

A new Standard DNA Damage (SDD) data format

Authors

J. Schuemann¹, A.L. McNamara¹, J. W. Warmenhoven², N. T. Henthorn², K. Kirkby², M. J. Merchant², S. Ingram², H. Paganetti¹, K.D. Held¹, J. Ramos-Mendez³, B. Faddegon³, J. Perl⁴, D. T. Goodhead⁵, I. Plante⁶, H. Rabus^{7,8}, H. Nettelbeck^{7,8}, W. Friedland^{8,9}, P. Kunderát⁹, A. Ottolenghi¹⁰, G. Baiocco^{8,10}, S. Barbieri^{8,10}, M. Dingfelder¹¹, S. Incerti^{12,13}, C. Villagrasa^{8,14}, M. Bueno¹⁴, M. A. Bernal¹⁵, S. Guatelli¹⁶, D. Sakata¹⁶, J. M. C. Brown¹⁷, Z. Francis¹⁸, I. Kyriakou¹⁹, N. Lampe¹², F. Ballarini^{10,20}, M. P. Carante^{10,20}, M. Davidková²¹, V. Štěpán²¹, X. Jia²², F. A. Cucinotta²³, R. Schulte²⁴, R. D. Stewart²⁵, D. J. Carlson²⁶, S. Galer²⁷, Z. Kuncic²⁸, S. Lacombe²⁹, J. Milligan, S. H. Cho³⁰, G. Sawakuchi³⁰, T. Inaniwa³¹, T. Sato³², W. Li^{9,33}, A. V. Solov'yov³⁴, E. Surdutovich³⁵, M. Durante³⁶, K. Prise³⁷ and S. J. McMahon³⁷

¹ Department of Radiation Oncology, Massachusetts General Hospital & Harvard Medical School, Boston, MA

² Division of Cancer Sciences, The University of Manchester, Manchester, UK

³ Department of Radiation Oncology, University of California San Francisco, San Francisco, CA

⁴ SLAC National Accelerator Laboratory, Menlo Park, CA

⁵ Medical Research Council, Harwell, UK

⁶ KBRwyle, 2400 NASA Parkway, Houston, TX

⁷ Physikalisch-Technische Bundesanstalt (PTB), Braunschweig, Germany

⁸ Task Group 6.2 “Computational Micro- and Nanodosimetry”, European Radiation Dosimetry Group e.V., Neuherberg, Germany

⁹ Institute of Radiation Protection, Helmholtz Zentrum München – German Research Center for Environmental Health, Neuherberg, Germany

¹⁰ Physics Department, University of Pavia, Pavia, Italy

¹¹ Department of Physics, East Carolina University, Greenville, NC

¹² CNRS, IN2P3, CENBG, UMR 5797, F-33170 Gradignan, France

¹³ University of Bordeaux, CENBG, UMR 5797, F-33170 Gradignan, France

¹⁴ IRSN, Institut de Radioprotection et Sûreté Nucléaire. PSE-SANTE/SDOS/LDRI. BP 17, F-92262 Fontenay aux Roses Cedex, France.

¹⁵ Applied Physics Department, Gleb Wataghin Institute of Physics, State University of Campinas, Campinas, SP, Brasil

¹⁶ Centre for Medical Radiation Physics, University of Wollongong, Wollongong, NSW, Australia

¹⁷ Department of Radiation Science and Technology, Delft University of Technology, Delft, The Netherlands

¹⁸ Department of Physics, Faculty of Science, Saint Joseph University, Beirut, Lebanon

¹⁹ Medical Physics Laboratory, University of Ioannina Medical School, Ioannina, Greece

²⁰ INFN (Italian National Institute of Nuclear Physics), Section of Pavia, via Bassi 6, I-27100 Pavia, Italy

²¹ Department of Radiation Dosimetry, Nuclear Physics Institute of the CAS, Řež, Czech Republic

²² Department of Radiation Oncology, University of Texas Southwestern Medical Center, Dallas, TX

²³ Health Physics and Diagnostic Sciences, University of Nevada Las Vegas, Las Vegas, NV

²⁴ Division of Biomedical Engineering Sciences, School of Medicine, Loma Linda University, Loma Linda, CA

²⁵ Department of Radiation Oncology, University of Washington, Seattle, WA

²⁶ Department of Therapeutic Radiology, Yale University School of Medicine, New Haven, CT

- ²⁷ Medical Radiation Science Group, National Physical Laboratory, Teddington, UK
- ²⁸ School of Physics, University of Sydney, Sydney, NSW, Australia
- ²⁹ Institut des Sciences Moléculaires d'Orsay (UMR 8214) University Paris-Sud, CNRS, University Paris-Saclay, 91405 Orsay cedex, France
- ³⁰ Department of Radiation Physics & Imaging Physics, The University of Texas MD Anderson Cancer Center, Houston, Texas
- ³¹ Department of Accelerator and Medical Physics, National Institute of Radiological Sciences, Chiba, Japan
- ³² Japan Atomic Energy Agency, Nuclear Science and Engineering Center, Shirakata 2-4, Tokai, Ibaraki 319-1196, Japan
- ³³ Task Group 7.7 "Internal Micro- and Nanodosimetry", European Radiation Dosimetry Group e.V., Neuherberg, Germany
- ³⁴ MBN Research Center, Altenhöferallee 3, 60438 Frankfurt am Main, Germany
- ³⁵ Department of Physics, Oakland University, Rochester MI 48309, USA
- ³⁶ GSI Helmholtzzentrum fuer Schwerionenforschung, Biophysics Department, Darmstadt, Germany
- ³⁷ Centre for Cancer Research & Cell Biology, Queens University Belfast, Belfast, UK

* Corresponding authors:

J. Schuemann: jschuemann@mgh.harvard.edu

M. J. Merchant: michael.merchant@manchester.ac.uk

S. J. McMahon: stephen.mcmahon@qub.ac.uk

Abstract

Our understanding of radiation-induced cellular damage has greatly improved over the past few decades. Despite this progress, there are still many obstacles to fully understand how radiation interacts with biologically relevant cellular components, such as DNA, to cause observable endpoints such as cell kill. Damage in DNA is identified as a major route of cell killing. One hurdle when modeling biological effects is the difficulty in directly comparing results generated by members of different research groups. Multiple Monte Carlo codes have been developed to simulate damage induction at the DNA-scale, while at the same time various groups have developed models that describe DNA repair processes with varying levels of detail. These repair models are intrinsically linked to the damage model employed in their development, making it difficult to disentangle systematic effects in either part of the modeling chain.

These modeling chains typically consist of track-structure Monte Carlo simulations of the physics interactions creating direct damages to the DNA; followed by simulations of the production and initial reactions of chemical species causing so-called 'indirect' damages. After the DNA damage induction, DNA repair models combine the simulated damage patterns with biological models to determine the biological consequences of the damage. Currently, the effect of the environment such as molecular oxygen (normoxic versus hypoxic) is poorly considered.

We propose a new Standard DNA Damage (SDD) data format to unify the interface between the simulation of damage induction in DNA and the biological modeling of DNA repair processes, and introduce the effect of the environment (molecular oxygen or other compounds) as a flexible parameter. Such a standard greatly facilitates inter-model comparisons, providing an ideal environment to tease out model assumptions and identify persistent, underlying mechanisms. Through inter-model comparisons, this unified standard has the potential to greatly advance our understanding of the underlying mechanisms of

radiation-induced DNA damage and the resulting observable biological effects when radiation parameters and/or environmental conditions change.

1. Introduction

Cellular responses to radiation damage have been studied for many decades, showing the dependency of DNA damage on the delivered dose, the delivery timeframe, and the irradiation particle type and energy. Numerous models have been developed to explain these responses across a range of endpoints, including DNA damage, mutations, micronuclei formation, chromosome aberrations and cell survival. Many of these are phenomenological macroscopic models, and simply relate cellular endpoints to the delivered dose and empirical parameters expressing cell sensitivity, which can depend on the cell line, irradiation conditions, and radiation quality. Such phenomenological approaches can capture the overall population-based trends in cell survival that are necessary to describe the effects of radiation therapy, or to estimate effects of exposure to environmental or space radiation. The most common example is the linear quadratic (LQ) cell survival model, which is widely used both experimentally and clinically. To more systematically include the observed dependence of cell survival on the ionization pattern of the radiation modality, i.e. the particle type and energy, various models have been proposed that explicitly include additional physics properties in order to describe effects relative to a reference radiation. Some models consider the linear energy transfer (LET) (1-4) or other properties related to the structure of the primary irradiating particles and the tracks of surrounding secondary particles (the track structure) in the cell survival calculation – such as the Local Effect Model (5) and the microdosimetric kinetic model (6). The latter two models are used clinically in Carbon therapy (7-10). However, these models are also primarily phenomenological and their parameters are dependent on fitting to a selected data set, rather than being based on more fundamental radiobiology.

In order to advance the field towards more individualised therapies we must study the underlying biological mechanisms of cellular response to radiation exposure and develop effective multi-scale models of radiation action that combine physics, chemistry and biology. Efforts to model cell response focus on damages to the nuclear DNA, which has long been established as the primary radiation target determining cell viability. The response of cells to radiation has been shown to correlate with the pattern of energy depositions within the nucleus; this correlation is attributed to the resulting differences in patterns and types of DNA damage.

Several decades ago, the first studies using Monte Carlo simulations were performed to link the track structure of different radiation modalities with DNA geometries and the probability of damage induction (11-22). These studies represent the first attempts to apply track structure Monte Carlo simulations to mechanistically understand how radiation energy depositions lead to DNA damage. In an ideal scenario, one would use track structure simulations of the incident radiation to simulate the physics interactions to obtain nanometer scale energy depositions and ionizations in accurate geometric models of the cells and its sub-components (nucleus and DNA). Following the physics interactions, the resulting radiolysis products and other ionized molecules react in a physicochemical stage, which is followed by migration of the chemical species. At this stage chemical species can react with each other, be scavenged inside the cells, or react with components of the cell, such as the DNA. The simulation finishes by determining the direct (caused by physical interactions) and indirect (caused by chemical reactions) damages to DNA. Finally, the DNA damage patterns can be used in mechanistic models of DNA repair kinetics to calculate cell viability, accounting for the damage complexity, along with properties of the cell and the surrounding environment, such as repair deficiencies, cell cycle and oxygenation.

In recent years, several major developments have led to a surge in attempts to mechanistically describe

DNA damage and repair kinetics (23-47). An increase in the computational power of standard computers has enabled the simulation of particle tracks in DNA fragments and even whole nuclei (48-53). This has been accompanied by improvements in imaging techniques for studying the responses of cells to ionizing radiation, providing an abundance of data showing the importance of repair pathways and their effect on cell viability. Currently, several Monte Carlo simulation codes exist that can provide the nanoscale track structure of particles passing through a medium, which is typically simulated as water but more recently also includes DNA nucleotide material (23,54-60). Codes like Geant4-DNA (23,61), KURBUC (62), PARTRAC (60), MC4 (63), RITRACK (57), TRAX (64), and TOPAS-nBio (65) are used by various groups to simulate the track-structure of different types of radiation and then score the resulting initial damages to a cell nucleus. One can even go further down in scale and use a molecular dynamics approach such as in the MBN Explorer (66) to model the interactions between molecules. Increasing the simulation detail increases the complexity and makes simulations computationally more expensive. Thus, researchers often apply a multi-scale approach, using macroscopic simulations to capture realistic radiation fields for radiation therapy or mixed field exposures, and switching to the cell-scale simulations of particle tracks scoring DNA damages in selected cells. This approach can be achieved using a common simulation framework (e.g. Geant4/Geant4-DNA, TOPAS/TOPAS-nBio (67), or the MultiScale Approach (68-70)), or with the ad hoc coupling of different transport and track structure codes (e.g. PHITS and PARTRAC (71)).

