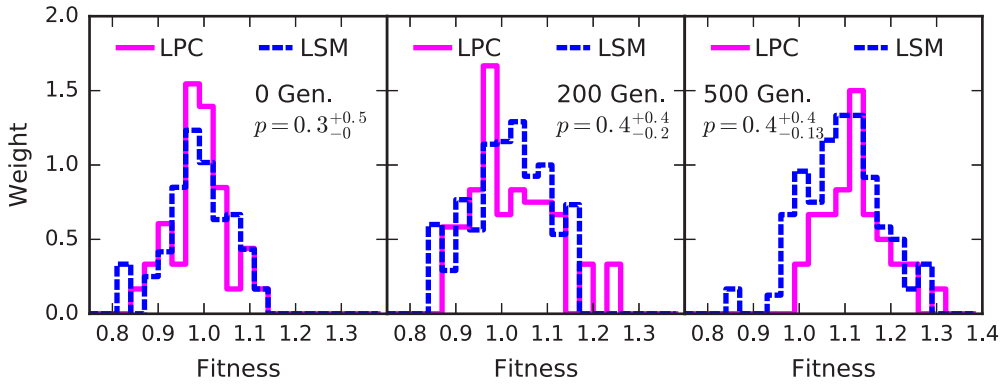


Fig. 4.13 Distributions of fitness in the LPC and LSM at 0, 200 and 500 generations. The p-value represents the likelihood that the distributions of fitness in each environment belong to the same mother distribution, measured via a bootstrapped KS test. Histograms were produced by weighting all individual measurements so that each lineage carries an equal weight.

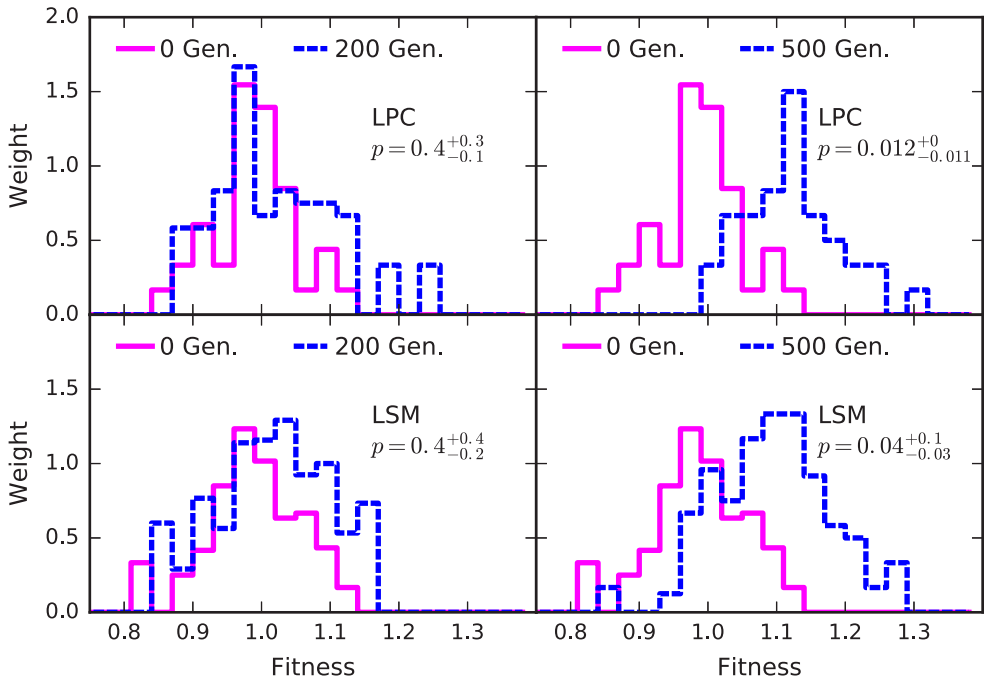


Fig. 4.14 Distributions of fitness at 200 and 500 generations compared to 0 generations, for the LPC and LSM, with p-values indicating the probability that the two distributions in each panel come from the same mother distribution, measured via a bootstrapped KS test. Histograms were produced by weighting all individual measurements so that each lineage carries an equal weight.

A more detailed statistical analysis is presented in Figures 4.13 and 4.14. Here, we show histograms of the distribution of fitnesses recorded at various times in the LSM and LPC. As a different number of replicates were sometimes used for each measurement (ideally six replicates were made, but sometimes the experiment was either repeated, yielding more than six, or some measurements failed), each replicate was weighted so that each independent line carried a weight of one in each environment at each time point. It can be seen that the distribution of fitnesses does not follow a Gaussian curve well, though it is possible that more observations would accelerate the convergence of these distributions towards a normal distribution. Ideally, a non-parametric test should be used to compare the LPC and LSM fitness measurements, as each individual measurement of colonies on a Petri dish has no uncertainty, and normality shouldn't be presumed. The two sample Kolmogorov-Smirnov (KS) test is often used in these cases however it cannot easily be used with weighted data points. To circumvent this, a bootstrapped version of the KS test was used, resampling the empirical distribution function generated by all our individual measurements.

Bootstrapping is the process whereby a distribution is randomly resampled with replace-

ment many times in order to estimate what other distributions are likely observable, when only one measured distribution exists. In the context of a a bootstrapped KS test, the empirical distribution function of the weighted measurements we have made is sampled randomly a number of times (the number of samplings corresponds to the number of independent lines in the distribution). This resampling is done for both distributions that are being considered, and the resampled values are compared using a two sample KS test. This procedure was repeated 2000 times to yield a distribution of KS statistics and their associated p-values (these distributions are discussed and presented in Appendix E). These p-values are not normally distributed, so when we present the likelihoods that the two samples compared in this way the media p-value is presented alongside its distance to the upper and lower quartile p-values.

The likelihoods that our measurements differ between the LSM and LPC are different from those presented above, but the conclusions from the data remain the same. No significant difference is noticed between the LSM and LPC at any time point measured. Due to the nature of a bootstrapped KS-test, p-values in agreement with the null hypothesis tend to produce a uniform distribution, which is the cause of the large interquartile ranges seen in Figure 4.13. We also find that the measured differences between the LSM population at 0 and 500 generations, and the LPC population at 0 and 500 generations are less significantly different than the normal distribution based analysis would suggest (c.f. Table 4.1). In particular, we find that the likelihood that the LSM population at 500 generations is distributed with a median of $p = 0.04$ and lower and upper quartiles $p = 0.14$ and $p = 0.01$, rather than $p = 8.7 \times 10^{-4}$ as the normally distributed analysis would indicate.

MALDI-TOF Mass Spectrometry Measurements The changes in fitness in the populations studied may have a few causes. Most likely, they follow the developments noticed in the populations studied by Lenski & Travisano, where early fitness gains are driven by changes in cell membrane proteins that optimise the transport of glucose into the cell (Travisano & Lenski, 1996). MALDI-TOF (Matrix Assisted Laser Desorption/Ionisation-Time Of Flight) mass spectrometry can be used to identify biomolecules including proteins. The method was used for four LPC lineages (REL606/607 lineage 1 and lineage 4) and four LSM lineages (REL606/607 lineage 1 and lineage 4). LPC measurements were made at 0 and 800 generations, and LSM measurements were made at 100 and 800 generations. While differences in the MALDI-TOF spectrum were noted between the REL606 and REL607 lineages, and between the early (0-100 generations) and late (800 generations) measurements, no significant difference was visible between the two environments. This strengthens our conjecture that the selection of beneficial mutations between the LSM and LPC environments

is, within the uncertainty of our experiments, the same.

4.5 Discussion

The main aim of this experiment was to determine whether the level of ionising radiation has a significant impact on the speed of evolution through the beneficial mutation rate. Such changes ought to be pronounced in a fitness experiment conducted under the conditions that we have used. Of course, earlier experiments showing under similar hypothetical changes in the beneficial mutation rate have revealed that the beneficial mutation rate likely scales with the overall mutation rate rather than the level of ionising radiation (de Visser et al., 1999). Nevertheless, given the propensity of ionising radiation to affect living systems in unexpected ways, this was a hypothesis that was in need of testing. From our simulations, we have already shown that ionising radiation likely causes mutations in *E. coli* 100 times less frequently than other processes (such as ROS attack and transcription errors), so it isn't too surprising that ionising radiation does not measurably affect the beneficial mutation rate.

While clear evidence exists that changing radiation environments can impact cell phenotypes, this is not clearly seen in any of our measurements. No clear sign of different biomolecule expression was seen in a preliminary MALDI-TOF analysis. Further phenotypic studies could be conducted to identify changes in cell behaviour between the environments, this could include more detailed growth analyses and measurements of cell size, following Elena et al. (1996).

Two important conclusions can however be reached based on the unchanged fitness behaviour between the two environments. The first is that it is unlikely that natural levels of ionising radiation significantly affect the beneficial mutation rate of microbial life. This is important in ecological studies of radiation affected regions, where ionising radiation becomes an evolutionary pressure. Adaptations noticed in cells in response to ionising radiation are not likely to be dependent on radiation as a cause of the mutation. Based on our results in Chapter 2, radiation levels around 20 μGy hr^{-1} are required in *E. coli* in order for the rate of radiation-cell interactions to be equivalent to the mutation rate from endogenous cell processes. At these levels, enhanced radio-resistance is inconsistently observed in cells chronically exposed to these dosages (Ruiz-Gonzalez et al., 2016; Zavilgelsky et al., 1998), suggesting that radiation dosages at this level are beginning to become an evolutionary pressure in bacteria. This however is likely independent of any change in the mutation rate. As has been previously mentioned, there is no strong *a priori* reason to think that ionising radiation should play a pivotal role in supplying beneficial mutations. It is more likely that DNA damage induced by ionising radiation causes a higher ratio of deadly and

detrimental mutations than endogenous cell processes, due to its high propensity to cause spatially co-located DNA damage and double strand breaks.

As dose rises above the background level, measurements of radiation's role in changing mutation rates become possible. If the impact on evolutionary dynamics was to be explored at higher doses than ambient, methods that specifically probe the mutation rate and the beneficial mutation rate, such as fluctuation analysis (Foster, 2006) are more appropriate than fitness experiments. Such methods are far more valuable than LTEEs here as (barring an LTEE with a population size specially chosen to avoid a cap on the selection rate from clonal interference) LTEEs can become saturated with mutations to the point where increasing the mutation rate will not change the evolutionary dynamics. One other possibility to work around this barrier is to use a sexually reproducing population. One advantage of sex is that beneficial mutations are shared between population members, removing the evolutionary bottleneck that occurs in an LTEE based upon a clonal population. *Drosophila melanogaster* the common fruit fly, is an excellent candidate for a study along these lines, as it has already been used in past evolution experiments (Burke et al., 2010).

The second important conclusion from this study is more subtle than the first. A major difficulty of underground biological experiments in the past has been determining the mechanisms by which cells change in response to their radiation environment. These mechanisms are sometimes often epigenetic, and their onset can occur across short and long durations. LTEEs provide a way of probing the evolutionary dynamics of a population. This could be applied to underground experiments to identify how and when adaptations to the underground environment occur. This could not be easily done however if radiation backgrounds strongly alter evolutionary behaviour or selection. At least across the first 500 generations of an LTEE, it seems here that a 7-fold reduction in the radiation background does not change evolutionary behaviour nor selection, so any changes in populations in response to similar changes in their radiation environment likely come from other sources.

4.6 Perspectives for Future Experiments

Small differences in radiation backgrounds seem to affect bacterial life, but this is not what we have seen here. Castillo et al. (2015) have seen changes in protein expression and growth rates in *S. oneidensis* and *D. radiodurans*, and Ruiz-Gonzalez et al. (2016) and Zavilgelsky et al. (1998) have both noticed increased radiation resistance in bacteria found in the chronically elevated radiation environment around Chernobyl. Here, we have shown a situation where reductions in the ionising radiation level do not change the evolutionary behaviour of *E. coli*. Is this because *E. coli* is robust to radiation changes, or because the changes that would

have occurred in the bacteria were not measured? If radiation environments do not affect selection, but only growth rates, this is likely to be unidentifiable in an LTEE. Establishing that *E. coli* does not show strong variation between radiation in terms of its growth rate and evolutionary behaviour would establish it as a strong control candidate for low background biology experiments.

Growth rate measurements of *E. coli* in different radiation environments are thus of significant interest in continuing the research done here. Some preliminary investigations have already been completed, growing bacteria that had spent time adapting to the LPC or LSM in both these environments. Little evidence for a change in growth rate was noticed, however when conducting experiments in the LSM, strains that had already grown in the LSM showed shorter lag times, and similarly strains grown in the LPC showed shorter lag times when grown in the LPC, than the LSM strain. As these changes could be linked to physically transporting cells between environments, it would be of interest to grow cells both inside and outside of lead shielding in the LSM, to note whether this changes cell growth rates. Whether or not this supports or contradicts the behaviours noticed in other cell lines by Planel et al. (1987) and Castillo et al. (2015) is of scientific interest for the field. In the case where growth rate does not change, we can more strongly affirm *E. coli*'s potential as a control in low radiation background studies.

In the case where growth rate does change, the causes of this could be well explored in *E. coli* where the mechanisms behind many cell behaviours are well known. Additionally, if there are traits that we have not explored that do change between radiation environments, studying their onset is important. Evolution experiments can effectively "save" bacterial cultures at regular intervals. This permits the origin of changes to be explored and also sometimes 'replayed' - from a starting point 100 generations before a change occurred, its reproducibility from that point can be ascertained. Seeing how small changes in radiation environment then affect these changes would allow researchers to see whether retardation of the growth rate due to the radiation environment is a binary phenomenon, or occurs gradually over a range of radiation levels. It would be interesting to see if there was one particular radiation level, or range, that living systems in general prefer.

One logical extension of this study is to conduct LTEEs at high radiation backgrounds, trying to force a change in the evolutionary behaviour. This has merit though only in a small sense. The dynamics of LTEEs are such that, working with the experimental method we have chosen here, increases in the radiation level would not increase the selection rate of beneficial mutations. LTEE's would however permit the emergence of mutations which favour radioresistance to be studied, in a high dose regime. Seeing where these adaptations occur in relation to the more canonical LTEE mutations would provide information about

the nature of radiation as an evolutionary pressure, and could complement the work of Harris et al. (2009), who studied the directed evolution of radioresistance in *E. coli*. In particular, they used comparative genome sequencing to identify that mutations that occurred in radioresistant *E. coli* correlated with genes present in *D. radiodurans*. Genome sequencing allows a very fine resolution of changes that occur in a population that may not necessarily manifest themselves in fitness assays. Applying it across our and future evolution experiments could yield information about environment dependent changes that were invisible to this study.

