JOINT INSTITUTE FOR NUCLEAR RESEARCH
The Laboratory of Radiation Biology

FINAL REPORT ON THE START PROGRAMME

New validation of Geant4-DNA physics for accurate modeling of accelerated charged particle interactions in biological media

Supervisor:
Dr. Batmunkh Munkhbaatar

Student:
Matvei Vologzhin, Russia
Tomsk State University

Participation period:
August 03 – September 27,
Summer Session 2025

Dubna, 2025

Contents

Abstract

Ion components in Galactic and Solar cosmic rays cause various radiation effects and health risks for astronauts during space missions. Reliable Monte Carlo simulations, like those performed with the Geant4 toolkit, are essential for accurately modeling ion transport and interactions at nanometer scales within the space radiation environment. Geant4 is widely utilized in space and medical applications, including hadrontherapy and heavy-ion irradiation studies.

Recent Geant4 physics models for ion transport have undergone significant updates and validation. This work presents a new validation of several Geant4 ion physics models, including our dynamic charge approach, by comparing simulated and measured radial dose distributions.

Some clustering algorithms have been considered in the context of secondary particles, namely K-Means, a modified version of K-Means using a genetic algorithm, and DBSCAN, which can be used for modeling DNA breaks.

Introduction

Currently, humanity is actively planning and implementing many space missions, particularly those involving humans. Such missions carry numerous risks, ranging from single-event upsets (SEUs) to the effects of radiation on humans. The cause of these risks is cosmic radiation. But how exactly does cosmic radiation affect the human body?

Cosmic radiation consists of approximately 85% protons, 10% alpha particles, 4-5% heavier ions, and the remaining 1% are electrons and positrons. However, in this work, the primary object of study is electrons, which might seem illogical, as their contribution to cosmic radiation is insignificant.

In fact, we are interested not in electrons of cosmic origin (primary radiation), but in delta-electrons (secondary radiation) that arise when ions pass through a medium, particularly water, which is the main chemical component of human cells (60-80% depending on the total cell mass).

Biological systems are inherently complex, and a complete theory of the effects of ionizing radiation on biological structures, representing a full set of analytical expressions, is absent. For this reason, the Geant4 software package will be used to investigate the effects of delta-electrons on an aqueous medium.

Geant4 (***Ge****ometry* ***an****d* ***t****racking*) is a software toolkit based on the Monte Carlo mathematical method, allowing simulations of various particles passing through matter. This software package is used in a wide range of tasks and projects. Within JINR, it is actively used and developed, for example, to describe the SPD detector at the NICA accelerator based on the GeoModel library. It is also used for modeling the effects of ionizing radiation on biological structures of various organizational levels at the Laboratory of Radiation Biology (LRB) [1, 2, 4, 11], with whose support this work was created.

Within the framework of the current work, two scientific and one educational goal are set:

1. (Sc.) Validation of the radial dose distribution originating from delta-electrons using the Geant4 software package version 11.3 (ref08)

2. (Sc.) Validation of Linear energy transfer (LET) values also using Geant4 11.3 (ref08)

3. (Ed.) Application of three clustering methods: K-Means, Modified K-Means by Genetic algorithm, DBSCAN

The characterization of the last goal as educational is due to the fact that it does not carry scientific novelty but is important for the subsequent connection between physical models of energy deposition and models of DNA damage in the DNA molecule.

I Radial dose distribution

Problem Statement

Radial dose distribution can be understood as the dependence of the ratio of energy deposited in a cylinder layer to the mass of that layer (dose) on the radius of this cylindrical layer.

Such a distribution can be used to estimate the energy contribution of delta-electrons near the track of an incident ion. In fact, this distribution helps to understand the damage radius of a certain particle with a specific energy in the context of human radiation safety.

To provide such estimates, a tool capable of performing such simulations is required. The Geant4 software package, which was mentioned in the introduction, can serve as a potential tool. However, before using it, it is necessary to validate the physical models used in Geant4. More precisely, it is necessary to validate the *physics lists*.