Each of these codes includes models of DNA structures within the nucleus or cell that are used to obtain the initial patterns of DNA damage. Most of these codes also include the first chemical reactions, i.e. the physicochemical generation of radiolysis products and their subsequent diffusion and interaction (72-77). Thus, these Monte Carlo codes can provide estimates of DNA damages induced both directly (from the initial particle track) and indirectly (from chemical reactions). While these codes frequently differ in their underlying assumptions, have slightly different implementations of particle transport and physics handling and developed their own data structure and damage pattern definitions, complicating inter-code comparisons of damage induction, most track structure codes predict reasonably similar yields of double strand breaks (DSBs), in part because the number of DSBs is often used as a reference dataset.

In order to fully describe the impact of DNA damage induction and repair on cell survival, chromosome aberrations, mutations or other endpoints of interest, the simulated patterns of damage along the DNA strands, as well as their complexity, must then be combined with models that describe the mechanisms of DNA repair (78). Various groups are working on models to describe these DNA repair kinetics, and to better understand the dependencies of predicted endpoints on uncertainties and assumptions made in each part of this modeling chain, a direct comparison between models and simulation results from different groups would be immensely useful. However, because of the differences in damage model outputs and dependencies among different damage and repair models, these comparisons are arduous and complex. Repair modeling approaches typically either use an assumed (often random) distribution of damage or are designed specifically to interface with one of the available track structure codes in an *ad hoc* fashion (30,31,37,38,79-83). While typically based on similar principles, these models often employ different approaches and make different assumptions about the underlying repair processes. Inter-comparison between repair models is often complicated by their close links with underlying damage models, which introduces implicit assumptions and dependencies that may not be apparent on simple inspection (84). Providing a common interface would offer much more flexibility and scope to testing different combinations of models, and comparing implicit assumptions and uncertainties.

In addition, there can be important differences among communities in the way DNA damages are defined. Examples include differences in what constitutes a single strand break (SSB) or a double strand break (DSB) and how they are categorized into different lesion complexities; what factors are considered when describing the nuclear environment; or at what time point the damages are recorded(85). Providing a standard data format that all groups can refer to will help to highlight these differences among the groups and disciplines and provide a platform with which to reconcile them.

In this paper, we propose such a new Standard for DNA Damage (SDD) to facilitate cross-comparisons between the various track-structure Monte Carlo codes and their implementations of first chemical reactions within the cell nucleus, and to link these to mechanistic models of cell repair and the kinetics of DNA damage repair. The proposed standard data format, illustrated in Figure 1, provides a new method for cross-code comparisons and promotes collaborations among groups by promoting sharing of DNA damage patterns at selected stages in time, i.e. after the initial energy depositions (direct damages) or after the chemical stage (including indirect damages), as input to calculate the biological endpoint(s) of interest. By developing a standard data format that various codes can write or read, we provide the means to not only compare the results of different codes and models, but also investigate the influence of each model assumption and cross validate between models. Testing the dependencies of various observable outcomes on model parameters and their implementation in different models can help us to understand which parts of the models are most sensitive and which parts have only a minor effect on the outcome. In combination with new experimental data of repair processes, in particular with higher temporal or spatial resolution from new microscopy technologies, this can further help to test the models at various stages along the repair process and identify key experiments to advance the field of radiation biology research.

This standard is primarily designed to collect nuclear damage information for eukaryotic cells after radiation damage. However, one can also apply the standard to other sources of DNA damage, e.g. from chemotherapeutic drugs, or to any other organism with DNA, such as bacterial/viral DNA damage (86,87). In that case, some of the cell specific information listed in the standard may be omitted. We indicate in some fields where bacterial/viral information can be used instead. While non-nuclear damages can also result in cells becoming non-viable, the proposed standard focuses on the main pathway of cell damage, i.e. damage to the DNA, to provide a compact and easily transferable format.

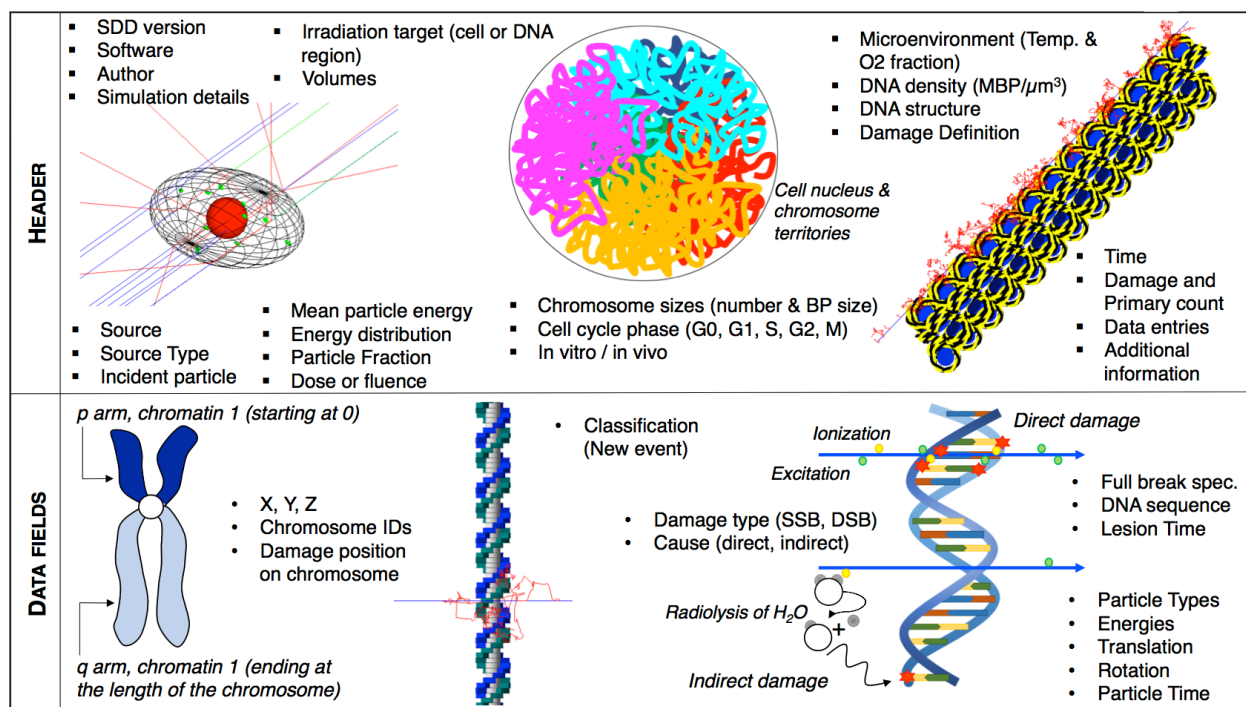


Figure 1: Illustration of the header and data structure of the proposed Standard for DNA Damage (SDD). The information common to all recorded damages is listed in the header and the information relevant to each damage is recorded in the data section of the SDD file.

2. The new Standard for DNA Damage

The data format for the proposed SDD is based on the format of a typical tuple, i.e. a finite ordered list or sequence of elements. The file format for each damage specification consists of two sections combined in one file that should have the suffix “.sdd”, for example Filename.sdd. The two sections are:

1. A header consisting of a series of factors common to all damage sites in the data block.
2. A series of fields defining individual damage sites within the modelled volume.

Our method intends to offer a standard suitable to accommodate a wide range of underlying simulations and DNA repair model designs. To achieve this, the header requires some basic information about the recorded damage patterns for automated read-in of standardized data, while at the same time providing free text sections to expand on the details of the simulation tools. For wide-spread readability, SDD-files employ a comma-separated value format in the header section with each field ending with a semicolon. For the data section, values are separated by a comma or forward-slash and a semicolon is used to indicate a new field. These are the only field separators. Spaces and new lines should be ignored by SDD readers. However, for better readability, we strongly recommend to start a new line for each field.

SDD files are written entirely as plain text (UTF-8 encoding). Due to the variable size of the damage definitions and the sparsity of the data even for radiation exposures of several Gy, a binary format for the data section is considered unnecessary.

2.1. Website and Updates

The SDD data format anticipates that with increasing use cases, numbering schemes will need to be expanded to define additional details or options in some fields. In order to keep the numbering scheme unique and continue to allow users to share their SDD files without ambiguity, we recommend that requests for new numbering schemes are submitted to the SDD collaboration (represented by the authors of this manuscript, headed by the groups at Massachusetts General Hospital / Harvard Medical School, University of Manchester and Queen's University Belfast) following the steps detailed on the SDD collaboration website <http://standard-for-dna-damage.readthedocs.org/>. Each new specification for fields in the header or data block will be assigned a specified number and documentation about all fields will be provided and updated.

2.2. The Header Section

The header provides information defining the conditions common to all entries in the data section. The structure of the header is presented in table 1. The header is designed to offer comprehensive information for a large variety of possible damage simulations. A side-effect of this flexibility is that many simulations may not be able to fill all header fields. However, we strongly recommend including all information available. In order to ensure reproducibility when sharing SDD files, the header contains information on the modelled geometry as well as the irradiation that caused the DNA damage. While the design of the SDD format and the description below is focussed on radiation-induced damage, the data structure is flexible enough to allow scoring of other sources of DNA damage, for example from chemotherapeutic drugs. In that case, some of the fields in the header have no value provided and the damage induction may be described in the additional information fields.

Comments can be added to provide additional information on the irradiation, simulation or modeling details. They should be denoted by a block of text, e.g. a new line, starting and ending with #. Any text between these characters should be ignored by the reader codes. Comment lines can be inserted anywhere in the header.

FIELD	VALUE	NOTES	TYPE
1	SDD version	Version number of SDD definition	String ("SDDv1.0")
2	Software	Program name, version and access link if any	String (Free text)
3	Author	Corresponding author, date, references	String (Free text)
4	Simulation details	Description of details of simulation settings and parameters	String (Free text)
5	Source	Description of source properties	String (Free text)
6	Source Type	Monoenergetic, distribution, phase space, GCR, ...	Int
7	Incident particles	Definition of primary incident irradiation particle(s) in PDG code format	Int(s)
8	Mean Particle Energy	Mean incident energy for each particle in MeV	Float(s)
9	Energy Distribution	Full energy distribution specification	String(s) + Floats
10	Particle fraction	Fraction of fluence of each particle in field	Float(s)
11	Dose or fluence	Define dose or fluence in each exposure, or note that the simulation was for a single track	Int + Float (+Float)
12	Dose Rate	Dose rate of irradiation field	Float
13	Irradiation Target	Description of simulated cell or target (DNA) region and microenvironment	String (Free text)

14	Volumes	Shape parameter plus X,Y,Z extents (μm)	2x (Int + 6 Float)
15	Chromosome sizes	Number and base pair size of chromosomes	Int + (Int) Floats
16	DNA Density	Density of base pairs in volume ($\text{MBP}/\mu\text{m}^3$)	Float
17	Cell Cycle Phase	Cell cycle phase index and progression	Int + Float
18	DNA structure	Additional field to define DNA structure	2 Ints
19	In vitro / in vivo	Experimental condition	Int
20	Proliferation status	Proliferative or quiescent and status details	Int + String (Free text)
21	Microenvironment	Temperature ($^{\circ}\text{C}$) and molar O_2 concentration	2 Float
22	Damage definition	Define how types of damage were determined	1 Float + 1 Int + 1 Bool + 2 Float
23	Time	Time point at which damages were recorded	Float
24	Damage and primary count	Number of distinct damage lesions scored & primaries simulated	2 Int
25	Data Entries	Number of fields included in the data section	14 Bool
26	Additional Information	Field for additional information that may be relevant	String (Free text)
27	***EndOfHeader***	Empty field to mark end of header.	