Conclusions and Perspectives

One hundred and something pages after writing: *this thesis considers just one simple question*; it's probably time to write something of an answer. Does ambient ionising radiation have a measurable impact on the pace of evolution? As far as this work is concerned, and the measurements made within it are concerned, no. We could not measure any impact of the natural radiation background upon the evolutionary behaviour of bacteria.

Through Chapter 2, we explored in simulation the physical limits on how much the radiation background can interact with cells, and we compared this to the mutation rate. From this, we showed that even in the event that all interactions between the radiation background and cells cause mutations, mutations caused by endemic processes in bacterial cells will still dominate the mutation rate. We extended this by developing a computational representation of bacterial DNA, and then simulating the relationship between dose and strand break yield for electrons in bacteria in Chapter 3. This allowed us to verify that the small number of energy depositions from ionising radiation in bacterial cells do not cause an unexpectedly large amount of DNA damage. As part of this work, we developed a flexible simulation platform for mechanistic DNA simulations, to be released as part of the Geant4-DNA project.

Simulation can only tell us so much about the real world - experiments must be done to verify our presumptions. This was the subject of Chapter 4 of this thesis. We conducted controlled evolution experiments in *E. coli* in the Modane Underground Laboratory, and Clermont-Ferrand Particle Physics Laboratory in order to see if the radiation background could have a measurable impact on the evolutionary behaviour of bacteria across 500 generations. Evolutionary differences were measured by considering changes in competitive fitness between evolved and ancestral bacterial strains every 100 generations. Despite an 8-fold difference in the radiation background between the two environments, and a 7.3 fold difference in the estimated radiation-induced mutation rate, no significant change was able to be observed between the cell lines grown in the low radiation background provided by the underground laboratory, and the reference radiation background in Clermont-Ferrand.

None of this is to say though that pursuing biological experiments in underground envi-

ronments is without merit. Our quantification of the rate of interactions between the radiation background and cells highlights why further research is needed in this field. Namely, the impact of ionising radiation from the background is so small that cells should be hardly sensitive to it. Even where cells are sensitive they shouldn't respond quickly to the background. Nevertheless, cell populations have been observed, repeatedly, to rapidly respond to reductions in the radiation background. This hints at the existence of a rich regulatory network that determines the radiation response of cells, that could implicate cell signalling and epigenetics. Understanding and explaining these processes is a significant challenge for the field, and requires careful, repeated experiments to assess under what conditions the radiation background can impact cellular growth, gene expression and antioxidant regulation.

Cell communication and epigenetic regulation are increasingly seen as important in a variety of low dose radiation effects, many of which have implications for human health. As an example of cell communication, the bystander effect provides evidence that cells communicate information about their radiation environment with each other. Exploring how this communication manifests itself at low doses may explain observations of the radiation background behaving hormetically, apparently accelerating the growth of unicellular organisms (Castillo et al., 2015; Planel et al., 1987). Cells grown underground have shown a memory of their radiation environment (Fratini et al., 2015), possibly an epigenetic response to the radiation background. In addition to providing a mechanism for cellular responses to radiation environments, epigenetics deserves exploration as at low doses it can be implicated in the formation of cancers, where cells inherit a memory of radiation damage that doesn't manifest itself until many generations after the initial radiation event.

The simulation tools developed in Chapter 3 represent a rich terrain for future simulation development. Up to now, mechanistic simulations of radiation induced DNA damage have remained difficult for most researchers to access, requiring both programming expertise, and access to a closed-source simulation toolkit. By creating a Geant4 application controllable entirely through user facing macro commands, both of these limitations are lifted. A flexible interface for DNA geometries has been created, that is generic enough to accept a variety of DNA confirmations, while remaining fast enough to simulate early stage DNA damage in a reasonable time period on high end consumer-grade computers. Advancing such simulations requires simulators and experimentalists to work together to determine what parameters are most relevant in such simulations, and then to develop ways of bridging the gap between very early stage cellular damage calculations, and biological cellular outcomes, particularly beneficial mutations, harmful mutations and cell death. The next step in this direction is the incorporation of DNA repair modelling, based upon break complexity.

Acknowledgements

In some ways, I've waited three years to write this last section of my thesis. This manuscript concludes the most challenging, at times frustrating, and (though I may not always have realised it) rewarding part of my life so far. It's only fair that I thank the many people who've been a part of a journey that took an Australian, who'd really only ever lived in one city, to a new country, with a new language and new customs.

An adventure that started in Montluçon, I discovered *La France profonde* in almost the clumsiest way possible. My thanks go especially to David Sarramia, who, in addition to directing this thesis, served as my guide, and through many afternoons writing French on the lunch room white board introduced me to the French language, so I wouldn't be so carelessly adrift. The next stop was the *Laboratoire Souterrain de Modane*, nestled in the French Alps. For the months that I was travelling regularly to the LSM and back, the team at the underground lab made me feel incredibly welcome, and I owe all of them my gratitude.

Sometime around here the thesis took me to Clermont-Ferrand where we began our biological experiments in earnest, and I'm indebted in particular to Pierre Marin, who was the best biology instructor anyone could ask for. To all of my friends and colleagues at the LPC, and especially our group in the PPSE, thanks once again for tolerating my French and sharing your knowledge. A special thank you to Arianna and Sonia for bravely facing the CROUS every day, to see what culinary wonders awaited us in the university restaurant.

Around the start of 2016, yet another *déménagement* took place taking me to Bordeaux, in order to build my cell level simulations at the *Centre d'Etudes Nucléaires de Bordeaux-Gradignan*. Here, I have to thank Sébastien Incerti, who co-supervised this last part of my thesis, and whose friendship, support and encouragement have been invaluable. I'll definitely miss our busy little office, which I also had the honour of sharing with Dousatsu Sakata - Dousatsu, my education in Japanese cuisine at your hand has been second to none, and for this you have my eternal gratitude. To the rest of the Iribio group, and the wider staff of the CENBG, thank you once again for your warm welcome, and for making me feel a part of the lab.

I am indebted to many people beyond those listed above chronologically. To Vincent Breton, you have enriched this thesis so much with your direction and your keen grasp of

the ‘right question’. I don’t think either of us could have seen the path this thesis would take at the outset, and a large part of its success (I hope) is due to your encouragement to be diligent, thorough and rigourous. David Biron, it has been a pleasure to exploring the world of biology with you. Thanks must go to Jeremy Brown, for both his technical guidance and his friendship over the last three years - I wouldn’t have even embarked on this adventure if it weren’t for your encouragement Jeremy.

To all the friends I’ve made along the way. To Laurin Williams, thanks for always encouraging me, especially when things were hardest. To those who’ve helped make Bordeaux feel like a second home, especially Rob and Craig Bradley-West, and Soizic Brohan - in large part thanks to you *la rédaction n’a pas été trop de galère*. To Mathieu Karamitros, thank you for dealing with all my bug reports, and for your support and wisdom throughout my PhD. To my many friends from the Monash University days, where my academic journey started, thanks go to you, and to all the lecturers who prepared me for research, especially Duncan Galloway, whose guidance during my honours prepared me well for my thesis.

To my family back home. Mum, thanks for letting me escape to Europe (amongst everything else). Dad, thanks for always encouraging me. Rachael, Naomi, Aaron, you’re the best siblings anyone could ask for. Though I’ve been away, seeing you guys all grow these last three years has been wonderful. I love you all.

And lastly, thanks to you, dear reader, for having made it this far. Whether I’ve been able to mention your name above or not, this last sentence is especially for you.

References

Agostinelli, S., Allison, J., Amako, K., et al. 2003, GEANT4 - A simulation toolkit, Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 506, 250

Allison, J., Amako, K., Apostolakis, J., et al. 2006, Geant4 developments and applications, IEEE Transactions on Nuclear Science, 53, 270

—. 2016, Recent developments in Geant4, Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 835, 186

Arnott, S., & Hukins, D. W. L. 1972, Optimised parameters for A-DNA and B-DNA, Biochemical and Biophysical Research Communications, 47, 1504

Atri, D. 2016, On the possibility of galactic cosmic ray-induced radiolysis-powered life in subsurface environments in the Universe, Journal of The Royal Society Interface, 13

Aydogan, B., Marshall, D. T., Swarts, S. G., et al. 2002, Site-specific OH attack to the sugar moiety of DNA: a comparison of experimental data and computational simulation, Radiation Research, 157, 38

Azzam, E. I., De Toledo, S. M., & Little, J. B. 2003, Oxidative metabolism, gap junctions and the ionizing radiation-induced bystander effect, Oncogene, 22, 7050

Balasubramanian, B., Pogozelski, W. K., & Tullius, T. D. 1998, DNA strand breaking by the hydroxyl radical is governed by the accessible surface areas of the hydrogen atoms of the DNA backbone, Proceedings of the National Academy of Sciences of the United States of America, 95, 9738

Baldwin, W. W., Myer, R., Kung, T., Anderson, E., & Koch, A. L. 1995, Growth and buoyant density of Escherichia coli at very low osmolarities, Journal of Bacteriology, 177, 235

Barrick, J. E., & Lenski, R. E. 2013, Genome dynamics during experimental evolution., Nature reviews Genetics, 14, 827

Barrick, J. E., Yu, D. S., Yoon, S. H., et al. 2009, Genome evolution and adaptation in a long-term experiment with Escherichia coli, Nature, 461, 1243

Barrios, R., Skurski, P., & Simons, J. 2002, Mechanism for Damage to DNA by Low-Energy Electrons, Journal of Physical Chemistry B, 106, 7991

Battista, J. R. 1997, AGAINST ALL ODDS:The Survival Strategies of Deinococcus radiodurans, Annual Review of Microbiology, 51, 203

Bé, M., Chisté, V., Dulieu, C., et al. 2010, Table of Radionuclides, Vol. 5, Table of Radionuclides, 256

Bé, M.-M., Duchemin, B., Lamé, J., et al. 1999, Table de Radionucléides, Vol. 1, Table de Radionucléides (DAMRI/LPRI BP 52, F-91193 Gif-sur-Yvette Cedex, France: Commissariat à l'Énergie Atomique)

Beltrán-Pardo, E., Jönsson, K. I., Wojcik, A., et al. 2013, Effects of Ionizing Radiation on Embryos of the Tardigrade Milnesium cf. tardigradum at Different Stages of Development, PLoS ONE, 8, e72098

Berger, M. J., Coursey, J., Zucker, M., & Chang, J. 1998, Stopping-power and range tables for electrons, protons, and helium ions, Stopping-power and range tables for electrons, protons, and helium ions (NIST Physics Laboratory)

Bernal, M., Bordage, M., Brown, J., et al. 2015, Track structure modeling in liquid water: A review of the Geant4-DNA very low energy extension of the Geant4 Monte Carlo simulation toolkit, Physica Medica, 1

Bernal, M. A., Sikansi, D., Cavalcante, F., et al. 2013, An atomistic geometrical model of the B-DNA configuration for DNA-radiation interaction simulations, Computer Physics Communications, 184, 2840

Bertucci, A., Pocock, R. D. J., Randers-Pehrson, G., & Brenner, D. J. 2009, Microbeam Irradiation of the C. elegans Nematode Antonella, Journal of Radiation Research, 50, A49

Bettini, A. 2014, New underground laboratories: Europe, Asia and the Americas, Physics of the Dark Universe, 4, 36

Bondi, A. 1964, van der Waals Volumes and Radii, The Journal of Physical Chemistry, 68, 441

Bonura, T., Town, C. D., Smith, K. C., & Kaplan, H. S. 1975, The influence of Oxygen on the Yield of DNA Double-Strand Breaks in X-Irradiated Escherischia coli K-12, Radiation Research, 63, 567

Bordage, M., Bordes, J., Edel, S., et al. 2016, Implementation of new physics models for low energy electrons in liquid water in Geant4-DNA, Physica Medica, 32, 1833

Boubriak, I. I., Grodzinsky, D. M., Polischuk, V. P., et al. 2008, Adaptation and Impairment of DNA Repair Function in Pollen of Betula verrucosa and Seeds of Oenothera biennis from Differently Radionuclide-contaminated Sites of Chernobyl, Annals of Botany, 101, 267

Boudaiffa, B., Cloutier, P., Hunting, D., Huels, M. A., & Sanche, L. 2000, Resonant Formation of DNA Strand Breaks by Low-Energy (3 to 20 eV) Electrons, Science, 287, 1658

Bouville, A., & Lowder, W. M. 1988, Human Population Exposure to Cosmic Radiation, Radiation Protection Dosimetry, 24, 293

Brown, J. M. C., & Currell, F. J. 2017, A local effect model-based interpolation framework for experimental nanoparticle radiosensitisation data, Cancer Nanotechnology, 8, 1

Brown, J. M. C., Dimmock, M. R., Gillam, J. E., & Paganin, D. M. 2014, A low energy bound atomic electron Compton scattering model for Geant4, Nuclear Instruments and Methods in Physics Research, Section B: Beam Interactions with Materials and Atoms, doi:10.1016/j.nimb.2014.07.042

Burke, M. K., Dunham, J. P., Shahrestani, P., et al. 2010, Genome-wide analysis of a long-term evolution experiment with Drosophila, Nature, 467, 587

Butterworth, K. T., Wyer, J. A., Brennan-fournet, M., et al. 2008, Variation of Strand Break Yield for Plasmid DNA Irradiated with High- Z Metal Nanoparticles, Radiation Research, 387, 381

Buxton, G. V., Greenstock, C. L., Helman, W. P., & Ross, A. B. 1988, Critical Review of rate constants for reactions of hydrated electrons, hydrogen atoms and hydroxyl radicals (OH/O) in Aqueous Solution, Journal of Physical and Chemical Reference Data, 17, 513

Caamano, D., Tracol, R., Guillotin, L., Davezac, H., & Loyen, J. 2011, La qualité radiologique de l ' eau du robinet en France 2008-2009, Tech. rep., Ministère de la Santé et des Sports

Calabrese, E. J. 2013, Hormetic mechanisms, Critical Reviews in Toxicology, 43, 580

Calabrese, E. J., & Baldwin, L. A. 2003, Toxicology rethinks its central belief, Nature, 421, 691

Cameron, J. A., & Singh, B. 2004, Nuclear Data Sheets for A=40, Nuclear Data Sheets, 102, 293

Carbone, M. C., Pinto, M., Antonelli, F., et al. 2009, The Cosmic Silence experiment: on the putative adaptive role of environmental ionizing radiation, Radiation and Environmental Biophysics, 48, 189

Carlton, B. C., & Brown, B. J. 1981, in Manual of methods for general bacteriology, ed. P. Gerhardt (Washington D.C.: American Society for Microbiology), 222–242

Castillo, H., Schoderbek, D., Dulal, S., et al. 2015, Stress induction in the bacteria Shewanella oneidensis and Deinococcus radiodurans in response to below-background ionizing radiation, International Journal of Radiation Biology, 3002, 1

—. 2016, Response to Dr Katz, International Journal of Radiation Biology, 92, 169

Charlton, D. E., & Humm, J. L. 1988, A Method of Calculating Initial DNA Strand Breakage Following the Decay of Incorporated 125I, International Journal of Radiation Biology, 53, 353

Chazal, V., Brissot, R., Cavaignac, J., et al. 1998, Neutron background measurements in the Underground Laboratory of Modane, Astroparticle Physics, 9, 163

Chirkin, D. 2004, Fluxes of Atmospheric Leptons at 600 GeV - 60 TeV, arXiv:hep

Chivian, D., Brodie, E. L., Alm, E. J., et al. 2008, Environmental Genomics Reveals a Single-Species Ecosystem Deep Within Earth, Science, 322

Cohen, B. 2012, in Radiation Dose from Multidetector CT, ed. D. Tack, M. K. Kalra, & P. A. Gevenois (Berlin Hiedelberg: Springer-Verlag), 61–79

Daly, M. J. 2012, Death by protein damage in irradiated cells, DNA Repair, 11, 12

Daly, M. J., Gaidamakova, E. K., Matrosova, V. Y., et al. 2004, Accumulation of Mn(II) in Deinococcus radiodurans facilitates gamma-radiation resistance, Science, 306, 1025

Dawkins, R. 2006, The Selfish Gene, The Selfish Gene (Oxford University Press)

de Lara, C. M., Hill, M. A., Jenner, T. J., Papworth, D., & Neill, P. O. 2001, Dependence of the Yield of DNA Double-Strand Breaks in Chinese Hamster V79-4 Cells on the Photon Energy of Ultrasoft X Rays, Radiation Research, 155, 440

de Visser, J. A. G. M., Zeyl, C. W., Gerrish, P. J., et al. 1999, Diminishing Returns from Mutation Supply Rate in Asexual Populations, Science, 283, 404

Delage, E., Pham, Q., Karamitros, M., et al. 2015, PDB4DNA: Implementation of DNA geometry from the Protein Data Bank (PDB) description for Geant4-DNA Monte-Carlo simulations, Computer Physics Communications, 192, 282

Deryabina, T., Kuchmel, S., Nagorskaya, L., et al. 2015, Long-term census data reveal abundant wildlife populations at Chernobyl, Current Biology, 25, R824

Dillon, S. C., & Dorman, C. J. 2010, Bacterial nucleoid-associated proteins, nucleoid structure and gene expression, Nature reviews Microbiology, 8, 185

Dos Santos, M., Villagrasa, C., Clairand, I., & Incerti, S. 2013, Influence of the DNA density on the number of clustered damages created by protons of different energies, Nuclear Instruments and Methods in Physics Research, Section B: Beam Interactions with Materials and Atoms, 298, 47

Douglass, M., Bezak, E., & Penfold, S. 2012, Development of a randomized 3D cell model for Monte Carlo microdosimetry simulations, Medical Physics, 39, 3509

Drake, J. W. 1991, A constant rate of spontaneous mutation in DNA-based microbes, Proceedings of the National Academy of Sciences of the United States of America, 88, 7160

Drake, J. W., Charlesworth, B., Charlesworth, D., & Crow, J. F. 1998, Rates of spontaneous mutation, Genetics, 148, 1667

Dubrova, Y. E. 2003, Radiation-induced transgenerational instability, Oncogene, 22, 7087

Dyson, F. 2004, A meeting with Enrico Fermi How one intuitive physicist rescued a team from fruitless research, Nature, 427, 8540

El Albani, A., Bengtson, S., Canfield, D. E., et al. 2010, Large colonial organisms with coordinated growth in oxygenated environments 2.1 Gyr ago, Nature, 466, 100

—. 2014, The 2.1 Ga old Francevillian biota: Biogenicity, taphonomy and biodiversity, PLoS ONE, 9, doi:10.1371/journal.pone.0099438

El Naqa, I., Pater, P., & Seuntjens, J. 2012, Monte Carlo role in radiobiological modelling of radiotherapy outcomes, Physics in Medicine and Biology, 57, R75

Eldredge, N., & Gould, S. J. 1972, in Models in Paleobiology, ed. T. J. M. Schopf (San Francisco: Freeman, Cooper and Company), 82–115

Elena, S. F., Cooper, V. S., & Lenski, R. E. 1996, Punctuated evolution caused by selection of rare beneficial mutations, Science, 272, 1802

Elsässer, T., Krämer, M., & Scholz, M. 2008, Accuracy of the Local Effect Model for the Prediction of Biologic Effects of Carbon Ion Beams In Vitro and In Vivo, International Journal of Radiation Oncology Biology Physics, 71, 866

Elsässer, T., & Scholz, M. 2007, Cluster effects within the local effect model., Radiation Research, 167, 319

Emfietzoglou, D., Cucinotta, F. A., & Nikjoo, H. 2005, A complete dielectric response model for liquid water: a solution of the Bethe ridge problem, Radiation Research, 164, 202

Enger, S. A., Landry, G., D'Amours, M., et al. 2012, Layered mass geometry: a novel technique to overlay seeds and applicators onto patient geometry in Geant4 brachytherapy simulations, Physics in Medicine and Biology, 57, 6269

Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. 1996, in Proceedings of 2nd International Conference on Knowledge Discovery and Data Mining (AAAI Press), 226–231

Feinendegen, L., Pollycove, M., & Sondhaus, C. 2004, Responses To Low Doses of Ionizing Radiation in Biological Systems, Nonlinearity in Biology, Toxicology, and Medicine, 2, 143

Feinendegen, L. E. 2005, Evidence for beneficial low level radiation effects and radiation hormesis, British Journal of Radiology, 78, 3

Folkard, M., Prise, K., Vojnovic, B., et al. 1993, Measurement of DNA Damage by Electrons with Energies between 25 and 4000 EV, International Journal of Radiation Biology, 64, 651

Foster, P. L. 2006, Methods for determining spontaneous mutation rates, Methods In enzymology, 409, 195

Francis, Z., Villagrasa, C., & Clairand, I. 2011, Simulation of DNA damage clustering after proton irradiation using an adapted DBSCAN algorithm, Computer Methods and Programs in Biomedicine, 101, 265

Françon, J. 1997, The algorithmic beauty of plants, Plant Science, 122, 109

Fratini, E., Carbone, C., Capece, D., et al. 2015, Low-radiation environment affects the development of protection mechanisms in V79 cells, Radiation and Environmental Biophysics, 54, 183

Friedberg, E. C. 2003, DNA damage and repair, Nature, 421, 436

Friedburg, E. C., Walker, G. C., Siede, W., & Wood, R. D. 2005, DNA Repair and Mutagenesis, DNA Repair and Mutagenesis (American Society for Microbiology Press), 1118

Friedland, W., Dingfelder, M., Kundrát, P., & Jacob, P. 2011, Track structures, DNA targets and radiation effects in the biophysical Monte Carlo simulation code PARTRAC, Mutation Research, 711, 28

Friedland, W., Jacob, P., Bernhardt, P., Paretzke, H. G., & Dingfelder, M. 2003, Simulation of DNA damage after proton irradiation, Radiation Research, 159, 410

Fritsche, M., Li, S., Heermann, D. W., & Wiggins, P. a. 2012, A model for Escherichia coli chromosome packaging supports transcription factor-induced DNA domain formation, Nucleic Acids Research, 40, 972

Galván, I., Bonisoli-Alquati, A., Jenkinson, S., et al. 2014, Chronic exposure to low-dose radiation at Chernobyl favours adaptation to oxidative stress in birds, Functional Ecology, doi:10.1111/1365-2435.12283

Gauthier-Lafaye, F., Holliger, P., & Blanc, P.-L. 1996, Natural fission reactors in the Franceville basin, Gabon: A review of the conditions and results of a "critical event" in a geologic system, Geochimica et Cosmochimica Acta, 60, 4831

Geant4 Collaboration. 2015, Geant4 User's Guide for Application Developers, version Geant4 10.2

Geissmann, Q., Costa, C., Yang, S., et al. 2013, OpenCFU, a New Free and Open-Source Software to Count Cell Colonies and Other Circular Objects, PLoS ONE, 8, e54072

Gerrish, P., & Lenski, R. 1998, The fate of competing beneficial mutations in an asexual population, Genetica, 102-103, 127

Gevertz, D., Friedman, A. M., Katz, J. J., & Kubitschek, H. E. 1985, Biological effects of background radiation: mutagenicity of 40K, Proceedings of the National Academy of Sciences of the United States of America, 82, 8602

Ghiassi-nejad, M., Mortazavi, S. M. J., Cameron, J. R., Niroomand-rad, A., & Karam, P. a. 2002, Very high background radiation areas of Ramsar, Iran: preliminary biological studies, Health physics, 82, 87

Gibson, K. D., & Scherage, H. A. 1987, Volume of the Intersection of Three Spheres of Unequal Size. A Simplified Formula, Journal of Physical Chemistry, 91, 4121

Godin, M., Bryan, A. K., Burg, T. P., Babcock, K., & Manalis, S. R. 2007, Measuring the mass, density, and size of particles and cells using a suspended microchannel resonator, Applied Physics Letters, 91, 3

Gordon, M., Goldhagen, P., Rodbell, K., et al. 2004, Measurement of the flux and energy spectrum of cosmic-ray induced neutrons on the ground, IEEE Transactions on Nuclear Science, 51, 3427

Gould, S. J., & Eldredge, N. 1977, Punctuated Equilibria : The Tempo and Mode of Evolution Reconsidered, Paleobiology, 3, 115

Green, N. J. B., Pilling, M. J., Pimblott, S. M., & Clifford, P. 1990, Stochastic modeling of fast kinetics in a radiation track, The Journal of Physical Chemistry, 94, 251

Groom, D. E., Mokhov, N. V., & Striganov, S. I. 2001, Muon Stopping Power and Range Tables 10 MeV - 100 TeV, Atomic Data and Nuclear Data Tables, 78, 183

Harris, D. R., Pollock, S. V., Wood, E. a., et al. 2009, Directed evolution of ionizing radiation resistance in Escherichia coli, Journal of Bacteriology, 191, 5240

Hirayama, R., Ito, A., Tomita, M., et al. 2009, Contributions of Direct and Indirect Actions in Cell Killing by High-LET Radiations, Radiation Research, 171, 212

Ilnytskyy, Y., & Kovalchuk, O. 2011, Non-targeted radiation effects-an epigenetic connection, Mutation Research, 714, 113

Incerti, S., Ivanchenko, A., Karamitros, M., et al. 2010, Comparison of GEANT4 very low energy cross section models with experimental data in water, Medical physics, 37, 4692

Itseez. 2015, Open source computer vision library, https://github.com/itseez/opencv

Jeong, H., Barbe, V., Lee, C. H., et al. 2009, Genome Sequences of Escherichia coli B strains REL606 and BL21(DE3), Journal of Molecular Biology, 394, 644

Joseph, S. B., & Hall, D. W. 2004, Spontaneous mutations in diploid Saccharomyces cerevisiae: More beneficial than expected, Genetics, 168, 1817

Kammeyer, C. W., & Whitman, D. R. 1972, Quantum Mechanical Calculation of Molecular Radii. I. Hydrides of Elements of Periodic Groups IV through VII, The Journal of Chemical Physics, 56, 4419

Karamitros, M. 2013, Extension de l'outil Monte Carlo généraliste Geant4 pour la simulation de la radiolyse de l'eau dans le cadre du projet Geant4-DNA, PhD thesis, Université de Bordeaux

Karamitros, M., Mantero, A., Incerti, S., et al. 2011, Modeling Radiation Chemistry in the Geant4 Toolkit, Progress in Nuclear Science and Technology, 2, 503

Karamitros, M., Luan, S., Bernal, M. A., et al. 2014, Diffusion-controlled reactions modeling in Geant4-DNA, Journal of Computational Physics, 274, 841

Katz, J. I. 2016, Comment on Castillo et al. (2015), International Journal of Radiation Biology, 3002, 1

Kawanishi, M., Okuyama, K., Shiraishi, K., et al. 2012, Growth Retardation of Paramecium and Mouse Cells by Shielding Them from Background Radiation, Journal of Radiation Research, 53, 404

Kreipl, M. S., Friedland, W., & Paretzke, H. G. 2009, Interaction of ion tracks in spatial and temporal proximity, Radiation and Environmental Biophysics, 48, 349

Krisko, A., & Radman, M. 2010, Protein damage and death by radiation in Escherichia coli and Deinococcus radiodurans, Proceedings of the National Academy of Sciences of the United States of America, 107, 14373

Kudryasheva, N. S., & Rozhko, T. V. 2015, Effect of low-dose ionizing radiation on luminous marine bacteria: Radiation hormesis and toxicity, Journal of Environmental Radioactivity, 142, 68

Kulp, A., & Kuehn, M. J. 2012, Biological Functions and Biogenesis of Secreted Bacterial Outer Membrane Vesicles, Annual Review of Microbiology, 163

Kyriakou, I., Incerti, S., & Francis, Z. 2015, Technical Note: Improvements in geant4 energy-loss model and the effect on low-energy electron transport in liquid water, Medical Physics, 42, 3870

Lampe, N. 2016, Multiscale geant4 simulation application, Online Repository, doi:10.5281/zenodo.163962

Lampe, N., Biron, D. G., Brown, J. M. C., et al. 2016, Simulating the impact of the natural radiation background on bacterial systems: Implications for very low radiation biological experiments, PLoS ONE, 11, 1

Lechtman, E., Mashouf, S., Chattopadhyay, N., et al. 2013, A Monte Carlo-based model of gold nanoparticle radiosensitization accounting for increased radiobiological effectiveness, Physics in Medicine and Biology, 58, 3075

Lenski, R. E., Mongold, J. a., Sniegowski, P. D., et al. 1998, Evolution of competitive fitness in experimental populations of E. coli: What makes one genotype a better competitor than another?, Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology, 73, 35

Lenski, R. E., Rose, M. R., Simpson, S. C., & Tadler, S. C. 1991, Long-Term Experimental Evolution in Escherichia coli. I. Adaptation and Divergence During 2,000 Generations, The American Naturalist, 138, 1315