Table 1: Field by field summary of the header fields and their type. “Bool” stands for entries that can be either 0 or 1. “Int” stands for parameters that enumerate different possibilities and can take values representing integer numbers between 0 and a maximum value that depends on the field as detailed in section 2.1.1. Entries designated by “Float” are decimal strings to be assigned to floating point variables when reading the SDD file content by a computer program.

2.1.1 Description of header fields

The header consists of 27 fields, each ending with a semicolon. For better readability, each field can be started on a new line, although this is not required per se. Table 1 summarizes the proposed fields and their format, additional details for each field are provided below. Each field starts with a string including the “value” tag in the table followed by a comma followed by the types defined in table 1. If the information for any field is not available, the value string and the field-ending semicolon should still be included in the header. Accordingly, free text sections should not include semicolons except to end the field.

Field 1, SDD version: SDD version number to allow tracking future modifications of the file structure and to enable automatic transformation of the information in the header and data after such modifications. The version detailed here is SDDv1.0. Thus, the first field should read: “SDD version, SDDv1.0;”.

Field 2, Software: Here the program name and version number that were used to obtain the DNA damage are described. This can be anything from a simple random sampling function to a combination of dedicated Monte Carlo codes. Due to the free text format, additional information about the software such as an access link (URL) to the software if available can be added.

Field 3, Author: Lists the corresponding author of the simulations to allow for communication about the data provided and the date of the file creation. The recommended minimum information is name, email address and date (comma separated). Additional references to publications relevant to the simulations can be listed here, recommended format: First author (et al), Title, Journal, edition, page, year, DOI. If

multiple references are included each reference should be separated by a forward-slash.

Field 4, Simulation details: Free text to describe simulation details, ideally providing sufficient information to potentially produce a similar simulation setup. For example, this field should include information about the physics settings, e.g. which secondary particles are included in the simulations with their respective energy cut-off or the names of the cross-section models, where relevant. Also, specifications of the world and transport medium corresponding to the interaction cross section models (e.g., liquid water, vapour water, DNA-like material) can be supplied.

Field 5, Source: Free text to describe the particle source used for the simulation. Particularly for scenarios that include multiple particle irradiations, use phase spaces or other functional forms such as galactic cosmic rays (GCR), this field can be used to add references describing the source following the structure of field 3, or to provide the URL of a website that defines the source. Additional information relevant to the source that is not covered by the structured data in fields 6-12 should also be added here. Each piece of information should be comma separated.

Field 6, Source Type: Integer giving a first overview of the incident particle source, defining the source as (1) single or multiple monoenergetic particles, (2) single or multiple particles with energy distributions, (3) a phase space source, or (4) GCRs. For cases 3&4, the source should be described in field 5, or users may need to contact the author (field 3) for full source definitions. In addition, for these two options, fields 7-10 may be insufficient and can be left blank, however, if these fields can be used to (roughly) describe the particle distributions, we suggest adding the information. Suggestions for additional options can be submitted to the collaboration website.

Field 7, Incident Particles: Defines the radiation type(s) of the incident particle(s), using the particle specification by the [Particle Data Group](#) (PDG)(88) to provide flexibility and a comprehensive handling of all known particle types, including (charged) ions and excited states of ions. The radiation source can be an external beam or radionuclides. Each incident particle type can be fully described by a single PDG code (integer). This field lists all incident particle types in the same order that further source definitions should be provided in subsequent fields. Each particle type is comma-separated. Resulting chemical species, using for example PubChem IDs, are not included at this point as chemical species typically are created as a result of the irradiation, i.e. from the primary particle.

Field 8, Mean particle energy: Lists the mean incident particle energy in units of MeV as a single float for each particle type listed in field 7 following the same order.

Field 9, Energy distribution: Further specifies the energy distribution of each incident particle in the same order as listed in field 7. For monoenergetic beams (indicated in field 6), this field should take the form "M, 0" for each particle defined in field 7. For other beams, the expected format is a letter specifying distribution ("G" for Gaussian, "B" for bifurcated Gaussian) followed by a comma and a series of forward-slash separated distribution parameters. The mean, μ , is given by field 8. Values required are the variance (G), and left and right variance (B). This field should define one set of parameters for each particle type, using comma separation. For other source formats such as user defined distributions or distributions from data tables or for radionuclides, the spectrum should be defined either using these functions or the free text section (field 5), where a link (URL) to a website can be included. Alternatively, users may need to contact the author (field 3), or submit suggestions for additional options of functional forms to be included to the collaboration website.

Field 10, Particle Fraction: Defines the fraction of the fluence represented by each particle type, as a single comma separated number per particle type (defined in field 7, same order).

Field 11, Dose or Fluence: The field contains first an integer specifying whether each field in the data block is for a single track irradiation (0), a delivered dose (1), or a fluence (2). For the latter two options, the second entry is a float given in Gy for dose, or particles per μm^2 for fluence. A third value can be added to provide the standard deviation of the mean averaged dose or fluence for multiple exposures. For a single track, the field reduces to “Dose or Fluence, 0;”.

Field 12, Dose Rate: Lists the dose rate in Gy/min. This field provides an easy distinction between space radiation and other low dose scenarios that can be treated as separate events per incident particle, radiation therapy treatments (in the order of 1 Gy/min), and high dose rate deliveries including FLASH therapy and microbeam or grid therapy (> 300 Gy/min) (44,89-92).

Field 13, Irradiation Target: Free text providing a detailed description of the irradiation target: the cell type, size, cell cycle stage and other properties relevant to the damage induction; size of the nucleus or sub-nuclear region simulated; other geometrical features like mitochondria; and the potential presence of additional factors, such as nanoparticles for radio-enhancement or chemotherapeutic drugs. In case of bacterial/viral or mitochondria irradiations, their DNA content can be defined here. Similar to the free text field for the source (field 5), this field should contain information that is not captured by the structured data in fields 14-21.

Field 14, Volumes: Defines the extent of the simulation volume (i.e. the simulated world), and the relevant scoring volume using two sets of comma-separated lists of an integer and 6 floats. For both volumes, the integer defines the shape of the bounding volume, such as the cell, as either a box (0), an ellipsoid (1), or a cylinder (2); other volume shapes can be added by submitting a request to the SDD collaboration website to extend the SDD by assigning higher value integers. The shape definition is followed by three floats in the order X, Y, then Z, specifying the bounds of the volume in μm . For a box, the values are given in half lengths, i.e. from (+X,+Y,+Z) to (-X, -Y, -Z); for an ellipsoid, the floats define the half axes of the ellipsoid along each of these three axes, i.e. for the special case of a spherical bounding volume, X, Y, Z are identical; for a cylinder, X and Y define the half axes of the ellipsoid along these axes, and Z defines the half length of the cylinder extent (from +Z to -Z). The bounding box thereby also defines the origin of the coordinate system as the center of the bounding box (i.e. the center of the nucleus or cell). The second group of 3 floats defines Euler rotations, ϕ , θ , ψ , respectively, to allow orienting the target in space according to the simulation setup.

The second set of volume definitions (int + 6 floats) follows the same rules as above and defines the scoring volume, e.g. the nucleus. If both volumes are identical, only one has to be defined.

Field 15, Chromosome sizes: Lists the number N of chromosomes in the nucleus (or in bacteria/virus), followed by N floats for the size of the chromosome in mega base pairs (MBPs). The order of chromosomes listed here should be consistent with the chromosome ID used in field 3 of the data block. Each chromosome should be listed, i.e. a total of 46 for a normal human cell. This allows for the inclusion of cells with missing or multiploid chromosomes. Optionally, if only N is provided, the chromosomes are assumed to be uniform in size based on the density stored in field 15.

Field 16, DNA density: Describes the density of the DNA base pairs in the scoring volume in units of mega base pairs (MBP) per μm^3 as a single float value. Here the average density over the entire scoring

volume is considered, i.e. an average of heterochromatin and euchromatin regions if both are present.

Field 17, Cell Cycle Phase: Defines the cell cycle and the progression through the phase using an integer and a float. The integer defines the cell cycle numerically as G0 (1), G1 (2), S (3), G2 (4), and M (5). Progression through a phase can be denoted by providing an additional (comma separated) float with value between 0 and 1 – for example, “3, 0.7;” indicates a cell 70% of the way through S phase. This optional float is included to allow more granular inclusion of asynchronous cell populations.

For simulations without a specific cell cycle phase, the value can be set to 0. The cell cycle phase is important to determine the presence of sister chromatids. It further influences the number of chromosome base pairs (BPs) listed in field 14; for cells in (late) S or G2, the number of BPs in a chromosome should only be half the total number of BPs as they are repeated and identified by their chromatid number (CR) in field 3 of the data block.

It should be noted that the DNA damage format is designed assuming that each file records responses in a single defined cell type at a particular point in the cell cycle. However, in in vitro or in vivo experiments or clinical treatment, the cell population being exposed is typically heterogeneous, with only features such as the composition of the cell population being available, e.g. what fraction of cells are in a given cell cycle phase. Thus, in order to fully describe biological experiments, it may be necessary to assemble a representative cell population from simulations of different cell cycle stages into a population-level response, with each particular condition stored as an individual SDD file. One can then represent any mixture of cell populations by an adequate assembly of SDD files/scenarios.

Field 18, DNA structure: Defines the DNA structure by two comma-separated integers. The first integer defines the arrangement of DNA as whole nucleus (0), a heterochromatin region (1), euchromatin region (2), a mixed (heterochromatin and euchromatin) region (3), single DNA fiber (4), DNA wrapped around a single histone (5), DNA plasmid (6), or a simple circular (7) or straight (8) DNA section. Details about higher order DNA assumptions can be added as descriptions in fields 4 or 24, for example by providing the URL of a website or by referring the reader to the author (field 3).

To facilitate cross-code comparisons, options for specially defined geometries are also available. The currently defined reference geometries are a straight DNA section (100), a circular DNA plasmid (101) and a chromatin fiber (102). For exact definitions of these geometries please refer to the SDD website. The second integer indicates “naked” (0) or wet (1) DNA.

Additional values can be added and described by submitting a request to the SDD collaboration website.

Field 19, In vitro / in vivo: Describes the experimental conditions that are simulated. This field is important for both the geometry setup and for considerations of biological response. The condition is defined by two comma-separated integers, the first integer defines if the simulations refer to in vitro (0) or in vivo (1) conditions, the second integer further explains the conditions; it should be 0 for in vivo experiments, (1) for monolayers of cells, (2) for cell suspensions, (3) for 3D grown tissue models. Additional conditions can be added by submitting requests to the website.

Field 20, Proliferation status: Integer variable to determine the proliferation state of the cell(s) as quiescent (0) or proliferating (1). A second optional string (free text) can be added to describe the status of the scenario, including environmental cues like serum starvation and innate states like stemness.