Lenski, R. E., & Travisano, M. 1994, Dynamics of adaptation and diversification: a 10,000-generation experiment with bacterial populations, Proceedings of the National Academy of Sciences of the United States of America, 91, 6808

Liang, Y., Yang, G., Liu, F., & Wang, Y. 2016, Monte Carlo simulation of ionizing radiation induced DNA strand breaks utilizing coarse grained high-order chromatin structures, Physics in Medicine and Biology, 61, 445

Lieberman-Aiden, E., van Berkum, N. L., Williams, L., et al. 2009, Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome, Science

Liljequist, D., & Nikjoo, H. 2014, On the validity of trajectory methods for calculating the transport of very low energy (<1keV) electrons in liquids and amorphous media, Radiation Physics and Chemistry, 99, 45

Limoli, C. L., Corcoran, J. J., Milligan, J. R., Ward, J. F., & Morgan, W. F. 1999, Critical target and dose and dose-rate responses for the induction of chromosomal instability by ionizing radiation, Radiation Research, 151, 677

Little, M. P., Wakeford, R., Tawn, E. J., Bouffler, S. D., & Berrington de Gonzalez, A. 2009, Risks associated with low doses and low dose rates of ionizing radiation: why linearity may be (almost) the best we can do, Radiology, 251, 6

Ljungman, M., Nyberg, S., Nygren, J., Eriksson, M., & Ahnström, G. 1991, DNA-Bound Proteins Contribute Much More than Soluble Intracellular Compounds to the Intrinsic Protection against Radiation-Induced DNA Strand Breaks in Human Cells, Radiation Research, 127, 171

Lynch, M. 2010, Evolution of the mutation rate, Trends in Genetics, 26, 345

Malczewski, D., Kisiel, J., & Dorda, J. 2012, Gamma background measurements in the Laboratoire Souterrain de Modane, Journal of Radioanalytical and Nuclear Chemistry, 292, 751

Martell, E. A. 1992, Radionuclide-induced evolution of DNA and the origin of life, Journal of Molecular Evolution, 35, 346

Martin, R. F., & Haseltine, W. A. 1981, Range of radiochemical damage to DNA with decay of iodine-125, Science, 213, 896

Maynard Smith, J., & Price, G. R. 1973, The logic of animal conflict, Nature, 246, 15

McMahon, S. J., Schuemann, J., Paganetti, H., et al. 2016, Mechanistic Modelling of DNA Repair and Cellular Survival Following Radiation-Induced DNA Damage, Scientific Reports, 6, 33290

McNamara, A. L., Guatelli, S., Prokopovich, D. a., Reinhard, M. I., & Rosenfeld, A. B. 2012, A comparison of X-ray and proton beam low energy secondary electron track structures using the low energy models of Geant4, International Journal of Radiation Biology, 88, 164

Merrifield, M., & Kovalchuk, O. 2013, Epigenetics in radiation biology: A new research frontier, Frontiers in Genetics, 4, 1

Meylan, S. 2016, Development of a multi-scale simulation tool for early radio-induced damage assessment in cells exposed to light ions irradiations (proton and alpha), PhD thesis, Université de Bordeaux

Meylan, S., Vimont, U., Incerti, S., Clairand, I., & Villagrasa, C. 2016, Geant4-DNA simulations using complex DNA geometries generated by the DnaFabric tool, Computer Physics Communications, 204, 159

Milligan, J. R., Aguilera, J. A., & Ward, J. F. 1993, Variation of single-strand break yield with scavenger concentration for plasmid DNA irradiated in aqueous solution, Radiation Research, 133, 151

Møller, A. P., Mousseau, T. A., Milinevsky, G., et al. 2005, Condition, reproduction and survival of barn swallows from Chernobyl, Journal of Animal Ecology, 74, 1102

Moore, F. D., & Sastry, K. S. 1982, Intracellular potassium: 40K as a primordial gene irradiator, Proceedings of the National Academy of Sciences of the United States of America, 79, 3556

Morgan, W. F. 2003a, Non-targeted and delayed effects of exposure to ionizing radiation: I. Radiation-induced genomic instability and bystander effects in vitro, Radiation Research, 159, 567

—. 2003b, Non-targeted and delayed effects of exposure to ionizing radiation: II. Radiation-induced genomic instability and bystander effects in vivo, Radiation Research, 159, 581

Mothersill, C., & Seymour, C. B. 2004, Radiation-induced bystander effects - implications for cancer., Nature reviews Cancer, 4, 158

Mozumder, A., & Magee, J. L. 1966, Model of Tracks of Ionizing Radiation Reaction Mechanisms, Radiation Research, 28, 203

Nakano, M., Tateishi-Karimata, H., Tanaka, S., et al. 2016, Local thermodynamics of the water molecules around single- and double-stranded DNA studied by grid inhomogeneous solvation theory, Chemical Physics Letters, 660, 250

Neiss, V. 2016, personal communication

Nikjoo, H., Emfietzoglou, D., Liamsuwan, T., et al. 2016, Radiation track, DNA damage and response—a review, Reports on Progress in Physics, 79, 116601

Nikjoo, H., O'Neill, O., Goodhead, T., & Terrissol, M. 1997, Computational modelling of low-energy electron-induced DNA damage by early physical and chemical events, International Journal of Radiation Biology, 71, 467

Nikjoo, H., Uehara, S., Emfietzoglou, D., & Cucinotta, F. 2006, Track-structure codes in radiation research, Radiation Measurements, 41, 1052

Niklas, K. J. 1994, Morphological evolution through complex domains of fitness, Proceedings of the National Academy of Sciences of the United States of America, 91, 6772

Nowak, M. a., & Sigmund, K. 2004, Evolutionary dynamics of biological games, Science, 303, 793

Olive, K., Agashe, K., Amsler, C., et al. 2014, Review of Particle Physics, Chinese Physics C, 38, 090001

Orlando, T. M., Oh, D., Chen, Y., & Aleksandrov, A. B. 2008, Low-energy electron diffraction and induced damage in hydrated DNA, Journal of Chemical Physics, 128, 1

Perfeito, L., Fernandes, L., Mota, C., & Gordo, I. 2007, Adaptive mutations in bacteria: high rate and small effects, Science, 317, 813

Piquemal, F. 2012, Modane underground laboratory: Status and project, The European Physical Journal Plus, 127, 110

Planel, H., Soleilhavoup, J. P., Tixador, R., et al. 1987, Influence on Cell Proliferation of Background Radiation Or Exposure to Very Low, Chronic Gamma Radiation., Health physics, 52, 571

Prise, K., & Schettino, G. 2011, Microbeams in Radiation Biology : Review and Critical Comparison, Radiation Protection Dosimetry, 143, 335

Prise, K. M., Pinto, M., Newman, H. C., & Michael, B. D. 2001, A Review of Studies of Ionizing Radiation-Induced Double-Strand Break Clustering, Radiation Research, 156, 572

Radicella, J. P., Clark, E. A., & Fox, M. S. 1988, Some mismatch repair activities in Escherichia coli, Proceedings of the National Academy of Sciences of the United States of America, 85, 9674

Rohde, R. A., & Muller, R. A. 2005, Cycles in fossil diversity, Nature, 434, 208

Roots, R., & Okada, S. 1975, Estimation of Life Times and Diffusion Distances of Radicals DNA Strand Breaks Involved in X-Ray-Induced or Killing of Mammalian Cells ', Radiation Research, 64, 306

Ruiz-Gonzalez, M. X., Czirjak, G. A., Genevaux, P., et al. 2016, Resistance of Feather-Associated Bacteria to Intermediate Levels of Ionizing Radiation near Chernobyl, Scientific reports, 6, 22969

Samuel, A. H., & Magee, J. L. 1953, Theory of Radiation Chemistry. II. Track Effects in Radiolysis of Water, The Journal of Chemical Physics, 21, 1080

Satta, L., Augusti-Tocco, G., Ceccarelli, R., et al. 1995, Low environmental radiation background impairs biological defence of the yeast Saccharomyces cerevisiae to chemical radiomimetic agents, Mutation Research, 347, 129

Satta, L., Antonelli, F., Belli, M., et al. 2002, Influence of a low background radiation environment on biochemical and biological responses in V79 cells, Radiation and environmental biophysics, 41, 217

Savvidis, I., Giomataris, I., Bougamont, E., et al. 2010, Underground low flux neutron background measurements in LSM using a large volume (1m3) spherical proportional counter, Journal of Physics: Conference Series, 203, 012030

Schmidt, B., Armengaud, E., Augier, C., et al. 2013, Muon-induced background in the EDELWEISS dark matter search, Astroparticle Physics, 44, 28

Scholes, G., Willson, R. L., & Ebert, M. 1969, Pulse radiolysis of aqueous solutions of deoxyribonucleotides and of DNA: reaction with hydroxy-radicals, Journal of the Chemical Society D: Chemical Communications, 17

Scholz, M., Kellerer, A. M., Kraft-Weyrather, W., & Kraft, G. 1997, Computation of cell survival in heavy ion beams for therapy. The model and its approximation, Radiation and environmental biophysics, 36, 59

Schwarz, H. A. 1969, Applications of the spur diffusion model to the radiation chemistry of aqueous solutions, The Journal of Physical Chemistry, 73, 1928

Shaviv, N. J. 2002, Cosmic ray diffusion from the galactic spiral arms, iron meteorites, and a possible climatic connection, Physical Review Letters, 89, 051102

Shiina, T., Watanabe, R., Shiraishi, I., et al. 2013, Induction of DNA damage, including abasic sites, in plasmid DNA by carbon ion and X-ray irradiation, Radiation and Environmental Biophysics, 52, 99

Smith, G. B., Grof, Y., Navarrette, A., & Guilmette, R. A. 2011, Exploring biological effects of low background radiation from the other side of the background, Health physics, 100, 263

Sniegowski, P. D., & Gerrish, P. J. 2010, Beneficial mutations and the dynamics of adaptation in asexual populations, Philosophical Transactions of the Royal Society B: Biological Sciences, 365, 1255

Sniegowski, P. D., Gerrish, P. J., Johnson, T., & Shaver, A. 2000, The evolution of mutation rates: separating causes from consequences, BioEssays, 22, 1057

Sniegowski, P. D., Gerrish, P. J., & Lenski, R. E. 1997, Evolution of high mutation rates in experimental populations of E . coli, Nature, 387, 703

Souici, M., Khalil, T. T., Boulanouar, O., et al. 2016, DNA strand break dependence on Tris and arginine scavenger concentrations under ultra-soft X-ray irradiation: the contribution of secondary arginine radicals, Radiation and Environmental Biophysics, 55, 215

Steinhauser, G. 2015, Ionizing radiation – an evolutionary threat?, Hypothesis, 13, 1

Sy, D., Savoye, C., Begusova, M., et al. 1997, Sequence-dependent variations of DNA structure modulate radiation-induced strand breakage, International Journal of Radiation Biology, 72, 147

Tenaillon, O., Barrick, J. E., Ribeck, N., et al. 2016, Tempo and mode of genome evolution in a 50,000-generation experiment, Nature, 536, 165

Terrissol, M., & Beaudré, A. 1990, Simulation of Space and Time Evolution of Radiolytic Species Induced by Electrons in Water, Radiation Protection Dosimetry, 31, 175

Tilley, D. R., Cheves, C. M., Kelley, J. H., Raman, S., & Weller, H. R. 2012, Energy Levels of Light Nuclei, Revised Manuscript, 40441

Travisano, M., & Lenski, R. E. 1996, Long-term experimental evolution in Escherichia coli. IV. Targets of selection and the specificity of adaptation, Genetics, 143, 15

Tubiana, M., Feinendegen, L. E., Yang, C., & Kaminski, J. M. 2009, The Linear No-Threshold Relationship Is Inconsistent with Radiation Biologic and Experimental Data, Radiology, 251, doi:10.1148/radiol.2511080671

Udovicić, L., Mark, F., & Bothe, E. 1994, Yields of single-strand breaks in double-stranded calf thymus DNA irradiated in aqueous solution in the presence of oxygen and scavengers, Radiation Research, 140, 166

Ulmer, K. M., Gomez, R. F., & Sinskey, A. J. 1979, Ionizing radiation damage to the folded chromosome of Escherichia coli K-12: Sedimentation properties of irradiated nucleoids and chromosomal deoxyribonucleic acid, Journal of Bacteriology, 138, 475

UNSCEAR. 2000, Annex B Exposures from natural radiation sources, United Nations Scientific Comittee on the Effects of Atomic Radiation, Annex B

—. 2008, Sources and effects of ionizing radiation: Exposures of the public and workers from various sources of radiation, Tech. rep., Vienna

Varley, F. S. 1992, Neutron scattering lengths and cross sectioirn, Neutron News, 3, 29

Ward, J. F. 1988, DNA Damage Produced by Ionizing Radiation in Mammalian Cells: Identities, Mechanisms, Progress in nucleic acid research and molecular biology, 35, 95

Watson, J. D., & Crick, F. H. C. 1953, Molecular Structure of Nucleic Acids, Nature, 171, 737

Wiser, M. J., Ribeck, N., & Lenski, R. E. 2013, Long-term dynamics of adaptation in asexual populations, Science, 342, 1364

Wloch, D. M., Szafraniec, K., Borts, R. H., & Korona, R. 2001, Direct estimate of the mutation rate and the distribution of fitness effects in the yeast Saccharomyces cerevisiae, Genetics, 159, 441

Wright, S. 1932, The roles of mutations, inbreeding, crossbreeding and selection in evolution, Proceedings of the Sixth International Congress of Genetics, 1, 356

Zavilgelsky, G. B., Abilev, S. K., Sukhodolets, V. V., & Ahmad, S. I. 1998, Isolation and analysis of UV and radio-resistant bacteria from Chernobyl, Journal of Photochemistry and Photobiology B: Biology, 43, 152

Zhou, H., Ivanov, V., Lien, Y.-C., Davidson, M., & Hei, T. 2008, Mitochondrial Function and NF-κB Mediated Signaling in Radiation-Induced Bystander Effects, Cancer Research, 68, 2233

Appendix A

DNA level simulation commands

Here we briefly detail the commands made available in our application for simulating DNA damage mechanistically. Optional commands are placed in square brackets ([]).

A.1 Command Line Options

The application is launched from the command line with the following syntax:

`molecular -t <int> -p <int> -m <str>`

This permits the user (when running in Geant4 multithreaded mode) to set the number of threads (`-t`), the Geant4-DNA physics list option (`-p`, 0 is the default list), and the macro file to be executed (`-m`).

A.2 Geometry Related Commands

- `/world/worldSize <s> <unit>` Side length for the world.
- `/dnageom/setVerbose <int>` Print verbose debugging information in related to the DNA geometry.
- `/dnageom/definitionFile <filepath>` Path to file that defines placement locations.
- `/dnageom/placementVolume <name> <filepath> [<twist>]` Set a placement volume, twist is an optional boolean parameter (written as true or false).
- `/dnageom/fractalScaling <x> <y> <z> <unit>` Scaling and units for the fractal along each axis.

- `/dnageom/placementSize <x> <y> <z> <unit>` Side length for each placement.
- `/dnageom/checkOverlaps <bool>` Check overlaps of molecules and fractal placements being placed for debugging.
- `/dnageom/setSmartVoxels <int>` Change the amount of voxelisation in the Geant4 geometry optimisation for a faster simulation initialisation, but slower overall simulation (1 refers to maximal optimisation in initialisation).
- Chromosomes can be added to define regions of interest. For all chromosome types, a name is required. The `x`, `y` and `z` variables refer to the translation of the chromosome, and the optional rotations in x, y and z are Euler rotations.
  - `/chromosome/add sphere <name> <rad> <x> <y> <z> <unit> [<rx> <ry> <rz>]` Add a spherical chromosome with a specified radius.
  - `/chromosome/add cyl <name> <rad> <height> <x> <y> <z> <unit> [<rx> <ry> <rz>]` Add a cylindrical chromosome with a specified height and radius.
  - `/chromosome/add rod <name> <rad> <height> <x> <y> <z> <unit> [<rx> <ry> <rz>]` Add a rod shaped chromosome. This is a cylinder of a specified height, with two hemispherical end caps. The radius of the cylinder and end caps is specified.
  - `/chromosome/add ellipse <name> <sx> <sy> <sz> <x> <y> <z> <unit> [<rx> <ry> <rz>]` Add an ellipsoidal chromosome, with semi-major axes `<sx>` `<sy>` and `<sz>`.
- `/chromosome/plotData <filename>` Save a scatter plot (x,y,z data points) of all chromosome positions.

A.3 Damage Model Related Commands

- `/dnageom/interactionDirectRange <d> <unit>` Distance from DNA molecules at which energy deposits count towards DNA damage.
- `/dnageom/radicalKillDistance <d> <unit>` Distance from DNA at which to stop tracking radicals.
- `/dnadamage/directDamageLower <d>` Minimum Energy required for an SSB.

- `/dnadamage/directDamageUpper <d>` Energy required for an SSB to definitely occur.
- `/dnadamage/indirectOHBaseChance <d>` Chance $\in [0,1]$ of a $^{\bullet}$OH damaging a base.
- `/dnadamage/indirectOHStrandChance <d>` Chance $\in [0,1]$ of a $^{\bullet}$OH damaging a sugar-phosphate moiety.
- `/dnadamage/inductionOHChance <d>` Chance $\in [0,1]$ of a reaction between a base and $^{\bullet}$OH yielding a strand break.
- `/dnadamage/indirectHBaseChance <d>` Chance $\in [0,1]$ of a $H^{\bullet}$ damaging a base.
- `/dnadamage/indirectHStrandChance <d>` Chance $\in [0,1]$ of a $H^{\bullet}$ damaging sugar-phosphate moiety.
- `/dnadamage/inductionHChance <d>` Chance $\in [0,1]$ of a reaction between a base and $H^{\bullet}$ yielding a strand break.
- `/dnadamage/indirectEaqBaseChance <d>` Chance $\in [0,1]$ of a e^{-}_{aq} damaging a base.
- `/dnadamage/indirectEaqStrandChance <d>` Chance $\in [0,1]$ of a e^{-}_{aq} damaging sugar-phosphate moiety.
- `/dnadamage/inductionEaqChance <d>` Chance $\in [0,1]$ of a reaction between a base and e^{-}_{aq} yielding a strand break.

A.4 Analysis Related Commands

- `/analysisDNA/saveStrands <bool>` Boolean for whether text representations of strand damage ought be saved.
- `/analysisDNA/strandDir <directory>` Directory to save text representations of DNA damage fragments.
- `/analysisDNA/fragmentGap <int>` Integer for the gap required between DNA fragments, that they be classified as separate damage events (d_s in the text). A value of zero will use placement boundaries to separate fragments (useful for plasmids).

- `/analysisDNA/diagnosticChain <int>` Accepts an integer referring to the index of one chain. In saved histograms, only that strand is saved (useful for debugging).
- `/analysisDNA/dsbDistance <int>` Set the maximum separation of damage events (in base pairs) for two nearby SSBs to be considered a DSB (referred to as d_{DSB} in the text). Must be less than 31.

A.5 Testing related commands

- `/dnatests/chromosome` Test Chromosomes are correctly positioned.
- `/dnatests/basepairs` Test data structures for reconstructing base pairs all work.
- `/analysisDNA/testClassifier` Run unit test of the break classification routine.

Appendix B

DNA level simulation class summary

A brief description of the classes in our molecular level simulations is provided here.

- **MolecularActionInitialization** Geant4 required class, which instantiates custom action classes.
- **MolecularAnalysisManager** Handles all the analysis done in the application. Parameters inside this class can be altered based on messages from the analysis messenger. To aid analysis, the class file defines individual damage records which represent each damage event. A binary tree is also implemented in this class to rapidly identify spatially correlated damage events.
- **MolecularAnalysisMessenger** Defines the macro commands that can be used to interact with the analysis manager.
- **MolecularChemistryConstructor** Instantiates the chemistry processes, and defines the reaction rates used in the simulation.
- **MolecularChromosomeFactory** Interprets the chromosome definition mini-language. This class can take a chromosome definition string, and return a chromosome object based on the strings contents.
- **MolecularChromosomeHit** Geant4 hit class that holds energy depositions in chromosomes across an entire event.
- **MolecularChromosomeMapper** Manages chromosome definitions, mapping spatial positions onto chromosome locations.
- **MolecularChromosomeMessenger** Messenger class that handles macro commands for all calls related to chromosomes.

- **MolecularCylindricalChromosome** Defines a cylindrical chromosome.
- **MolecularDamageModelMessenger** Messenger class that handles macro commands related to the damage model.
- **MolecularDetectorConstruction** Implementation of the Geant4 detector construction class, responsible for instantiating the MolecularDNAGeometry class, which builds the DNA geometry.
- **MolecularDetectorMessenger** Messenger class for handling non-DNA related geometry macro commands.
- **MolecularDNAGeometry** Here the DNA geometry is built, and a large part of the information related to is stored, or is passed through. Can be accessed through the detector construction class.
- **MolecularDNAGeometryMessenger** Messenger class for handling DNA related geometry macro commands.
- **MolecularDNAHit** Hit class for hits on DNA molecules. Hits can be added together (via a method), which makes sure that two hits on the same base pair are properly combined to be consistent with the damage model.
- **MolecularDNAWorld** Definition of the parallel which contains the physically placed DNA geometry.
- **MolecularEllipticalChromosome** Defines an ellipsoidal chromosome.
- **MolecularEventAction** Implementation of the Geant4 event action class, which provides hooks at the start and end of each event.
- **MolecularOctreeNode** Defines the octree nodes that hold the position of each molecule in a placement volume.
- **MolecularParallelWorldPhysics** Physics list to ensure that parallel worlds function correctly.
- **MolecularPhysicsList** Geant4 required class, defining the physics to be used, and instantiating the chemistry processes.
- **MolecularPlacementVolumeInfo** DNA helper class that holds information related to each placement volume, so that base pairs can be properly indexed.

- **MolecularPrimaryGeneratorAction** Geant4 required class that generates primary particles. Instantiates a general particle source.
- **MolecularPrimaryGeneratorMessenger** Messenger class that can be used to store any macro commands related to the generation of primary particles.
- **MolecularRodChromosome** Defines a rod shaped chromosome.
- **MolecularRunAction** Class that allows hooks into the start and end of each run, used to initialise the analysis manager.
- **MolecularSphericalChromosome** Defines a spherical chromosome.
- **MolecularStackingAction** The stacking action provides hooks that run after all physics has been completed. These are used to start the chemistry simulation.
- **MolecularSteppingAction** Provides user hooks for every physical step. Here, physical damage is recorded and stored in MolecularDNAHit objects.
- **MolecularTimeStepAction** Provides user hooks for every chemistry time-step and chemical reaction. Here, chemical damage is recorded and stored in MolecularDNAHit objects.
- **MolecularUtilityFunctions** This file provides useful methods as free functions in the namespace `utility`
- **MolecularVirtualChromosome** Provides the abstract definition of a chromosome.

Appendix C

Model Parameter Sweep Tables

The following pages provide a number of tables that are referenced in text (and some that aren't) from the parameter sweeps we run.

Table C.1 Parameter sweep of direct damage for a 3 Å damage radius, using the physics constructor G4EmDNAPhysics_option4.

Primary	E (eV)	n –	Rad (Å)	E_{min} (eV)	E_{max} (eV)	None –	SSB –	SSB+ –	2SSB –	DSB –	DSB+ –	DSB++ –	SSB/DSB –
e^-	300	3333333	3	12.60	12.60	3347	1784	73	11	71	4	0	24.91
e^-	500	2000000	3	12.60	12.60	3312	1922	72	19	60	0	1	33.00
e^-	1000	1000000	3	12.60	12.60	3357	1933	76	13	55	5	0	33.70
e^-	3000	333333	3	12.60	12.60	3904	2100	60	15	38	4	0	51.79
e^-	4500	222222	3	12.60	12.60	4034	2190	59	10	45	5	0	45.18
e^-	300	3333333	3	15.00	15.00	4177	1107	30	4	26	0	0	43.88
e^-	500	2000000	3	15.00	15.00	4263	1118	38	6	13	2	0	77.47
e^-	1000	1000000	3	15.00	15.00	4428	1154	26	6	22	0	0	53.91
e^-	3000	333333	3	15.00	15.00	4785	1264	31	1	16	1	0	76.24
e^-	4500	222222	3	15.00	15.00	5016	1246	20	8	17	1	0	70.78
e^-	300	3333333	3	17.50	17.50	4612	772	13	0	14	0	0	56.07
e^-	500	2000000	3	17.50	17.50	4564	670	14	3	10	0	0	68.70
e^-	1000	1000000	3	17.50	17.50	4745	725	15	1	11	0	0	67.36
e^-	3000	333333	3	17.50	17.50	5292	713	8	0	7	1	0	90.12
e^-	4500	222222	3	17.50	17.50	5381	733	9	1	9	0	0	82.56
e^-	300	3333333	3	21.10	21.10	4863	440	5	1	4	0	0	111.50
e^-	500	2000000	3	21.10	21.10	4940	446	5	1	2	0	0	226.00
e^-	1000	1000000	3	21.10	21.10	5189	487	7	0	4	0	0	123.50
e^-	3000	333333	3	21.10	21.10	5584	439	2	0	2	0	0	220.50
e^-	4500	222222	3	21.10	21.10	5867	463	1	0	3	0	0	154.67
e^-	300	3333333	3	30.00	30.00	5272	166	0	0	0	0	0	-
e^-	500	2000000	3	30.00	30.00	5155	214	0	1	2	0	0	107.50
e^-	1000	1000000	3	30.00	30.00	5282	215	0	0	1	0	0	215.00
e^-	3000	333333	3	30.00	30.00	5971	257	0	0	0	0	0	-
e^-	4500	222222	3	30.00	30.00	6177	249	1	0	1	0	0	250.00
e^-	300	3333333	3	5.00	37.50	4304	1066	31	5	17	3	0	55.10
e^-	500	2000000	3	5.00	37.50	4343	1072	25	7	13	1	0	78.86
e^-	1000	1000000	3	5.00	37.50	4483	1116	21	7	11	0	0	104.00
e^-	3000	333333	3	5.00	37.50	4839	1163	29	3	13	1	0	85.36
e^-	4500	222222	3	5.00	37.50	5039	1197	21	3	10	1	0	111.00

Table C.2 Parameter sweep of direct damage for a 4 Å damage radius, using the physics constructor G4EmDNAPhysics_option4.

Primary	E (eV)	n –	Rad (Å)	E_{min} (eV)	E_{max} (eV)	None –	SSB –	SSB+ –	2SSB –	DSB –	DSB+ –	DSB++ –	SSB/DSB –
e^-	300	3333333	4	12.60	12.60	4768	3837	396	61	294	59	16	11.64
e^-	500	2000000	4	12.60	12.60	4699	3873	401	67	261	69	14	12.62
e^-	1000	1000000	4	12.60	12.60	5186	4014	431	81	286	46	12	13.16
e^-	3000	333333	4	12.60	12.60	6161	4577	300	55	226	45	7	17.74
e^-	4500	222222	4	12.60	12.60	6493	4837	268	39	215	39	9	19.56
e^-	300	3333333	4	15.00	15.00	6542	2705	169	34	136	17	0	19.01
e^-	500	2000000	4	15.00	15.00	6310	2665	179	36	126	15	3	20.00
e^-	1000	1000000	4	15.00	15.00	6915	2792	192	29	122	12	5	21.68
e^-	3000	333333	4	15.00	15.00	8094	2994	150	30	97	13	2	28.34
e^-	4500	222222	4	15.00	15.00	8671	3024	149	32	89	6	1	33.39
e^-	300	3333333	4	17.50	17.50	7319	1989	92	11	78	7	0	24.61
e^-	500	2000000	4	17.50	17.50	7251	1923	97	19	59	9	1	29.55
e^-	1000	1000000	4	17.50	17.50	7986	1874	87	10	56	6	0	31.79
e^-	3000	333333	4	17.50	17.50	9401	1880	75	14	40	2	0	46.88
e^-	4500	222222	4	17.50	17.50	9882	1979	60	5	28	2	0	68.13
e^-	300	3333333	4	21.10	21.10	7976	1355	54	4	37	2	0	36.23
e^-	500	2000000	4	21.10	21.10	8137	1257	42	3	33	1	2	36.17
e^-	1000	1000000	4	21.10	21.10	8682	1255	39	6	27	0	1	46.43
e^-	3000	333333	4	21.10	21.10	10101	1235	31	4	19	1	0	63.50
e^-	4500	222222	4	21.10	21.10	10500	1243	27	9	17	0	0	75.24
e^-	300	3333333	4	30.00	30.00	8874	546	2	0	10	0	0	54.80
e^-	500	2000000	4	30.00	30.00	8985	611	9	1	4	0	0	155.25
e^-	1000	1000000	4	30.00	30.00	9350	641	11	1	7	0	0	93.29
e^-	3000	333333	4	30.00	30.00	10720	644	5	2	6	0	0	108.50
e^-	4500	222222	4	30.00	30.00	11284	736	8	2	3	0	0	248.67
e^-	300	3333333	4	5.00	37.50	6672	2502	130	30	109	12	2	21.64
e^-	500	2000000	4	5.00	37.50	6930	2455	163	23	118	10	4	20.01
e^-	1000	1000000	4	5.00	37.50	7079	2423	131	32	100	14	0	22.68
e^-	3000	333333	4	5.00	37.50	8614	2749	114	26	85	6	1	31.40
e^-	4500	222222	4	5.00	37.50	9044	2798	92	31	76	4	0	36.51

Table C.3 Parameter sweep of direct damage for a 5 Å damage radius, using the physics constructor G4EmDNAPhysics_option4.