Field 21, Microenvironment: Contains two floats, the first value defines the temperature in degrees Celsius, the second the molar oxygen (O₂) concentration in the volume in molarity (M). If no values are pro-

vided, a standard room temperature of 25°C and normoxic conditions are assumed. Other relevant concentrations such as the concentration of various scavengers should be defined in the free text format of field 13; they are not included due to the wide range of potential scavenging agents. Potential additional fractions or other microenvironment factors can be added by sending a request for expanding the number of parameters given in this field to the SDD collaboration website.

Field 22, Damage definition: This field defines how damage is scored and accumulated into distinct damage sites in the data block. It consists of a list of the following values using comma separation:

1) Integer to define if damages were recorded as those resulting from direct effects only (0) or including chemistry (1). Other types are not currently explicitly included but can be defined by sending a corresponding request to the SDD collaboration website.

2) A Boolean flag to define if the following numbers are listed in number of BPs (0) or in nm (1).

3) This value sets the distance in base pairs or nm between backbone lesions that are considered double strand breaks (float).

4) If this value is set to -1, it indicates that base lesions are not scored. Non-negative values mean that damages to the bases add to the damage complexity and are stored in the data block. In that case, all base damages between backbone damages that form a DSB are stored. This value then determines the distance (in BP or nm) beyond the outer backbone damages where base damages are also stored in the same site (float).

5) Low energy threshold to induce a strand break (or base damage) in eV (float).

6) Optional field to define a linear probability function for damage induction as used in PARTRAC, with the probability $p(E < A) = 0$, $p(E > B) = 1$, and linearly increasing probability from 0 to 1 in the interval from A to B, where A is defined by the 5th value in field 22, and B is given in this field in eV (float).

Note: This field will influence the full break specification in the data part of the standard, as demonstrated also in figure 2. Fields 22.1, 22.5 and 22.6 influence which interactions are scored as damages, field 22.3 and 22.4 determine the distances between and around damages that are clustered in a single break record. However, together with the chromosome position (fields 3 and 4 in the data block), the data block can be post-processed to yield new break clustering using different distances as desired.

An example of field 22 would look like:

“Damage Definition, 0, 0, 10, 3, 17.5;”

Translating to: Only counting lesions from direct track interactions, distances are defined in number of BPs, a distance of 10 BP to call two opposite strand SSBs a DSB, base damages are considered, grouping base damages up to 3 BPs on either side of backbone damages in a single site, and only interactions depositing at least 17.5 eV are counted as lesions.

Field 23, Time: This specifies the total simulation time for each primary particle - that is, the time from when the source particle was created to the time at which the chemistry simulation ends, i.e. when the damage was recorded, in nanoseconds. For simulations that only consider direct (physics) interactions, this value should be set to 0.

Field 24, Damage and primary count: The first integer records the number of distinct damage lesions scored as a single integer and should be identical to the number of fields in the data block divided by the number of fields per damage site (sum of “true” values of field 25). The second integer is a counter of how many primary particles were simulated. This value is important to count particles that did not cause any damage to the DNA to accurately represent the probability of interactions and avoid overestimation of damage induction.

Field 25, Data entries: An array of 14 comma-separated Booleans to indicate which fields of the data block are filled. This field facilitates SDD-reader interfaces.

Field 26, Additional Information: Allows for additional comments about the simulation that may be relevant for the scored damages. This can, for example, include further details on the physics settings, simulated geometries, material compositions, the source, potential scavenger concentrations in the cell, or other descriptions of the simulation or irradiated target that may be helpful to better understand the simulations or improve the biological modeling.

This field can also be used to define new user-specified values for any field in the header or data block, however, we strongly recommend to rather submit a request to update the standard with such new settings to the SDD collaboration website (<http://standard-for-dna-damage.readthedocs.org/>) so that the new settings can be officially included in the standard to ensure that all users use the same uniquely defined values.

Field 27, *EndOfHeader***:** An empty field, used to denote the end of the header and beginning of main text. This field ends the header with: "***EndOfHeader***;"

2.3 The Data Section

The data block is recorded in text (UTF-8) format. Each damage site is stored as a group of up to 14 fields, each containing a series of comma- and/or forward-slash-separated fields to define the structure of the damage. Each field will end with a semicolon to indicate the start of the next field in the data block. Some fields are required to identify the original (primary) particle incident on the cell, the position and type of the damage; other fields are optional to provide additional information. Many fields are optional and may be omitted, with the fields used in a particular file defined in the header field 25. We acknowledge that most codes are currently not designed to supply all data fields; we consider the structure also as a motivation for future developments to improve reporting of relevant details. In general, if the information is available in a simulation, it should be added and all optional fields can be filled to increase the value of the data. The data structure is summarized in table 2 and detailed below.

FIELD	VALUE	NOTES	TYPE	REQ?
1	Classification	Is damage associated with new primary particle or new exposure? And event ID.	2 Ints	Y,N
2	X,Y,Z	Spatial X, Y, Z coordinates and extent (μm)	3x3 Floats	*
3	Chromosome IDs	ID of chromosome/chromatid where damage occurred and on which arm (long/short) or specification of non-nuclear DNA type.	4 Int	*
4	Chromosome Pos.	Location of damage within chromosome	Float	*
5	Cause	Cause of damage - direct or indirect and number	3 Int	N
6	Damage Types	Types of damage at site (Base damage, SSB, DSB)	3 Int	**
7	Full Break Spec	Full description of strand break structure	Special	**
8	DNA Sequence	DNA Base Sequence around break site	Special	N
9	Lesion Time	Time of each damage induction (in ns)	Special	N
10	Particle Types	PDG list of particles	Int(s)	N
11	Energies	List of kinetic energies for each particle (in MeV)	Float(s)	N
12	Translation	Starting position of each particle (in μm)	Floats	N
13	Direction	Starting direction of each particle (unit vector)	Floats	N
14	Particle Time	Starting time of each particle (in ns)	Floats	N

Table 2: Value by value definition of the data fields to score DNA damages. Of the fields indicated with “” either field 2 or 3&4 are required, similarly at least one of the “**” fields is required. “Int” stands for parameters that enumerate different possibilities and can take values representing integer numbers between 0 and a maximum value that depends on the field as detailed in section 2.1.1. Entries designated by “Float” are decimal strings to be assigned to floating point variables when reading the SDD file content by a computer program.*

Field 1, Classification: The first integer of the first field in the data block identifies each damage site as a damage from a new exposure, a damage by a new primary particle in the same exposure, or another damage from an already recorded primary particle. A new ‘exposure’ here means a separate set of the radiation dose or fluence defined in field 11 of the header (e.g. 1.8 Gy) and allows multiple instances of the same irradiation conditions to be recorded in the same file. In that case, the data block contains multiple instances of the same irradiation scheme; for example for an exposure of 1.8 Gy, a new exposure flag would be set every time the total dose in the target reaches 1.8 Gy. For example, a total of 42 simulations of 1.8 Gy could be recorded in one file, equivalent to a total simulated dose (in fractions) of 75.6 Gy. Similarly, a new ‘event’ refers to damages created by a new (single) primary particle. A new exposure flag also means a new event started. If a particle induces multiple damage sites, for example for particle irradiations with high LET, this flag is set to 0 for the second and subsequent damages, indicating damages were caused by the same single particle.

The first value of this field is defined by an integer with:

0: for a damage caused by the same primary particle as the previous row,

1: for a damage caused by a new primary particle within the same (user defined) exposure, and

2: for a damage which represents the start of a new exposure (which is also necessarily a new primary particle).

The second integer is optional and offers a place to add the event ID, i.e. the number of the primary particle that was simulated, typically counting from 0 or 1 for each exposure.

Field 2, X, Y, Z (*): Defines the spatial position X, Y, Z of the center and extent of each recorded damage, using coordinates within the bounding box specified by field 14 in the header. The first 3 values define positions specified as 3 comma-separated values with unit μm . All following fields are optional but should be included if available. The second set of 3 comma-separated values defines the maximal position value in X, in Y, and in Z and the last 3 comma-separated values list the minimal values of X, Y, Z, respectively, together defining a box that encompasses the damage. Each 3-tuple of values is separated by a forward-slash, for example field 2 could read “0.002, 0, 1.2 / 0.004, 0.002, 1.122 / 0.001, -0.001, 1.117;”.

*Either field 2 or fields 3&4 (Chromosome IDs and Position) have to be provided. While both should be listed if possible, the option to define either acknowledges the fact that, depending on the code, not all information may be available.

Field 3, Chromosome IDs (*): Stores the identity of the chromatid where the damage occurs. The entry consists of 4 integers. The first integer defines the ‘DNA structure’ as unspecified (0), hetero- (1) or euchromatin (2) regions of nuclear DNA, a free DNA fragment (3) or mitochondrial/bacterial/viral DNA (4). In the case of nuclear DNA, the next 3 integers are the chromosome and chromatid number and indication of long/short arm. The values are stored comma-separated as “CH, CR, CA;” where CH is the chromosome number, CR is the chromatid number and CA is the arm of the chromosome (short (0) or long

(1)). CR is specified as 1 for unduplicated chromosomes, and 1 or 2 to identify the two chromatids in the duplicated chromosome in later S and G2 (and early M) phases. For example, '12, 1, 1;' corresponds to the long arm of chromatid 1 on chromosome 12. Chromosome numbering is assumed to follow the order listed in header field 15. For cells without a specified cell phase or cells in G0 or G1 phase the chromatid number CR is always 1. In cases where the CA information of short vs. long arm is not available, the last number may be left empty.

*Either this value together with field 4 or the X, Y, Z information (field 2) is required.

Field 4, Chromosome Position (*): Defines the damage position along the chromosome's genetic length. This value is defined as the distance along the chromosome from the start of the short (p) arm towards the end of the long (q) arm. It can be stored either as a value between 0 and 1 (excluding 1) giving the fractional distance along the chromosome at which the break occurs, or, if the value is greater or equal to 1, as the distance in base pairs from the beginning of the short arm (p) to the damage site. In case of non-nuclear DNA, such as DNA fragments or mitochondrial, bacterial or viral DNA, the fraction simply refers to the size of DNA segment provided in the header or, if the value is greater than 1, the BP number along the defined DNA.

*Either this value together with field 3 or the X, Y, Z information (field 2) is required.

Field 5, Cause (optional): Offers a flag to identify the cause of the induced damage and a counter for how many damages were caused by direct or indirect events. The first integer classifies the damage type; currently included are options to identify whether the damage is a result of direct physical interactions (0), indirect interactions, i.e. the result of the propagation of any chemical species and following reactions with the DNA (1), caused by a combination of direct and indirect interactions (2), or caused by charge migration (3). Additional options can be included according to the needs of other codes, for example to represent damages induced by concomitant drug-based therapies. If additional values of this specifier are needed, a request to update the standard should be submitted to the SDD collaboration website. The second and third integers provide counters for the number of direct and indirect damages at the site, respectively.

Additional information about the damages, for example which damage was induced by which process, and more specification of the indirect damages (e.g. fixation by OH, stabilization of $R^\bullet$ by O_2 or O_2^-) can be recorded in field 7.