Primary	E (eV)	n –	Rad (Å)	$E_{\min}$ (eV)	$E_{\max}$ (eV)	None –	SSB –	SSB+ –	2SSB –	DSB –	DSB+ –	DSB++ –	SSB/DSB –
e^-	300	3333333	5	12.60	12.60	5272	5246	819	122	621	205	41	7.14
e^-	500	2000000	5	12.60	12.60	5238	5232	777	150	601	202	56	7.17
e^-	1000	1000000	5	12.60	12.60	5874	5664	773	170	557	180	43	8.47
e^-	3000	333333	5	12.60	12.60	7491	6649	666	115	492	127	29	11.47
e^-	4500	222222	5	12.60	12.60	8087	7028	640	105	482	120	25	12.40
e^-	300	3333333	5	15.00	15.00	7347	4110	465	63	330	80	18	10.84
e^-	500	2000000	5	15.00	15.00	7451	4003	463	79	310	67	22	11.39
e^-	1000	1000000	5	15.00	15.00	8386	4221	419	73	295	55	14	12.95
e^-	3000	333333	5	15.00	15.00	10412	4673	359	70	235	48	10	17.41
e^-	4500	222222	5	15.00	15.00	11149	4708	314	57	241	41	7	17.57
e^-	300	3333333	5	17.50	17.50	8559	3229	276	41	195	28	4	15.62
e^-	500	2000000	5	17.50	17.50	8732	3074	254	48	194	23	9	14.94
e^-	1000	1000000	5	17.50	17.50	9758	3054	217	52	153	14	2	19.66
e^-	3000	333333	5	17.50	17.50	12097	3166	204	32	113	21	1	25.20
e^-	4500	222222	5	17.50	17.50	13209	3151	161	31	116	24	0	23.88
e^-	300	3333333	5	21.10	21.10	9812	2292	136	23	113	10	1	19.77
e^-	500	2000000	5	21.10	21.10	9897	2170	163	20	89	9	2	23.53
e^-	1000	1000000	5	21.10	21.10	10659	2211	134	19	85	4	1	26.27
e^-	3000	333333	5	21.10	21.10	13331	2171	86	7	56	3	0	38.37
e^-	4500	222222	5	21.10	21.10	14019	2238	78	11	61	4	1	35.26
e^-	300	3333333	5	30.00	30.00	10896	1140	33	2	19	0	0	61.84
e^-	500	2000000	5	30.00	30.00	10994	1108	40	6	18	2	0	57.70
e^-	1000	1000000	5	30.00	30.00	12273	1126	28	9	22	1	0	50.57
e^-	3000	333333	5	30.00	30.00	14413	1185	17	5	14	1	0	80.47
e^-	4500	222222	5	30.00	30.00	15351	1205	23	1	10	0	0	122.90
e^-	300	3333333	5	5.00	37.50	7937	3784	355	49	289	45	8	12.25
e^-	500	2000000	5	5.00	37.50	7890	3648	333	61	276	53	8	11.99
e^-	1000	1000000	5	5.00	37.50	8636	3797	307	64	260	28	11	13.94
e^-	3000	333333	5	5.00	37.50	10779	4258	264	52	213	21	4	19.22
e^-	4500	222222	5	5.00	37.50	11691	4381	236	49	209	30	8	18.89

Table C.4 Parameter sweep of direct damage for a 6 Å damage radius, using the physics constructor G4EmDNAPhysics_option4.

Primary	E (eV)	n –	Rad (Å)	E_{min} (eV)	E_{max} (eV)	None –	SSB –	SSB+ –	2SSB –	DSB –	DSB+ –	DSB++ –	SSB/DSB –
e^-	300	3333333	6	12.60	12.60	5514	6334	1336	185	992	410	107	5.21
e^-	500	2000000	6	12.60	12.60	5382	6257	1280	208	980	439	115	5.05
e^-	1000	1000000	6	12.60	12.60	6352	6971	1192	246	854	336	85	6.60
e^-	3000	333333	6	12.60	12.60	8376	8431	1094	230	797	257	81	8.59
e^-	4500	222222	6	12.60	12.60	9092	9090	1084	182	729	232	60	10.14
e^-	300	3333333	6	15.00	15.00	7812	5072	823	115	587	186	40	7.39
e^-	500	2000000	6	15.00	15.00	7970	5027	746	136	531	155	39	8.15
e^-	1000	1000000	6	15.00	15.00	8967	5461	714	117	507	123	32	9.50
e^-	3000	333333	6	15.00	15.00	11762	5993	570	112	407	85	16	13.14
e^-	4500	222222	6	15.00	15.00	13153	6269	567	118	412	78	24	13.53
e^-	300	3333333	6	17.50	17.50	9528	4264	522	80	326	88	11	11.45
e^-	500	2000000	6	17.50	17.50	9692	4216	468	96	340	79	19	10.91
e^-	1000	1000000	6	17.50	17.50	11070	4207	440	84	284	48	16	13.59
e^-	3000	333333	6	17.50	17.50	14181	4451	335	75	243	43	15	16.15
e^-	4500	222222	6	17.50	17.50	15355	4370	288	49	220	35	8	17.90
e^-	300	3333333	6	21.10	21.10	11355	3136	294	39	194	30	5	15.15
e^-	500	2000000	6	21.10	21.10	11028	3054	275	43	193	27	6	14.92
e^-	1000	1000000	6	21.10	21.10	12550	3111	255	45	154	27	6	18.24
e^-	3000	333333	6	21.10	21.10	15748	3128	160	30	120	16	5	23.53
e^-	4500	222222	6	21.10	21.10	17110	3132	167	30	115	14	3	25.22
e^-	300	3333333	6	30.00	30.00	12889	1705	64	7	51	3	0	32.89
e^-	500	2000000	6	30.00	30.00	12814	1669	75	7	58	3	2	27.79
e^-	1000	1000000	6	30.00	30.00	14239	1695	60	12	39	5	0	40.16
e^-	3000	333333	6	30.00	30.00	17435	1765	35	6	29	1	0	60.20
e^-	4500	222222	6	30.00	30.00	18869	1799	48	6	26	2	0	66.18
e^-	300	3333333	6	5.00	37.50	8809	4752	647	87	476	100	26	9.11
e^-	500	2000000	6	5.00	37.50	8775	4745	585	115	448	99	33	9.39
e^-	1000	1000000	6	5.00	37.50	9856	4966	548	129	427	87	25	10.47
e^-	3000	333333	6	5.00	37.50	12664	5451	478	95	311	59	11	15.81
e^-	4500	222222	6	5.00	37.50	13869	5780	447	82	339	60	16	15.20

Table C.5 Parameter sweep of direct damage for a 7 Å damage radius, using the physics constructor G4EmDNAPhysics_option4.

Primary	E (eV)	n –	Rad (Å)	E_{min} (eV)	E_{max} (eV)	None –	SSB –	SSB+ –	2SSB –	DSB –	DSB+ –	DSB++ –	SSB/DSB –
e^-	300	3333333	7	12.60	12.60	5679	7113	1788	259	1338	726	185	4.07
e^-	500	2000000	7	12.60	12.60	5566	7165	1759	319	1269	671	219	4.28
e^-	1000	1000000	7	12.60	12.60	6509	7975	1626	328	1282	565	197	4.86
e^-	3000	333333	7	12.60	12.60	9140	10060	1503	279	1132	441	135	6.93
e^-	4500	222222	7	12.60	12.60	10249	10802	1404	258	1140	395	99	7.63
e^-	300	3333333	7	15.00	15.00	8317	6113	1186	172	799	318	74	6.27
e^-	500	2000000	7	15.00	15.00	8368	6171	1129	211	850	321	85	5.98
e^-	1000	1000000	7	15.00	15.00	9710	6525	1065	212	730	251	63	7.47
e^-	3000	333333	7	15.00	15.00	13314	7358	868	173	673	165	47	9.49
e^-	4500	222222	7	15.00	15.00	14834	7819	878	173	687	165	42	9.92
e^-	300	3333333	7	17.50	17.50	10281	5177	782	116	583	134	36	8.07
e^-	500	2000000	7	17.50	17.50	10465	4940	729	140	570	158	38	7.58
e^-	1000	1000000	7	17.50	17.50	11941	5077	647	128	463	94	23	10.09
e^-	3000	333333	7	17.50	17.50	16001	5448	539	101	383	72	18	12.87
e^-	4500	222222	7	17.50	17.50	17583	5512	475	98	365	75	15	13.37
e^-	300	3333333	7	21.10	21.10	12163	4104	459	53	333	69	18	10.99
e^-	500	2000000	7	21.10	21.10	12236	3934	440	81	322	57	20	11.17
e^-	1000	1000000	7	21.10	21.10	13698	3847	360	72	273	45	14	12.89
e^-	3000	333333	7	21.10	21.10	17993	4025	301	72	226	30	3	16.98
e^-	4500	222222	7	21.10	21.10	19653	4085	277	42	209	29	7	17.98
e^-	300	3333333	7	30.00	30.00	14606	2345	120	15	92	4	0	25.83
e^-	500	2000000	7	30.00	30.00	14219	2378	152	16	89	8	0	26.25
e^-	1000	1000000	7	30.00	30.00	15853	2252	111	20	71	7	2	29.79
e^-	3000	333333	7	30.00	30.00	20063	2319	79	14	63	3	0	36.55
e^-	4500	222222	7	30.00	30.00	21821	2424	84	15	62	5	2	36.57
e^-	300	3333333	7	5.00	37.50	9381	5792	858	144	599	168	48	8.34
e^-	500	2000000	7	5.00	37.50	9359	5654	874	168	645	191	51	7.55
e^-	1000	1000000	7	5.00	37.50	10848	6020	813	163	627	154	44	8.48
e^-	3000	333333	7	5.00	37.50	14202	6774	690	150	473	102	30	12.59
e^-	4500	222222	7	5.00	37.50	15696	7101	638	128	453	74	25	14.25

Table C.6 Parameter sweep of direct damage for a 8 Å damage radius, using the physics constructor G4EmDNAPhysics_option4.

Primary	E(eV)	n		Rad E1 (eV)	E2 (eV)	None	SSB	SSB+	2SSB	DSB	DSB+	DSB++	SSB/DSB
e^-	300	3333333	8	12.60	12.60	5807	7825	2224	313	1603	1041	321	3.49
e^-	500	2000000	8	12.60	12.60	5691	7852	2130	395	1655	968	370	3.47
e^-	1000	1000000	8	12.60	12.60	6646	8726	2136	441	1646	879	306	3.99
e^-	3000	333333	8	12.60	12.60	9397	11428	1945	368	1580	629	218	5.66
e^-	4500	222222	8	12.60	12.60	11041	12494	1803	349	1517	614	173	6.36
e^-	300	3333333	8	15.00	15.00	8605	7109	1539	236	1166	514	129	4.91
e^-	500	2000000	8	15.00	15.00	8537	6866	1465	276	1055	477	146	5.13
e^-	1000	1000000	8	15.00	15.00	10003	7332	1415	300	1048	397	135	5.73
e^-	3000	333333	8	15.00	15.00	14430	8637	1195	232	927	280	95	7.73
e^-	4500	222222	8	15.00	15.00	16205	9109	1122	251	837	284	82	8.71
e^-	300	3333333	8	17.50	17.50	10769	6164	1108	169	859	269	76	6.18
e^-	500	2000000	8	17.50	17.50	10892	5866	1059	189	755	256	73	6.56
e^-	1000	1000000	8	17.50	17.50	12703	6148	950	238	650	199	63	8.04
e^-	3000	333333	8	17.50	17.50	17864	6674	689	133	573	139	37	10.01
e^-	4500	222222	8	17.50	17.50	19573	6783	658	144	492	120	37	11.69
e^-	300	3333333	8	21.10	21.10	12753	4890	688	108	560	114	31	8.07
e^-	500	2000000	8	21.10	21.10	13023	4774	662	109	553	124	31	7.83
e^-	1000	1000000	8	21.10	21.10	14701	4885	569	127	439	75	26	10.34
e^-	3000	333333	8	21.10	21.10	20095	5105	443	107	341	57	20	13.53
e^-	4500	222222	8	21.10	21.10	21988	5210	416	75	305	42	14	15.79
e^-	300	3333333	8	30.00	30.00	15608	3002	235	28	164	17	2	17.84
e^-	500	2000000	8	30.00	30.00	15741	2930	226	42	165	20	3	17.01
e^-	1000	1000000	8	30.00	30.00	17528	3046	180	43	127	14	4	22.54
e^-	3000	333333	8	30.00	30.00	22656	3021	162	25	106	4	6	27.66
e^-	4500	222222	8	30.00	30.00	24671	3064	137	33	96	8	0	31.10
e^-	300	3333333	8	5.00	37.50	9672	6581	1173	225	916	329	100	5.93
e^-	500	2000000	8	5.00	37.50	9883	6587	1142	223	935	304	82	6.02
e^-	1000	1000000	8	5.00	37.50	11379	7069	1069	265	855	256	81	7.05
e^-	3000	333333	8	5.00	37.50	15867	8105	911	207	734	158	50	9.79
e^-	4500	222222	8	5.00	37.50	17517	8551	855	192	648	155	71	10.98

Table C.7 Parameter sweep of direct damage for a 9 Å damage radius, using the physics constructor G4EmDNAPhysics_option4.

Primary	E (eV)	n –	Rad (Å)	E_{min} (eV)	E_{max} (eV)	None –	SSB –	SSB+ –	2SSB –	DSB –	DSB+ –	DSB++ –	SSB/DSB –
e^-	300	3333333	9	12.60	12.60	5757	8496	2585	327	2000	1504	520	2.83
e^-	500	2000000	9	12.60	12.60	5614	8194	2539	458	2065	1430	537	2.78
e^-	1000	1000000	9	12.60	12.60	6695	9492	2540	495	1987	1181	533	3.38
e^-	3000	333333	9	12.60	12.60	10225	12772	2408	448	1965	953	372	4.75
e^-	4500	222222	9	12.60	12.60	11446	14218	2305	414	1885	830	309	5.60
e^-	300	3333333	9	15.00	15.00	8953	7742	1860	306	1520	773	236	3.92
e^-	500	2000000	9	15.00	15.00	8993	7539	1823	377	1472	691	268	4.01
e^-	1000	1000000	9	15.00	15.00	10560	8077	1776	393	1307	569	204	4.93
e^-	3000	333333	9	15.00	15.00	15333	10101	1527	306	1240	423	148	6.59
e^-	4500	222222	9	15.00	15.00	17562	10408	1442	287	1164	413	154	7.01
e^-	300	3333333	9	17.50	17.50	11018	6679	1475	233	1067	443	112	5.17
e^-	500	2000000	9	17.50	17.50	11071	6522	1396	267	1026	411	135	5.21
e^-	1000	1000000	9	17.50	17.50	13360	6988	1193	251	946	301	114	6.20
e^-	3000	333333	9	17.50	17.50	19018	7873	1057	206	765	243	68	8.49
e^-	4500	222222	9	17.50	17.50	21337	8097	918	198	685	195	70	9.70
e^-	300	3333333	9	21.10	21.10	13363	5689	989	133	778	174	56	6.76
e^-	500	2000000	9	21.10	21.10	13533	5636	963	167	703	188	47	7.21
e^-	1000	1000000	9	21.10	21.10	15331	5685	773	178	560	126	45	9.08
e^-	3000	333333	9	21.10	21.10	21667	5934	621	114	459	112	30	11.10
e^-	4500	222222	9	21.10	21.10	23919	6169	613	116	422	81	29	12.97
e^-	300	3333333	9	30.00	30.00	16616	3860	359	42	240	26	9	15.49
e^-	500	2000000	9	30.00	30.00	16786	3674	323	73	264	33	7	13.39
e^-	1000	1000000	9	30.00	30.00	18881	3582	305	61	207	26	8	16.38
e^-	3000	333333	9	30.00	30.00	25270	3647	217	57	160	18	3	21.66
e^-	4500	222222	9	30.00	30.00	27660	3736	197	35	140	12	2	25.77
e^-	300	3333333	9	5.00	37.50	10182	7541	1557	271	1273	482	167	4.87
e^-	500	2000000	9	5.00	37.50	10125	7339	1489	294	1150	425	152	5.28
e^-	1000	1000000	9	5.00	37.50	11862	7743	1419	331	1087	383	138	5.90
e^-	3000	333333	9	5.00	37.50	17149	9423	1288	290	921	270	79	8.66
e^-	4500	222222	9	5.00	37.50	19021	9875	1125	242	907	246	84	9.09

Table C.8 Parameter sweep of direct damage for a 5 Å damage radius, using the physics constructor G4EmDNAPhysics (Default). Fewer DSBs are recorded compared to the option4 constructor.

Primary	E (eV)	n –	Rad (Å)	E_{min} (eV)	E_{max} (eV)	None –	SSB –	SSB+ –	2SSB –	DSB –	DSB+ –	DSB++ –	SSB/DSB –
e^-	300	3333333	5	12.60	12.60	24943	4941	710	117	521	144	37	8.22
e^-	500	2000000	5	12.60	12.60	23033	5103	715	160	545	136	36	8.34
e^-	1000	1000000	5	12.60	12.60	23712	5478	663	147	470	106	40	10.21
e^-	3000	333333	5	12.60	12.60	27296	6278	601	112	445	104	21	12.26
e^-	4500	222222	5	12.60	12.60	28651	6553	555	120	388	82	23	14.66
e^-	300	3333333	5	15.00	15.00	27304	3739	387	62	249	32	7	14.54
e^-	500	2000000	5	15.00	15.00	25211	3644	374	76	269	41	2	13.12
e^-	1000	1000000	5	15.00	15.00	25962	3874	347	78	228	34	11	15.75
e^-	3000	333333	5	15.00	15.00	30406	4203	272	59	194	23	4	20.52
e^-	4500	222222	5	15.00	15.00	31618	4251	257	47	163	19	6	24.23
e^-	300	3333333	5	17.50	17.50	28083	2633	200	32	117	17	1	21.22
e^-	500	2000000	5	17.50	17.50	26654	2609	161	46	126	10	0	20.71
e^-	1000	1000000	5	17.50	17.50	27425	2501	140	28	94	7	3	25.66
e^-	3000	333333	5	17.50	17.50	32121	2547	115	27	77	6	1	32.01
e^-	4500	222222	5	17.50	17.50	33700	2565	127	22	85	11	3	27.41
e^-	300	3333333	5	21.10	21.10	29295	1889	104	16	68	4	1	27.52
e^-	500	2000000	5	21.10	21.10	27436	1760	96	15	58	7	0	28.78
e^-	1000	1000000	5	21.10	21.10	28445	1813	81	10	43	4	1	39.67
e^-	3000	333333	5	21.10	21.10	33068	1746	62	17	42	1	0	42.44
e^-	4500	222222	5	21.10	21.10	34537	1755	76	11	41	0	0	44.93
e^-	300	3333333	5	30.00	30.00	30063	899	20	1	18	0	0	51.11
e^-	500	2000000	5	30.00	30.00	28422	934	15	3	17	2	0	50.11
e^-	1000	1000000	5	30.00	30.00	29190	990	26	3	10	3	0	78.38
e^-	3000	333333	5	30.00	30.00	33892	981	22	3	8	0	0	125.75
e^-	4500	222222	5	30.00	30.00	35596	1045	19	1	5	0	0	213.00
e^-	300	3333333	5	5.00	37.50	27369	3631	286	62	227	28	9	15.07
e^-	500	2000000	5	5.00	37.50	25274	3508	303	68	202	26	10	16.30
e^-	1000	1000000	5	5.00	37.50	26409	3696	251	59	184	21	7	18.90
e^-	3000	333333	5	5.00	37.50	30359	3957	224	61	148	21	6	24.24
e^-	4500	222222	5	5.00	37.50	31718	4222	208	43	153	12	2	26.78

Table C.9 Parameter sweep of direct damage for a 6 Å damage radius, using the physics constructor G4EmDNAPhysics (Default). Fewer DSBs are recorded compared to the option4 constructor.

Primary	E (eV)	n –	Rad (Å)	E_{min} (eV)	E_{max} (eV)	None –	SSB –	SSB+ –	2SSB –	DSB –	DSB+ –	DSB++ –	SSB/DSB –
e-	300	3333333	6	12.60	12.60	26825	6135	1132	238	794	306	78	6.37
e-	500	2000000	6	12.60	12.60	24361	6155	1137	250	757	252	92	6.85
e-	1000	1000000	6	12.60	12.60	25327	6639	1049	278	690	233	80	7.94
e-	3000	333333	6	12.60	12.60	29846	8049	908	229	614	177	57	10.83
e-	4500	222222	6	12.60	12.60	31489	8533	825	200	643	152	47	11.35
e-	300	3333333	6	15.00	15.00	29114	4903	660	134	430	121	17	10.03
e-	500	2000000	6	15.00	15.00	26876	4703	636	159	392	104	25	10.55
e-	1000	1000000	6	15.00	15.00	28061	4896	553	147	373	71	22	12.01
e-	3000	333333	6	15.00	15.00	33700	5358	478	113	303	57	21	15.61
e-	4500	222222	6	15.00	15.00	35141	5618	471	120	321	55	18	15.76
e-	300	3333333	6	17.50	17.50	31223	3605	366	82	227	39	6	14.90
e-	500	2000000	6	17.50	17.50	29356	3442	337	83	239	40	9	13.41
e-	1000	1000000	6	17.50	17.50	30114	3514	284	78	181	29	13	17.38
e-	3000	333333	6	17.50	17.50	35890	3548	225	50	152	18	5	21.85
e-	4500	222222	6	17.50	17.50	37599	3568	206	53	148	18	8	21.99
e-	300	3333333	6	21.10	21.10	32100	2655	189	34	168	11	3	15.81
e-	500	2000000	6	21.10	21.10	30288	2655	205	28	156	14	1	16.89
e-	1000	1000000	6	21.10	21.10	31701	2540	176	34	114	10	1	22.00
e-	3000	333333	6	21.10	21.10	36834	2586	147	23	99	8	0	25.76
e-	4500	222222	6	21.10	21.10	39603	2538	116	24	78	11	4	28.80
e-	300	3333333	6	30.00	30.00	33643	1450	52	9	34	2	0	41.97
e-	500	2000000	6	30.00	30.00	31515	1437	62	9	35	0	0	43.09
e-	1000	1000000	6	30.00	30.00	32688	1424	52	14	28	1	0	51.38
e-	3000	333333	6	30.00	30.00	38478	1495	52	7	31	2	0	47.09
e-	4500	222222	6	30.00	30.00	40429	1485	30	7	28	3	0	49.10
e-	300	3333333	6	5.00	37.50	29562	4683	486	94	373	61	10	11.85
e-	500	2000000	6	5.00	37.50	27567	4596	539	127	353	55	26	12.12
e-	1000	1000000	6	5.00	37.50	28361	4966	485	127	311	48	18	14.80
e-	3000	333333	6	5.00	37.50	33852	5428	411	104	234	45	16	20.15
e-	4500	222222	6	5.00	37.50	35630	5646	374	88	263	37	7	19.90

Table C.10 Break classifications by source for 4.5 keV electrons in a test geometry.

Radius	SSB_i	SSB_d	SSB_m	DSB_i	DSB_d	$DSB_m + DSB_{hyb}$	SSBs	DSBs
0.0	0.0	100.0	0.0	0.0	100.0	0.0	2823	203
1.0	34.0	56.2	9.8	5.7	22.6	71.7	3746	508
2.0	66.8	23.6	9.6	25.8	10.1	64.0	6560	948
3.0	79.5	12.7	7.8	39.9	4.5	55.6	10123	1478
4.0	85.8	8.3	6.0	51.4	3.8	44.8	13695	1853
5.0	89.0	5.9	5.0	56.6	3.6	39.8	17102	2249
6.0	91.1	4.6	4.3	59.9	3.4	36.7	20764	2555
7.0	92.2	3.9	4.0	63.7	2.2	34.0	24363	2945
8.0	93.5	3.0	3.5	68.3	1.7	30.0	28035	3140
9.0	94.3	2.7	3.0	69.3	2.1	28.7	32132	3426
N97	37	62	1	15	74	11	4209	236

Table C.11 Break classifications by source for 4.5 keV electrons in a test geometry, ending the simulation at 1 ns.

Radius (nm)	SSB_i (%)	SSB_d (%)	SSB_m (%)	DSB_i (%)	DSB_d (%)	$DSB_m + DSB_{hyb}$ (%)	SSBs -	DSBs -
0.0	0.0	100.0	0.0	0.0	100.0	0.0	2740	182
1.0	36.3	54.8	8.9	5.9	25.1	69.0	3758	506
2.0	66.7	24.3	9.0	23.9	9.3	66.9	6495	893
3.0	78.0	14.2	7.9	40.9	5.5	53.7	9469	1314
4.0	82.5	11.4	6.1	44.9	5.1	50.0	11419	1583
5.0	82.9	10.7	6.4	44.9	3.8	51.2	12509	1602
6.0	83.3	10.4	6.3	45.2	5.3	49.5	12983	1576
7.0	83.6	10.5	5.9	44.3	4.8	50.9	13082	1541
8.0	83.9	10.1	5.9	46.7	4.3	48.9	13057	1568
N97	37	62	1	15	74	11	4209	236

Table C.12 Break classifications by complexity when reproducing Nikjoo et al. (1997)

Energy (eV)	None (%)	SSB (%)	SSB+ (%)	2SSB (%)	DSB (%)	DSB+ (%)	DSB++ (%)	$\frac{\text{Complex DSB}}{\text{Total DSB}}$ (%)	Hits -	SSBs (Gy^{-1} Mbp^{-1})	DSBs (Gy^{-1} Mbp^{-1})
300	82.41	12.59	2.07	0.68	1.56	0.38	0.31	30.46	105792	0.265	0.039
500	80.73	13.53	2.23	0.96	1.80	0.38	0.37	29.39	90105	0.246	0.038
1000	80.24	14.22	2.24	0.92	1.68	0.33	0.36	29.11	90236	0.257	0.035
3000	82.69	13.51	1.56	0.62	1.22	0.22	0.19	24.92	115684	0.297	0.031
4500	83.97	12.80	1.33	0.53	1.06	0.18	0.13	22.51	127151	0.305	0.028

Table C.13 Break classifications by source when reproducing Nikjoo et al. (1997)

Energy (eV)	SSB_i (%)	SSB_d (%)	SSB_m (%)	DSB_i (%)	DSB_d (%)	$DSB_m + DSB_{hyb}$ (%)	SSBs -	DSBs -
300	76.9	14.7	8.4	9.9	36.8	53.3	16229	2380
500	76.7	14.7	8.6	9.4	37.6	53.0	15060	2300
1000	76.5	15.2	8.4	8.6	37.7	53.7	15692	2140
3000	78.2	15.6	6.2	7.9	40.7	51.4	18152	1874
4500	78.7	15.5	5.8	8.3	39.3	52.4	18639	1737

Appendix D

Rejection of Data Points In Fitness Analysis

The rejection and acceptance of data points is often a contentious subject, and so we address the points rejected in our experiment in the following pages. We rejected a number of measurements in our analysis, as too many colonies had grown on some Petri dishes, which can bias the colony count. In particular, this causes the spread of fitnesses measured on a particular day within one lineage at a given time point to be very large. When this occurs, the noise in the count yields a very noisy fitness measurement ($\delta F > 0.1$) between replicates, and as this noise is bigger than the change we seek to observe in our experiment, it permits one or two very noisy measurements to bias the entire experiment.