Field 6, Damage types ():** Provides a high-level specification of the type of damage present at a given site in terms of base damages, backbone (single strand) and double strand breaks (defined as exactly two single strand damages within the separations defined in the header), or a combination of these. This classification can be seen to be a numerical description of many other damage classification metrics(21,93), which effectively groups these damages into broader categories according to the expected biological severity of the damage. Damages separated by less than the minimum distance of BPs defined by the damage definition in the header (field 22.4) are scored in a single data block, i.e. are considered to be a single cluster of damages. Repair codes can either convert these clusters to a lesion or use the information in field 7 (if provided) to define lesions. An example of how lesions are grouped based on the information provided by field 22 in the header is shown in Figure 2.

The damages are stored as 3 comma-separated integers, where the first integer lists the number of base damages, the second the total number of single backbone breaks including those contributing to the formation of a DSB, and the last number is a binary (0 or 1) indicating the presence of a double strand

break, i.e. if lesions occurred on both backbones within the BP range defined in the header. For example, '3, 2, 1;' would represent a damaged DNA site consisting of 3 base damages with 2 backbone damages that are on opposing strands within the BP limit and thus are counted as a double strand break. Additional examples are listed in Figure 4.

**Either field 6 or 7 are mandatory, but if available, the full damage structure should always be included as detailed below to provide more details of the break structure. This field is intended to provide a high-level summary and support models that do not calculate the full structure of individual breaks and rather rely on numbers of SSBs and DSBs and their distribution.

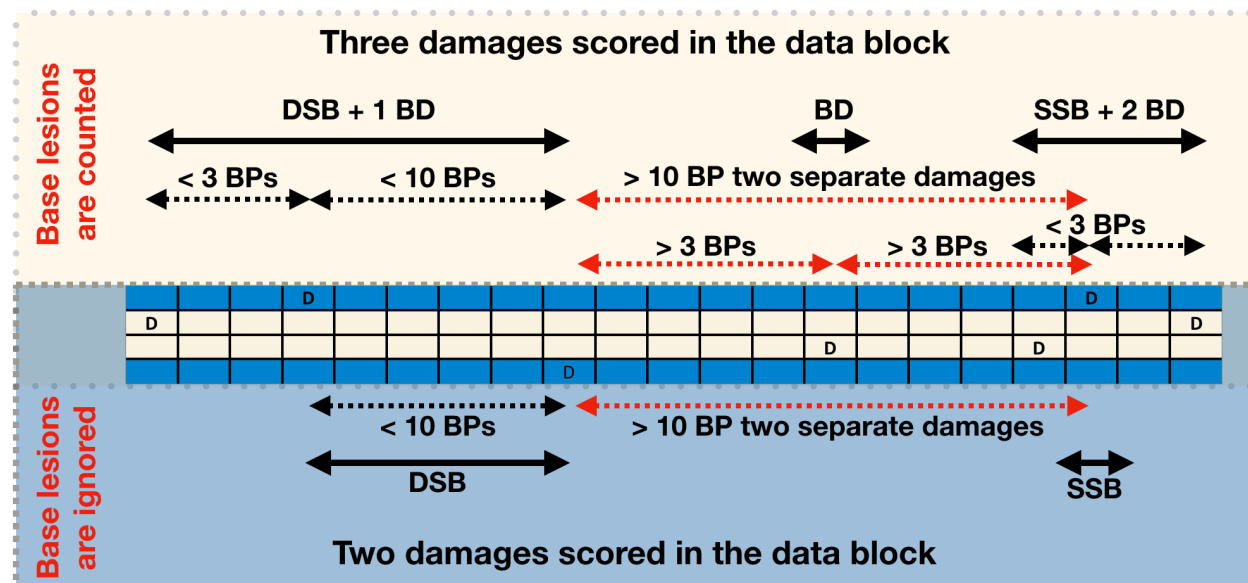


Figure 2. Example of a single DSB recorded with a 10 BP maximum backbone separation. The top half scores damages if base lesions are counted with a BP separation of up to 3 BPs as defined in the header field 22.4, scoring three entries in the data block as indicated by the solid arrows. The entries in field 6 would be a DSB "1, 2, 1;" a BD "1, 0, 0;" and a SSB "2, 1, 0;". If base damages are neglected (lower half), the same damage pattern will be scored as two separate damages: a DSB "0, 2, 1;" and a SSB "0, 1, 0;". The dashed lines demonstrate the separations considered for grouping, red indicates the distances being larger than the cutoff.

Field 7, Full break spec (**):

This field allows for a full specification of the structure of the damage. We apply a four-strand structure, using a 4 x N array, with the rows consisting of the backbone (row 1) and bases (row 2) of the 5' to 3' strand, and the bases (row 3) and backbone (row 4) of the 3' to 5' strand. The base positions (columns) are aligned reading from the short arm (p) towards the long arm (q), beginning from base 1 which is defined as the first involved alteration at the position corresponding to field 4 (if provided). Thus, increasing columns for strand 1 corresponds to the 5' to 3' direction and the 3' to 5' on strand 2. The design of the data structure is illustrated in Figure 3, with blue fields corresponding to the backbones and light orange to the base pair fields.

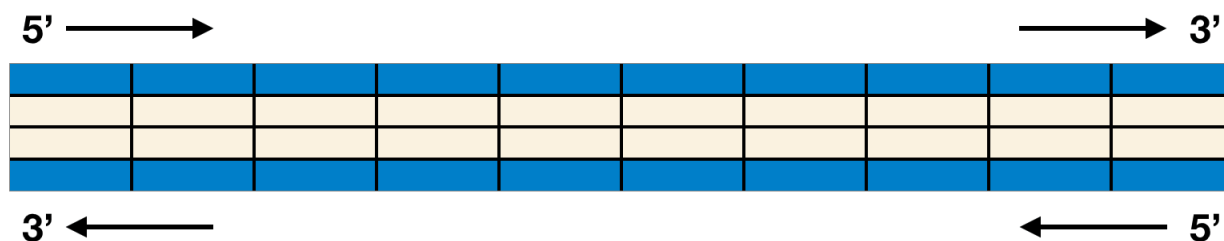


Figure 3. Structural design of the detailed damage scoring in field 7 of the data structure.

All unmarked sites are assumed to be unaffected, while damaged sites are marked numerically:

Strand Damages are denoted as a 1 for point breaks on a strand from direct effects, and 2 for lesions (losses or attachments) from a single indirect damage, 3 for multiple damages to the same base pair strand from two or more direct or indirect interactions. In addition, 0 can be used to denote non-damage inducing interactions, i.e. events that are below the damage induction threshold. 4+ can be used for adducts or other modifications, according to the source program, for example to define additional details such as what type of indirect reaction occurred. This may be relevant to account for differences in repair likelihood for different types of reactions from radiolysis such as dehydrogenation, OH addition or deoxyribose damage. All additional numbering schemes should be submitted to and detailed on the collaboration webpage to ensure a unique numbering scheme.

Base Damages are denoted in the same manner as for strand breaks, with 1 for point losses from direct effects, 2 for lesions (losses or attachments) from indirect effects, and 3 for multiple damages from direct or indirect damages. 0 can be used to indicate non-damage inducing interactions. Again, all additional suggested numbering schemes should be submitted to and detailed on the collaboration webpage to ensure a unique numbering scheme.

The damages are recorded in a comma-separated list of: Strand (row), Base Pair (column), Damage type, with each damage separated by the delimiter '/'. Damage events are recorded by row, beginning with the 5' to 3' backbone, then its accompanying bases, then the 3' to 5' bases, then finally their backbone. Within each row, damages are then recorded in the 5' to 3' direction.

To illustrate the syntax of this field, Figure 4 showcases example break types, ranging from a single base deletion to a highly complex damage with multiple losses in both bases and strands. These are presented both as schematic DNA sections as in Figure 3, and an accompanying structure definition below the image.

For example, for the case of multiple base double strand break with overhang, with both direct and indirect damages illustrated in Figure 4d, the definition reads: "1, 2, 1 / 1, 3, 1 / 1, 4, 1 / 1, 5, 1 / 1, 9, 0 / 2, 1, 1 / 2, 2, 1 / 2, 3, 2 / 2, 4, 2 / 2, 5, 3 / 3, 3, 1 / 3, 4, 2 / 3, 5, 1 / 3, 6, 2 / 3, 7, 0 / 4, 3, 2 / 4, 4, 2 / 4, 5, 2 / 4, 6, 1;". The base pair count of each damage site starts with the first occurring damage, which in this case is on the 5' to 3' strand base (2, 1, 1). However, the damages are stored starting with the 5' to 3' strand backbone, and the first damage for this strand occurs on the second BP, so the damage definition starts with (1, 2, 1). The first 1 indicates the strand, the middle 2 indicates that a damage occurred on the second BP, and the 1 for the last value shows that this was a direct damage. The second, third and fourth triplets (1, 3, 1 / 1, 4, 1 / 1, 5, 1) similarly define 3 more direct damages on the following three BPs. The

fifth triplet (1, 9, 0) shows that there was another interaction with the strand that did not result in a lesion. If additional damages on this strand would have occurred within this break, it would be listed next, for example an additional single direct backbone damage at BP 10 would add a (1, 10, 1) next.

The group '2, 1, 1 / 2, 2, 1' likewise defines a block of damages on the 5' to 3' bases, starting at BP 1 up to BP 2 caused by direct damages, and the next '2, 3, 2 / 2, 4, 2' indicates the next two bases were damaged from indirect processes. '2, 5, 3' shows that the fifth BP was hit by multiple events in any combination of direct and indirect lesions. The opposite bases start with 3 and the block '3, 3, 1 / 3, 4, 2 / 3, 5, 1 / 3, 6, 2' defines damages on BP 3-6 alternating between direct and indirect damages, followed by another interaction that did not result in a lesion '3, 7, 0'. The 3' to 5' backbone has the same positions damaged but with the first 3 damages from indirect processes recorded as '4, 3, 2 / 4, 4, 2 / 4, 5, 2', and an additional direct damage '4, 6, 1'.

a) Single base deletion from direct damage:

D									

Result: Field 3: 1, 0, 0;

Field 7: 2, 1, 1;

b) Single base deletion from direct damage and single strand break from indirect damage, plus two non-damaging events:

		I							
D									
					*				
							*		

Result: Field 3: 1, 1, 0;

Field 7: 1, 3, 2 / 2, 1, 1 / 3, 6, 0 / 4, 8, 0;

c) Point double strand break, all direct damage:

D									
D									
D									
D									

Result: Field 3: 2, 2, 1;

Field 7: 1, 1, 1 / 2, 1, 1 / 3, 1, 1 / 4, 1, 1;

d) Multiple base double strand break with overhang, with direct, indirect, multiple and non-damages:

	D	D	D	D				*	
D	D	I	I	M					
		D	I	D	I	*			
		I	I	I	D				

Result: Field 3: 9, 8, 1;

Field 7: 1, 2, 1 / 1, 3, 1 / 1, 4, 1 / 1, 5, 1 / 1, 9, 0 / 2, 1, 1 / 2, 2, 1 / 2, 3, 2 / 2, 4, 2 / 2, 5, 3 / 3, 3, 1 / 3, 4, 2 / 3, 5, 1 / 3, 6, 2 / 3, 7, 0 / 4, 3, 2 / 4, 4, 2 / 4, 5, 2 / 4, 6, 1;

e) Highly complex damage, with both direct and indirect damages and multiple damaged sites:

			I	I	M	I	D		
	I		I	D	D	D			
			D	M	I	I		D	
D			D	I	D				

Result: Field 3: 10, 9, 1;

Field 7: 1, 4, 2 / 1, 5, 2 / 1, 6, 3 / 1, 7, 2 / 1, 8, 1 / 2, 2, 2 / 2, 4, 2 / 2, 5, 1 / 2, 6, 1 / 2, 7, 1 / 3, 4, 1 / 3, 5, 3 / 3, 6, 2 / 3, 7, 2 / 3, 9, 1 / 4, 1, 1 / 4, 4, 1 / 4, 5, 2 / 4, 6, 1

Figure 4. Examples for the damage definition structure used in field 7 of the data structure. Here “” lists interactions that were not sufficient to cause a damage, i.e. below the cut-off defined in the header, “D” denotes direct damages, “I” stands for indirect damages, “M” for multiple damages from any combination of D and I events. These events are defined by values of 0, 1, 2, 3, respectively.*

**Either the full break spec or field 6 (damage types) need to be included, and ideally both will be provided. While this field can be omitted if field 6 is provided, all codes simulating the induction of DNA should strive to eventually provide the full break specification. This field, in combination with field 3&4 (chromosome position) can be used to identify exactly where along the strand the damage occurred, for example to obtain the number of non-hit BPs on either side of the lesion.

Field 8, DNA Sequence (optional): Provides a field to further specify the surrounding DNA sequence of the site that was damaged. Information about the structural geometry (e.g. heterochromatin or euchromatin) should be provided in field 3. This field consists of a $M \times N$ array to record the DNA sequence. Here M refers to the number of strands involved, i.e. 2 for a standard double helix DNA section. N is the number of BPs involved in the damage definition, for the example in figure 4c it would be 1, for 4e it would be 9. For a double helix without mismatched or modified bases, the array can also be reduced to a $1 \times N$ array without loss of information.

For models incorporating the actual physical structure of individual bases, the DNA sequence along the strand can be included in this field. The design uses the same layout as the break structures above, beginning from the 5' end at the position of the first damage. Bases are written in sequence for the 5' to 3' strand, stored as strings of integer values, with bases denoted as: Missing=0, A=1, C=2, T=3, G=4. Backbones are not specified in these data.

This field is kept optional as most codes do not yet consider the DNA sequence. However, evidence exists that specific types of individual lesions formed by ionizing radiation differ for A, G, T and C (94). There is also evidence in the literature that the larger-scale base sequence can have an impact on the types and quantities of individual lesions created by irradiation, and how this is repaired (95,96). An example of this latter effect would be hole migration along a DNA molecule. These effects are generally not yet considered in most current Monte Carlo DNA damage models, although the RADAMOL code (75) offers an option to include charge migration. Such migrations have potentially important implications for modeling of DNA repair.

Field 9, Lesion Time (optional): This field is provided to add a time for each induced damage in nanoseconds starting from the first recorded damage, using the same order as in field 7. If only a single value is given here, that is assumed to be the time at which the whole damage site enters the simulation. The values are recorded separated by '/', for example for the case shown in figure 4b denoted as (1, 3, 2 / 2, 1, 1 / 3, 6, 0 / 4, 8, 0;), the time structure could be "2.1 / 0 / 0.0000008 / 0.000001". This translates to an event with an initial direct damage (on 2, 1, 1), a direct base and a direct backbone damage below the break threshold 0.8 and 1 fs later (3, 6, 0) and (4, 8, 0), respectively, and an indirect damage 2.1 ns after the first break (1, 3, 2).

Field 10, Particle Types (optional): This and the following fields are optional fields to describe the primary particles from the irradiation source that actually caused the recorded damage by itself, through secondary particles or chemical reactions. The primaries should be defined in the header section. Field 10 defines the particle type for each involved source particle using comma-separated integers (PDG values). For single particle type irradiations, this field is already defined by header field 7 and can be omitted.

Field 11, Energies (optional): The corresponding initial particle energies of the particles defined above, in MeV, one comma-separated entry per particle defined in field 10. Similar to field 10, for monoenergetic irradiations, this field is already defined in header field 8.

Field 12, Translation (optional): The 3-vectors (X, Y, Z) of the starting points of the particles, in μm relative to the centre of the world volume, stored as a comma-separated list of one 3-tuple for each particle, with each value within the 3-tuple separated using '/ '.

Field 13, Direction (optional): Euler rotation angles (ϕ , θ , ψ) for the above particles following the same style as for the translations to define their direction. A rotation of 0/0/0 is defined as having the particles propagate along the +Z direction.

Field 14, Particle Time (optional): A list of comma-separated floats giving the start time in ns of each particle defined in field 10 with t=0 defined as the time when a new exposure starts (see field 1 and header field 11). This may be particularly important for very low dose rate exposures such as those from GCRs.

2.4 Dissemination and repository

We have set up a website (<https://standard-for-dna-damage.readthedocs.org/>) where the standard is documented and to which new requests for enumeration schemes can be submitted. In addition, we provide a link to our GitHub repository which offers selected example codes and provides a place to share SDD files. Such a repository will be useful for modellers who do not have the resources to perform their own full damage simulations, and to test (new) damage models against other (published) models that provided data here. This repository also includes a code to generate random damage distributions in the SDD format and example SDD files.

3. Discussion

The outlined standardized data format for DNA damages (SDD) is intended to provide the basis for cross-disciplinary investigations of DNA damage induction and ensuing kinetics of DNA repair mechanisms. By standardizing the recording format of the distribution of damages and their structural pattern for single cells and nuclei, we anticipate creating synergies between various developments in modeling cellular response to DNA damage.

The standard has been developed anticipating several future developments in the field. For example, modeling of the initial chemical reactions is becoming increasingly common, allowing to include more sophisticated effects such as the potential production of additional reactive species induced by shock waves from high-LET ions (97) or neutralization effects of Auger emitters (98). With an increased range of chemical reactions simulated, repair mechanisms and accuracy can be adjusted depending on the particular indirect effect classes of DNA damage. Similarly, radiation therapy is often only one of the delivered treatment modalities and combined with chemotherapy or immunotherapy. Modeling of multi-modality treatments is an emerging field, and good quality input data from each modality is essential. The effects of other treatments can be partially included in the SDD by adding DNA damages from drugs. For therapeutics that inhibit certain repair pathways the biological models will have to be adjusted, with the SDD providing detailed DNA damage maps.

While the nucleus is the primary target in radiation therapy, the standard is flexible enough to also be used to describe damages to mitochondrial DNA or DNA in viruses or bacteria (in a separate file). However, for these cases many of the optional fields in the SDD may not be relevant. The SDD has been designed to allow a high level of flexibility. Many of the entries are optional and are only included to encourage people to think about the concepts and if possible, include these details as they may become useful for repair kinetics.

Overall, we anticipate that the standardized DNA damage (SDD) data format will greatly reduce the burden of sharing analysis tools and thus, facilitate the formation of new collaborations. Using standardized

data will allow researchers to test the predictions from different models simply by feeding the SDD data to another code. The standard already is (or will soon be) supported by the following codes: DaMaRiS (39), gMicroMC, MC4 (63), MCDS (27,79), PARTRAC (60), PHITS (99), RADAMOL (75,100), RITRACK (57), TOPAS-nBio (65) as well as by users of Geant4-DNA (23,61). By providing a clearly defined standard and example codes of scorers for some of the models, we hope to provide an incentive for other existing and newly developed codes to offer interfaces to the SDD data format both for using it as scorer or as damage distribution input for repair models.

4. Conclusion

We have developed a new Standard DNA Damage data format. The SDD has been designed to interface at the point where physics and chemistry simulations, at the DNA scale, meet biological modeling efforts. With this standard, we hope to provide modellers with a new tool to test their model design and dependencies on underlying physics properties. In combination with the supported collaboration website, the SDD offers access to the most accurate available damage simulations and provides a platform for inter-code comparisons of the underlying track structure Monte Carlo simulation codes and their assumptions in the description of physics, chemistry and the geometrical arrangement of DNA. This standard will play a significant role in advancing our understanding of DNA response to radiation insults by creating the basis for a wide spread interdisciplinary collaborative effort.

5. Acknowledgements

JS, ALM, JWW, NTH, KK, MJM and SJM would like to acknowledge support from the STFC-funded Global Challenge Network+ in Advanced Radiotherapy Multi-Scale Monte Carlo Modeling for Radiotherapy Sandpit (ST/N002423/1). JS would further like to acknowledge the support from the NIH/NCI under R01CA187003 (“TOPAS-nBio: a Monte Carlo tool for radiation biology research”).

Supplemental Data:

6. Example SDD files

With many detailed options of scoring the DNA damage in the SDD format, we believe it is helpful to illustrate the format with example SDD files. We have created two files using the McMahon Empirical Model v0.3 to generate DNA damages for an irradiation of a cell to 1 Gy with a 0.975 MeV monoenergetic proton beam. The two files use identical irradiation setup, but the first (DNA Damage Proton0.975 MeV 1 Gy full.txt) fills all available scoring blocks (see the header field “Data entries”), including records of the full damage definition of data field 7. The second file (DNA Damage Proton0.975 MeV 1 Gy minimal.txt) showcases an SDD file that only fills a minimum of 3 of the 14 data sections. Nevertheless, even the minimal data format still offers useful information about the frequencies of various damage types. Links to additional examples and example codes to produce SDD files can be found on the SDD website:

[Readthedocs website for SDD to define new definitions/numbers \(in preparation\).](http://standard-for-dna-damage.readthedocs.org)
<http://standard-for-dna-damage.readthedocs.org>

6. References:

1. McNamara AL, Schuemann J, Paganetti H. A phenomenological relative biological effectiveness (RBE) model for proton therapy based on all published in vitro cell survival data. *Phys Med Biol*. 2015 Nov 7;60(21):8399–416.
2. Wedenberg M, Lind BK, Hårdemark B. A model for the relative biological effectiveness of protons: the tissue specific parameter α/β of photons is a predictor for the sensitivity to LET changes. *Acta Oncol*. 2013 Apr;52(3):580–8.
3. Carabe-Fernandez A, Dale RG, Jones B. The incorporation of the concept of minimum RBE (RBE_{min}) into the linear-quadratic model and the potential for improved radiobiological analysis of high-LET treatments. *Int J Radiat Biol*. 2009 Jul 3;83(1):27–39.
4. Wilkens JJ, Oelfke U. A phenomenological model for the relative biological effectiveness in therapeutic proton beams. *Phys Med Biol*. 2004 Jun 12;49(13):2811–25.
5. Elsässer T, Krämer M, Scholz M. Accuracy of the Local Effect Model for the Prediction of Biologic Effects of Carbon Ion Beams In Vitro and In Vivo. *International Journal of Radiation Oncology* Biology* Physics*. 2008 Jul;71(3):866–72.
6. Hawkins RB. A statistical theory of cell killing by radiation of varying linear energy transfer. *Radiation Research*. 1994 Dec;140(3):366–74.
7. Inaniwa T, Kanematsu N, Matsufuji N, Kanai T, Shirai T, Noda K, et al. Reformulation of a clinical-dose system for carbon-ion radiotherapy treatment planning at the National Institute of Radiological Sciences, Japan. *Phys Med Biol*. IOP Publishing; 2015 Apr 21;60(8):3271–86.
8. Inaniwa T, Furukawa T, Kase Y, Matsufuji N, Toshito T, Matsumoto Y, et al. Treatment planning for a scanned carbon beam with a modified microdosimetric kinetic model. *Phys Med Biol*. 2010 Oct 28;55(22):6721–37.
9. Krämer M, Scholz M. Treatment planning for heavy-ion radiotherapy: calculation and optimization of biologically effective dose. *Phys Med Biol*. 2000 Nov;45(11):3319–30.
10. Bolst D, Tran LT, Chartier L, Prokopovich DA, Pogossoy A, Guatelli S, et al. RBE study using solid state microdosimetry in heavy ion therapy. *Radiation Measurements*. 2017 Nov;106:512–8.
11. Goodhead DT, Charlton DE. Analysis of high-LET radiation effects in terms of local energy deposition. *Radiation Protection Dosimetry*. 1985.
12. Goodhead DT, Charlton DE, Wilson WE, Paretzke HG. Current biophysical approaches to the understanding of biological effects of radiation in terms of local energy deposition. In: Schraube H, Burger G, Booz JJ, editors. 1984. pp. 57–68.
13. Goodhead DT, Brenner DJ. Estimation of a single property of low LET radiations which correlates with biological effectiveness. *Phys Med Biol*. 2000 Dec 6;28(5):485–92.
14. Charlton DE, Humm JL. A method of calculating initial DNA strand breakage following the decay of incorporated ¹²⁵I. *Int J Radiat Biol Relat Stud Phys Chem Med*. 1988 Mar;53(3):353–65.