Thus, points were rejected when they showed large colony counts and large variations in fitness between replicates. Rejection of points primarily affected the LSM lines where a number of measurements were made (up to six lines) with high quantities of colonies on March 3rd, 2016 and March 10th, 2016. At the LPC, a measurement of from Line 1 in REL606 was rejected due to highly varied colony counts at 100 generations. For completeness, we show in Figure D.1 the variation in fitness between locations, without any rejection of data points. Here, a difference between the fitness at the LPC and LSM begins to become visible. This is because of the noisiness of the data that was added. Across the measurements on March 3rd and March 10th, the number of colonies in some replicates varied greatly. Some Petri dishes showed over 800 total colonies, whilst the same dilution on another dish would have less than 80. The variability of our fitness assays in these circumstances is ultimately what caused us to reject these points as unreliable. The high colony count we observed strengthens this rejection, as the accuracy of counts when there are more than 500 colonies on a plate begins to become questionable, as close colonies become difficult to distinguish. The correlation of these events in time also leads us to believe that

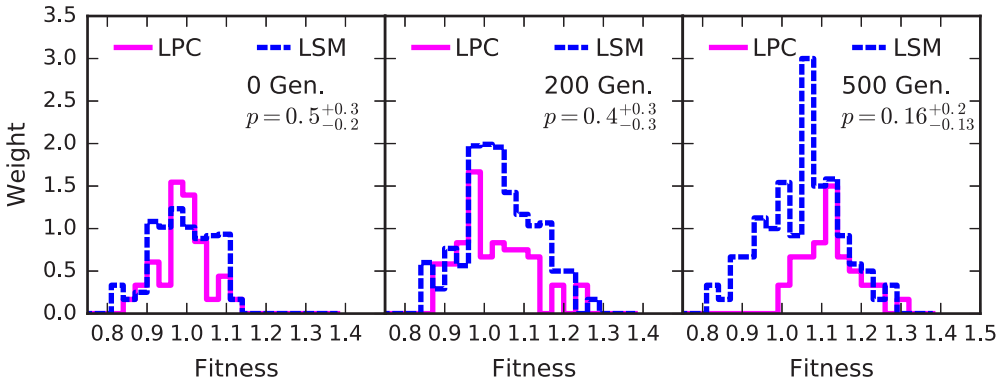


Fig. D.1 The top panel reproduces Figure 4.13, showing the distributions of fitness in the LPC and LSM, with the significance that they are different. No data points were rejected here, which causes the LSM measurements in particular to be flatter, and increases the distance between the LSM and LPC distributions at 500 generations.

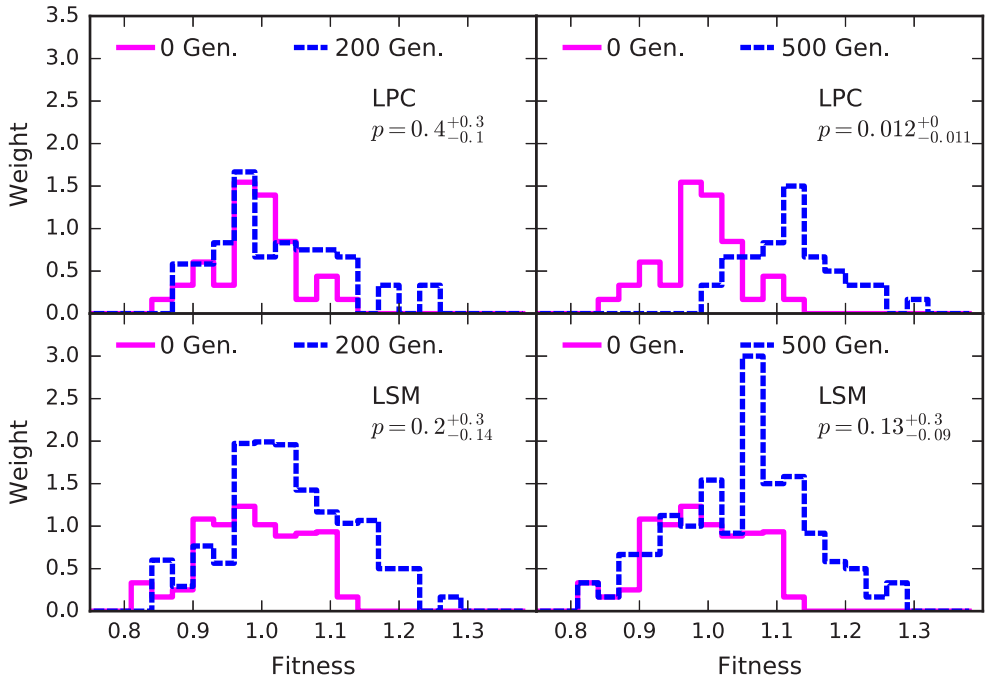


Fig. D.2 The top panel reproduces Figure 4.14, showing the likelihood that the evolved and ancestral fitnesses in each environment have changed. The flatter distribution of fitness at 500 generations at the LSM means that a majority of bootstrapped KS tests have $p > 0.15$, indicating the change in fitness between 0 and 500 generations is more consistent with these two distributions being drawn from the same mother.

these points are experimental errors, rather than a biological signal.

In Figure D.2, the changes in fitness with time at each location are considered. The distribution of fitness in the LSM at 500 generations is closer to the distribution at 0 generations than it was when the data points mentioned above were rejected. The changes in the fitness distributions when the rejected points are included in our analysis is consistent with the addition of very noisy data, flattening the distribution of fitness, though the increase in histogram points at $F = 1.1$ in the 500 generation data set may indicate that their is still some signal buried amongst the noise. Nevertheless, it seems clear that the inclusion of the rejected points adds a set of data that doesn't follow the same distribution of the other points we considered. This second distribution is quite possibly tied to whatever introduced the enormous variations in cell colony counts on March 3rd and 10th.

Appendix E

P-values from Bootstrapped KS tests

In order to estimate the likelihood that fitness had changed between the LPC and the LSM in Section 4.4, a bootstrapping was applied to a two sample Kolmogorov-Smirnov (KS) test so that each independent line could be modelled statistically with the same weight. Bootstrapping resamples (with replacement) a distribution many hundreds to thousands of times to assess how the selection of data from an underlying distribution impacts a measurement. It considers the idea that if the empirical distribution of values we have measured represents the true distribution, then sampling that distribution differently by a random process could impact our results. In the context of a KS test, bootstrapping gives a range of possible probability values (p-values), one for each test that was run, by resampling the empirical distribution function of the observed sample. The distribution of these values is shown in Figures E.1 and E.2. The two sample KS test, which tries to ascertain whether two measured distributions were drawn from the same mother distribution. The KS statistic is distributed in two sample tests such that sampling the same measured distribution repeatedly via bootstrapping will produce p-values of 1, while sampling different distributions will give p-values of 0. If two different measured distributions drawn from the same mother distribution are considered however, the randomness inherent in the bootstrapped samplings will yield p-values that are uniformly distributed.

This is clear considering the p-values in Figure E.1 where the measurements at the LPC and LSM are compared at 0, 200 and 500 generations. Here, no significant difference is seen in the three comparisons made, so we expect to see a uniform distribution of p-values, or even a distribution skewing towards $p = 1$. Immediately though, it's clear that neither a uniform, nor 'continuous' distribution of p-values is realised. This comes from the nature of the empirical distribution function, which can have a stepped behaviour[1]. This structure

[1] We linearly interpolate between fitness measurements when sampling the empirical distribution function, but when two measurements measure the same fitness, a step results

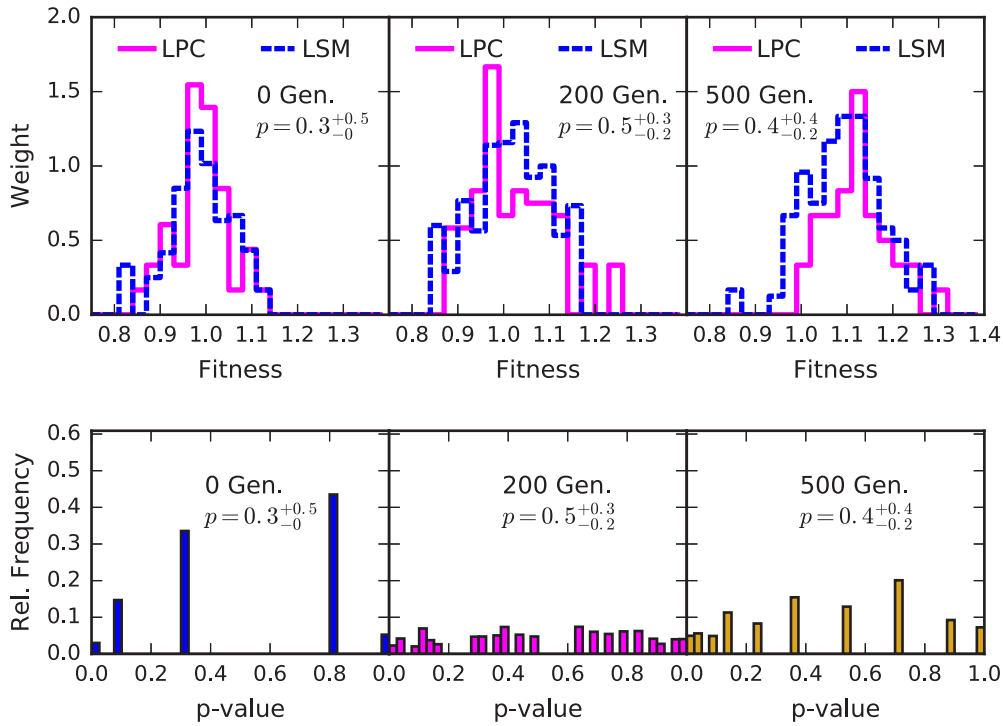


Fig. E.1 The top panel reproduces Figure 4.13, showing the distributions of fitness in the LPC and LSM, with the significance that they are different. The p-values used here show the median and upper and lower quartiles of the p-value distribution from 2000 bootstrapped KS tests. The distribution of p-values is shown in the bottom panel.

strongly favours some KS statistics over others, which leads to some p-values being strongly favoured compared to others, giving a non-uniform distribution. The errors in the p-values shown on the graphs better illustrate the uniformity or not of the p-value distribution, as they indicate the first and third quartiles of the p-value distribution, with the primary value representing the median.

Figure E.2 shows the differences between the strains at 0, 200 and 500 generations grown at the LPC and LSM. A clear difference is difference in the fitness is apparent at 500 generations. This is reflected in the p-values, which are clustered towards zero. The LSM case at 500 generations is interesting because the median p-value is $p = 0.04$, slightly below the threshold for canonical acceptance, with an upper quartile near $p = 0.14$. While this is not as strong a separation as is seen in the LPC at 500 generations, it is worthwhile to note that the measured median p-values and interquartile ranges are robust. In the LSM at 500 generations for example, even after removing all data points with fitnesses of $F \geq 1.2$, we still recover $p = 0.04^{+0.1}_{0.03}$.

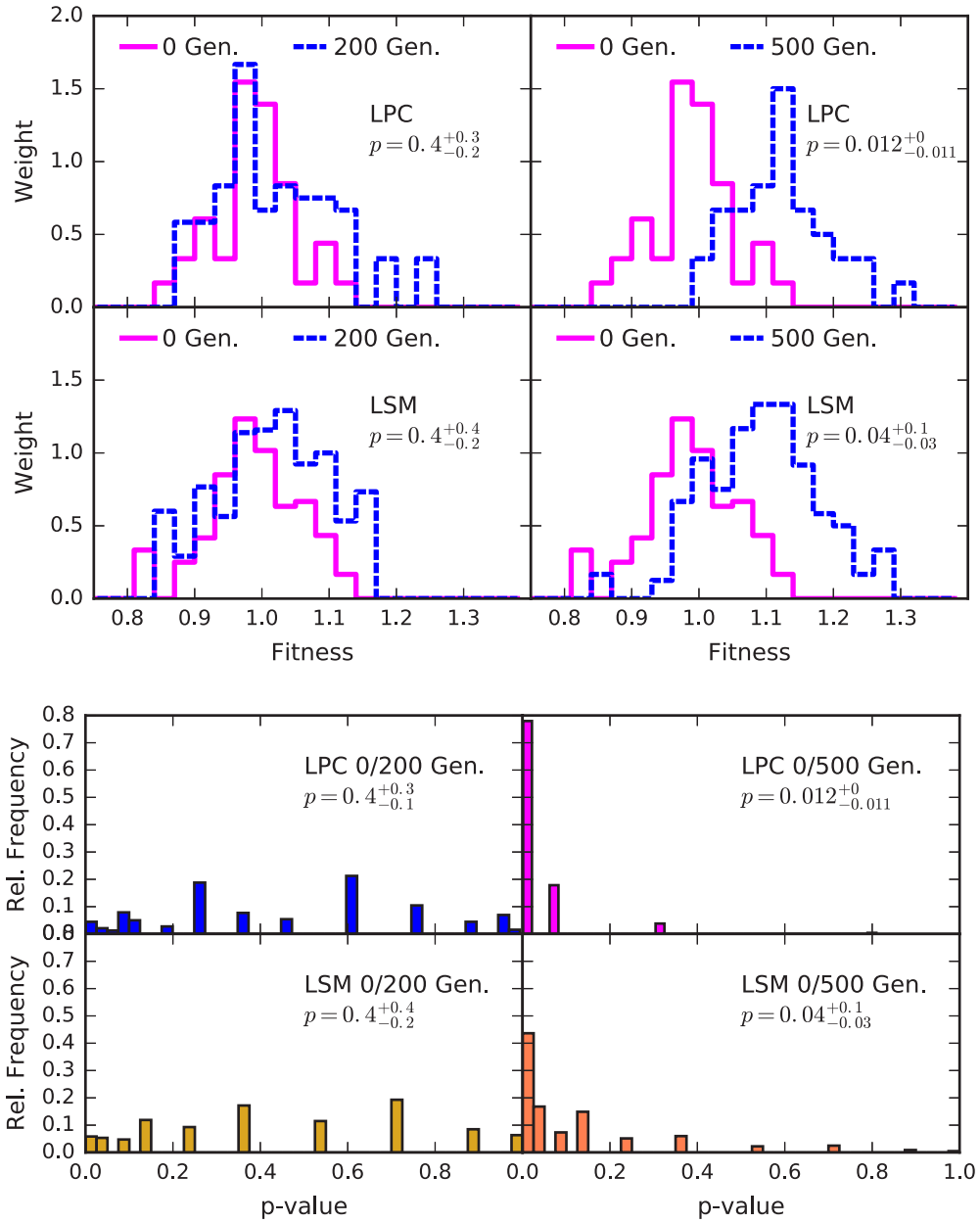


Fig. E.2 The top panel reproduces Figure 4.14, showing the likelihood that the evolved and ancestral fitnesses in each environment have changed. The distributions of p-values from these measurements, based on 2000 bootstrapped KS tests are shown in the bottom panel.