15. Goodhead DT, Nikjoo H. Track structure analysis of ultrasoft X-rays compared to high- and low-LET radiations. *Int J Radiat Biol.* 1989 Apr;55(4):513–29.
16. Nikjoo H, Goodhead DT, Charlton DE, Paretzke HG. Energy deposition in small cylindrical targets by ultrasoft x-rays. *Phys Med Biol.* 1989 Jun;34(6):691–705.
17. Charlton DE, Nikjoo H, Humm JL. Calculation of initial yields of single- and double-strand breaks in cell nuclei from electrons, protons and alpha particles. *Int J Radiat Biol.* 1989 Jul;56(1):1–19.
18. Paretzke HG. Physical events in the track structure of heavy ions and their relation to alterations of biomolecules. *Advances in Space Research.* 1989 Jan;9(10):15–20.
19. Pomplun E. A New DNA Target Model for Track Structure Calculations and Its First Application to I-125 Auger Electrons. *Int J Radiat Biol.* Taylor & Francis; 1991 Jul 3;59(3):625–42.
20. Terrissol M, Pomplun E. Computer Simulation of DNA-Incorporated 125I Auger Cascades and of the Associated Radiation Chemistry in Aqueous Solution. *Radiation Protection Dosimetry.* Oxford University Press; 1994 Apr 1;52(1-4):177–81.
21. Nikjoo H, O'Neill P, Goodhead DT, Terrissol M. Computational modelling of low-energy electron-induced DNA damage by early physical and chemical events. *Int J Radiat Biol.* 1997 May;71(5):467–83.
22. Friedland W, Jacob P, Paretzke HG, Stork T. Monte Carlo simulation of the production of short DNA fragments by low-linear energy transfer radiation using higher-order DNA models. *Radiation Research.* 1998;150(2):170.
23. Bernal MA, Bordage MC, Brown JMC, Davidkova M, Delage E, Bitar El Z, et al. Track structure modeling in liquid water: A review of the Geant4-DNA very low energy extension of the Geant4 Monte Carlo simulation toolkit. *Phys Med.* 2015 Dec;31(8):861–74.
24. Tello JJ, Incerti S, Francis Z, Tran H, Bernal MA. Numerical insight into the Dual Radiation Action Theory. *Phys Med.* 2017 Nov;43:120–6.
25. Tello Cajiao JJ, Carante MP, Bernal Rodriguez MA, Ballarini F. Proximity effects in chromosome aberration induction by low-LET ionizing radiation. *DNA Repair.* 2017 Oct;58:38–46.
26. Stewart RD, Streitmatter SW, Argento DC, Kirkby C, Goorley JT, Moffitt G, et al. Rapid MCNP simulation of DNA double strand break (DSB) relative biological effectiveness (RBE) for photons, neutrons, and light ions. *Phys Med Biol.* IOP Publishing; 2015 Nov 7;60(21):8249–74.
27. Stewart RD, Yu VK, Georgakilas AG, Koumenis C, Park JH, Carlson DJ. Effects of Radiation Quality and Oxygen on Clustered DNA Lesions and Cell Death. *Radiation Research.* 2011 Nov;176(5):587–602.
28. Semenenko VA, Stewart RD. Fast Monte Carlo simulation of DNA damage formed by electrons and light ions. *Phys Med Biol.* 2006 Apr 7;51(7):1693–706.

29. Semenenko VA, Stewart RD. A fast Monte Carlo algorithm to simulate the spectrum of DNA damages formed by ionizing radiation. *Radiation Research*. 2004 Apr;161(4):451–7.
30. Frese MC, Yu VK, Stewart RD, Carlson DJ. A Mechanism-Based Approach to Predict the Relative Biological Effectiveness of Protons and Carbon Ions in Radiation Therapy. *International Journal of Radiation Oncology*Biophysics*. 2012 May;83(1):442–50.
31. Kamp F, Cabal G, Mairani A, Parodi K, Wilkens JJ, Carlson DJ. Fast Biological Modeling for Voxel-based Heavy Ion Treatment Planning Using the Mechanistic Repair-Misrepair-Fixation Model and Nuclear Fragment Spectra. *Int J Radiat Oncol Biol Phys*. 2015 Nov 1;93(3):557–68.
32. Ballarini F, Carante MP. Chromosome aberrations and cell death by ionizing radiation: Evolution of a biophysical model. *Radiat Phys Chem*. 2016 Nov 1;128:18–25.
33. Carante MP, Aimè C, Cajiao JJT, Ballarini F. BIANCA, a biophysical model of cell survival and chromosome damage by protons, C-ions and He-ions at energies and doses used in hadrontherapy. *Phys Med Biol*. IOP Publishing; 2018 Mar 26;63(7):075007.
34. Li Y, Cucinotta FA. Modeling non-homologous end joining. *Journal of Theoretical Biology*. 2011 Aug 21;283(1):122–35.
35. Li Y, Reynolds P, O'Neill P, Cucinotta FA. Modeling Damage Complexity-Dependent Non-Homologous End-Joining Repair Pathway. Lustig AJ, editor. *PLoS ONE*. Public Library of Science; 2014 Feb 10;9(2):e85816.
36. McMahon SJ, Butterworth KT, McGarry CK, Trainor C, O'Sullivan JM, Hounsell AR, et al. A computational model of cellular response to modulated radiation fields. *Int J Radiat Oncol Biol Phys*. 2012 Sep 1;84(1):250–6.
37. McMahon SJ, McNamara AL, Schuemann J, Paganetti H, Prise KM. A general mechanistic model enables predictions of the biological effectiveness of different qualities of radiation. *Sci Rep*. Nature Publishing Group; 2017 Sep 7;7(1):688.
38. Henthorn NT, Warmenhoven JW, Sotiropoulos M, Mackay RI, Kirkby NF, Kirkby KJ, et al. In Silico Non-Homologous End Joining Following Ion Induced DNA Double Strand Breaks Predicts That Repair Fidelity Depends on Break Density. *Sci Rep*. Nature Publishing Group; 2018 Feb 8;8(1):2654.
39. Henthorn NT, Warmenhoven JW, Sotiropoulos M, Mackay RI, Kirkby KJ, Merchant MJ. Nano-dosimetric Simulation of Direct Ion-Induced DNA Damage Using Different Chromatin Geometry Models. *Radiation Research*. 2017 Dec;188(6):770–83.
40. Lampe N, Karamitros M, Breton V, Brown JMC, Kyriakou I, Sakata D, et al. Mechanistic DNA damage simulations in Geant4-DNA part 1: A parameter study in a simplified geometry. *Phys Med*. 2018 Apr 5.
41. McMahon SJ, Schuemann J, Paganetti H, Prise KM. Mechanistic Modelling of DNA Repair and Cellular Survival Following Radiation-Induced DNA Damage. *Sci Rep*. 2016;6:33290.

42. Ballarini F, Altieri S, Bortolussi S, Carante M, Giroletti E, Protti N. The role of DNA cluster damage and chromosome aberrations in radiation-induced cell killing: a theoretical approach. *Radiation Protection Dosimetry*. 2015 Sep;166(1-4):75–9.
43. P Carante M, Ballarini F, 1 Department of Physics, University of Pavia, via Bassi 6, I-27100 Pavia, Italy, 2 INFN-National Institute of Nuclear Physics, Section of Pavia, via Bassi 6, I-27100 Pavia, Italy, Pavia I-2, Italy. Modelling cell death for cancer hadrontherapy. *AIMS Biophysics*. 2017;4(3):465–90.
44. Testa A, Ballarini F, Giesen U, Gil OM, Carante MP, Tello J, et al. Analysis of Radiation-Induced Chromosomal Aberrations on a Cell-by-Cell Basis after Alpha-Particle Microbeam Irradiation: Experimental Data and Simulations. *Radiation Research*. 2018 Jun;189(6):597–604.
45. Tommasino F, Friedrich T, Scholz U, Taucher-Scholz G, Durante M, Scholz M. A DNA double-strand break kinetic rejoining model based on the local effect model. *Radiation Research*. 2013 Nov;180(5):524–38.
46. Tommasino F, Friedrich T, Scholz U, Taucher-Scholz G, Durante M, Scholz M. Application of the local effect model to predict DNA double-strand break rejoining after photon and high-LET irradiation. *Radiation Protection Dosimetry*. 2015 Sep;166(1-4):66–70.
47. Tommasino F, Friedrich T, Jakob B, Meyer B, Durante M, Scholz M. Induction and Processing of the Radiation-Induced Gamma-H2AX Signal and Its Link to the Underlying Pattern of DSB: A Combined Experimental and Modelling Study. *PLoS ONE*. 2015;10(6):e0129416.
48. Friedland W, Li WB, Jacob P, Paretzke HG. Simulation of exon deletion mutations induced by low-LET radiation at the HPRT locus. *Radiation Research*. 2001 May;155(5):703–15.
49. Ponomarev AL, Costes SV, Cucinotta FA. Stochastic properties of radiation-induced DSB: DSB distributions in large scale chromatin loops, the HPRT gene and within the visible volumes of DNA repair foci. *Int J Radiat Biol*. 2008 Nov;84(11):916–29.
50. Meylan S, Vimont U, Incerti S, Clairand I, Villagrasa C. Geant4-DNA simulations using complex DNA geometries generated by the DnaFabric tool. *Computer Physics Communications*. 2016 Mar.
51. Bueno M, Schulte R, Meylan S, Villagrasa C. Influence of the geometrical detail in the description of DNA and the scoring method of ionization clustering on nanodosimetric parameters of track structure: a Monte Carlo study using Geant4-DNA. *Phys Med Biol*. 2015 Nov 7;60(21):8583–99.
52. Santos Dos M, Villagrasa C, Clairand I, Incerti S. Influence of the DNA density on the number of clustered damages created by protons of different energies. *Nuclear Inst and Methods in Physics Research, B*. 2013 Mar 1;298(C):47–54.
53. Nikjoo H, Girard P. A model of the cell nucleus for DNA damage calculations. *Int J Radiat Biol*. 2012 Jan;88(1-2):87–97.

54. Nikjoo H, Uehara S, Emfietzoglou D, Cucinotta FA. Track-structure codes in radiation research. *Radiation Measurements*. 2006 Oct 1;41(9):1052–74.
55. Meylan S, Incerti S, Karamitros M, Tang N, Bueno M, Clairand I, et al. Simulation of early DNA damage after the irradiation of a fibroblast cell nucleus using Geant4-DNA. *Sci Rep. Nature Publishing Group*; 2017 Sep 20;7(1):11923.
56. Bug MU, Yong Baek W, Rabus H, Villagrasa C, Meylan S, Rosenfeld AB. An electron-impact cross section data set (10 eV–1 keV) of DNA constituents based on consistent experimental data: A requisite for Monte Carlo simulations. *Radiat Phys Chem. Pergamon*; 2017 Jan 1;130:459–79.
57. A F. Monte-Carlo Simulation of Particle Diffusion in Various Geometries and Application to Chemistry and Biology. In: *Theory and Applications of Monte Carlo Simulations*. InTech; 2013.
58. Plante I, Cucinotta FA. Binary-Encounter-Bethe ionisation cross sections for simulation of DNA damage by the direct effect of ionising radiation. *Radiation Protection Dosimetry. Oxford University Press*; 2015 Sep 1;166(1-4):19–23.
59. Friedland W, Dingfelder M, Kunderát P, Jacob P. Track structures, DNA targets and radiation effects in the biophysical Monte Carlo simulation code PARTRAC. *Mutat Res*. 2011 Jun 3;711(1-2):28–40.
60. Friedland W, Schmitt E, Kunderat P, Dingfelder M, Baiocco G, Barbieri S, et al. Comprehensive track-structure based evaluation of DNA damage by light ions from radiotherapy-relevant energies down to stopping. *Sci Rep*. 2017 Mar 27;7(1):45161.
61. Incerti S, Ivanchenko A, Karamitros M, Mantero A, Moretto P, Tran HN, et al. Comparison of GEANT4 very low energy cross section models with experimental data in water. *Med Phys*. 2010 Sep;37(9):4692–708.
62. Nikjoo H, Emfietzoglou D, Liamsuwan T, Taleei R, Liljequist D, Uehara S. Radiation track, DNA damage and response-a review. *Rep Prog Phys. IOP Publishing*; 2016 Nov;79(11):116601.
63. Emfietzoglou D, Papamichael G, Nikjoo H. Monte Carlo Electron Track Structure Calculations in Liquid Water Using a New Model Dielectric Response Function. *Radiation Research. The Radiation Research Society*; 2017 Jun 26;188(3):355–68.
64. Wälzlein C, Krämer M, Scifoni E, Durante M. Low-energy electron transport in non-uniform media. *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms*. 2014 Feb 1;320:75–82.
65. McNamara A, Geng C, Turner R, Méndez JR, Perl J, Held K, et al. Validation of the radiobiology toolkit TOPAS-nBio in simple DNA geometries. *Phys Med*. 2017 Jan;33:207–15.
66. Solov'yov IA, Yakubovich AV, Nikolaev PV, Volkovets I, Solov'yov AV. MesoBioNano Explorer--a universal program for multiscale computer simulations of complex molecular structure and dynamics. *J Comput Chem. Wiley-Blackwell*; 2012 Nov 15;33(30):2412–39.

67. Sung W, Ye S-J, McNamara AL, McMahon SJ, Hainfeld J, Shin J, et al. Dependence of gold nanoparticle radiosensitization on cell geometry. *Nanoscale*. 2017 May 11;9(18):5843–53.
68. Solov'yov AV, Surdutovich E, Scifoni E, Mishustin I, Greiner W. Physics of ion beam cancer therapy: a multiscale approach. *Phys Rev E Stat Nonlin Soft Matter Phys*. American Physical Society; 2009 Jan;79(1 Pt 1):011909.
69. Surdutovich E, Solov'yov AV. Multiscale approach to the physics of radiation damage with ions. *Eur Phys J D*. Springer Berlin Heidelberg; 2014;68(11):353.
70. Verkhovtsev A, Surdutovich E, Solov'yov AV. Multiscale approach predictions for biological outcomes in ion-beam cancer therapy. *Sci Rep*. Nature Publishing Group; 2016 Jun 14;6(1):27654.
71. Baiocco G, Barbieri S, Babini G, Morini J, Alloni D, Friedland W, et al. The origin of neutron biological effectiveness as a function of energy. *Sci Rep*. 2016 Sep 22;6(1):34033.
72. Karamitros M, Luan S, Bernal MA, Allison J, Baldacchino G, Davidkova M, et al. Diffusion-controlled reactions modeling in Geant4-DNA. *Journal of Computational Physics*. 2014 Oct;274:841–82.
73. Plante I, Devroye L. Considerations for the independent reaction times and step-by-step methods for radiation chemistry simulations. *Radiat Phys Chem*. Pergamon; 2017 Oct 1;139:157–72.
74. Incerti S, Douglass M, Penfold S, Guatelli S, Bezak E. Review of Geant4-DNA applications for micro and nanoscale simulations. *Phys Med*. 2016 Oct;32(10):1187–200.
75. Štěpán V, Davidková M. RADAMOL tool: Role of radiation quality and charge transfer in damage distribution along DNA oligomer. *Eur Phys J D*. 2014 Aug 14;68(8):409.
76. Boscolo D, Krämer M, Durante M, Fuss MC, Scifoni E. TRAX-CHEM: A pre-chemical and chemical stage extension of the particle track structure code TRAX in water targets. *Chem Phys Lett*. 2018 Apr;698:11–8.
77. Ramos-Méndez J, Perl J, Schuemann J, McNamara A, Paganetti H, Faddegon B. Monte Carlo simulation of chemistry following radiolysis with TOPAS-nBio. *Phys Med Biol*. 2018 May 17;63(10):105014.
78. Nikjoo H, Taleei R, Liamsuwan T, Liljequist D, Emfietzoglou D. Perspectives in radiation biophysics: From radiation track structure simulation to mechanistic models of DNA damage and repair. *Radiat Phys Chem*. Pergamon; 2016 Nov 1;128:3–10.
79. Carlson DJ, Stewart RD, Semenenko VA, Sandison GA. Combined use of Monte Carlo DNA damage simulations and deterministic repair models to examine putative mechanisms of cell killing. *Radiation Research*. 2008 Apr;169(4):447–59.
80. Taleei R, Nikjoo H. The non-homologous end-joining (NHEJ) pathway for the repair of DNA double-strand breaks: I. A mathematical model. *Radiation Research*. 2013 May;179(5):530–9.

81. Taleei R, Girard PM, Sankaranarayanan K, Nikjoo H. The non-homologous end-joining (NHEJ) mathematical model for the repair of double-strand breaks: II. Application to damage induced by ultrasoft X rays and low-energy electrons. *Radiation Research*. 2013 May;179(5):540–8.
82. Semenenko VA, Stewart RD, Ackerman EJ. Monte Carlo simulation of base and nucleotide excision repair of clustered DNA damage sites. I. Model properties and predicted trends. *Radiation Research*. 2005 Aug;164(2):180–93.
83. Semenenko VA, Stewart RD. Monte Carlo Simulation of Base and Nucleotide Excision Repair of Clustered DNA Damage Sites. II. Comparisons of Model Predictions to Measured Data. *Radiation Research*. 2005 Aug;164(2):194–201.
84. Georgakilas AG, O'Neill P, Stewart RD. Induction and repair of clustered DNA lesions: what do we know so far? *Radiation Research*. 2013 Jul;180(1):100–9.
85. Adams GE, Jameson DG. Time effects in molecular radiation Biology. *Radiat Environ Biophys*. Springer-Verlag; 1980 Feb;17(2):95–113.
86. Lampe N, Karamitros M, Breton V, Brown JMC, Sakata D, Sarramia D, et al. Mechanistic DNA damage simulations in Geant4-DNA Part 2: Electron and proton damage in a bacterial cell. *Phys Med*. 2018 Jan 19.
87. Shuryak I, Brenner DJ. Effects of radiation quality on interactions between oxidative stress, protein and DNA damage in *Deinococcus radiodurans*. *Radiat Environ Biophys*. 2010 Nov;49(4):693–703.
88. Tanabashi M, (Particle Data Group). The Review of Particle Physics. *Phys Rev D* [Internet]. 2018;98. Available from: <http://pdg.lbl.gov/>
89. Favaudon V, Fouillade C, Vozenin M-C. Ultrahigh dose-rate, “flash” irradiation minimizes the side-effects of radiotherapy. *Cancer Radiother*. 2015 Oct;19(6-7):526–31.
90. Favaudon V, Caplier L, Monceau V, Pouzoulet F, Sayarath M, Fouillade C, et al. Ultrahigh dose-rate FLASH irradiation increases the differential response between normal and tumor tissue in mice. *Science Translational Medicine*. American Association for the Advancement of Science; 2014 Jul 16;6(245):245ra93–3.
91. Meyer J, Stewart RD, Smith D, Eagle J, Lee E, Cao N, et al. Biological and dosimetric characterisation of spatially fractionated proton minibeam. *Phys Med Biol*. IOP Publishing; 2017 Nov 21;62(24):9260–81.
92. Incerti S, Barberet P, Villeneuve R, Aguer P, Gontier E, Michelet-Habchi C, et al. Simulation of cellular irradiation with the CENBG microbeam line using GEANT4. *IEEE Trans Nucl Sci*. 2004;51(4):1395–401.
93. Nikjoo H, O'Neill P, Wilson WE, Goodhead DT. Computational approach for determining the spectrum of DNA damage induced by ionizing radiation. *Radiation Research*. 2001 Nov;156(5 Pt 2):577–83.

94. Téoule R. Radiation-induced DNA Damage and Its Repair. *Int J Radiat Biol Relat Stud Phys Chem Med*. Taylor & Francis; 2009 Jul 3;51(4):573–89.
95. Gu J, Lu H, Tsai AG, Schwarz K, Lieber MR. Single-stranded DNA ligation and XLF-stimulated incompatible DNA end ligation by the XRCC4-DNA ligase IV complex: influence of terminal DNA sequence. *Nucleic Acids Res*. Oxford University Press; 2007 Sep 1;35(17):5755–62.
96. Audebert M, Salles B, Calsou P. Effect of double-strand break DNA sequence on the PARP-1 NHEJ pathway. *Biochem Biophys Res Commun*. 2008 May 9;369(3):982–8.
97. Surdutovich E, Solov'yov AV. Shock wave initiated by an ion passing through liquid water. *Phys Rev E Stat Nonlin Soft Matter Phys*. American Physical Society; 2010 Nov;82(5 Pt 1):051915.
98. Li WB, Friedland W, Jacob P, Panyutin IG, Paretzke HG. Simulation of ¹²⁵I decay in a synthetic oligodeoxynucleotide with normal and distorted geometry and the role of radiation and non-radiation actions. *Radiat Environ Biophys*. 2004;43(1):23–33.
99. Sato T, Iwamoto Y, Hashimoto S, Ogawa T, Furuta T, Abe S-I, et al. Features of Particle and Heavy Ion Transport code System (PHITS) version 3.02. *Journal of Nuclear Science and Technology*. Taylor & Francis; 2018 Jan 5;55(6):684–90.
100. Štěpán V, Davidková M. Understanding radiation damage on sub-cellular scale using RADAMOL simulation tool. *Radiat Phys Chem*. 2016 Nov;128:11–7.