Physics Lists in Geant4

Geant4 offers the capability to simulate processes such as: particle transport, electromagnetic and hadronic processes, decay, and optical phenomena. For any given task, a specific set of particles and processes, described by a certain collection of models, is chosen. Such a combination is called a *physics list*.

For example, in the context of our task, we are interested in electromagnetic physics lists, as they enable the modeling of the ionization process, which is the source of delta-electrons.

Among all electromagnetic physics lists, the following five are investigated in this work: EM Standard Option 4, DNA Option 2/4/6/8. This choice is due to the following reasons:

1. To confirm the necessity of developing DNA physics lists: standard physics lists (EM Standard Option 4) are currently not adapted to model ionization losses at low energies and small scales.

2. To evaluate and compare the offered simulation quality for different DNA physics lists

The difference between the proposed physics lists is expressed not only in different models and their associated particles but also in the approach on which the EM Standard and DNA lists are based. EM Standard Physics lists are developed based on the Condensed History approach, while DNA Physics Lists are based on the Detailed Track Structure approach. Let's consider the content of these physics lists in more detail.

Models used in EM Standard Physics lists are presented in Table 1. And models used in DNA Physics lists (for options 2,4,8) are presented in Table 2.

Thus, each physics list has its own features in the form of a list of unique models. Two physics lists are of particular interest: DNA option 6 & DNA option 8[1].

The interest in the first case is due to the fact that the models used in this configuration of the DNA physics list are specifically designed for electron modeling. The second case is interesting due to its

[1]Based on Rudd Dynamic model

Model	Particle	Based on data
G4BetheBlochModel	default high-energy model for heavy particles	–
G4LindhardSorensenModel	using for ions Z > 2	ICRU73 & ICRU90
G4BraggModel	for protons, backup for all ions, effective charge for ions	ICRU90 & PSTAR
G4BraggIonModel	for alpha-particles	ICRU90 & ASTAR

Table 1 – Models in EM Standard Physics list

<table>
<tr><th rowspan="2">Energy</th><th colspan="3">Particle</th></tr>
<tr><th>Protons</th><th>Alpha-type</th><th>Other Ions</th></tr>
<tr><td>$E < 0.5$ MeV/amu</td><td>Rudd model</td><td colspan="2" rowspan="3">Rudd model</td></tr>
<tr><td>$0.5 < E < 100$ MeV/amu</td><td>Born model</td></tr>
<tr><td>$100 < E < 300$ MeV/amu</td><td>RPWA model</td></tr>
<tr><td>$E > 300$ MeV/amu</td><td colspan="3">Standard models</td></tr>
</table>

Table 2 – Models in DNA Physics list (2, 4, 8)

novelty: a new concept in Geant4 modeling, *dynamic charge*, is used, which will be tested for protons and alpha particles[2].

Validation

In this work, the validation of ionization models is carried out through the radial dose distribution, represented by relation (1),

$$\text{Dose} = \frac{E}{\rho \cdot V}, \tag{1}$$

where [Dose] = Gy is the dose in Grays, $[E]$ = keV is the energy in keV, $\rho = 1.000$ kg/m^3 is the density of water, and $[V] = \text{m}^3$ is the volume of the cylindrical layer in cubic meters.

Validation of Linear Energy Transfer (LET) is also performed through the ratio of the total energy deposited in the cylinder to the length of the investigated cylinder. The formal expression can be seen in relation (2).

$$\text{LET} = \frac{E}{l}, \tag{2}$$

where $[E]$ = keV is the total energy deposited in the water cylinder, $[l] = \mu$m is the actual length of the cylinder.

It should be noted that the contribution of delta-electrons to LET can be calculated separately by using formula (3) from work [5] and the dose dependencies on the distance to the incident ion track, which will be provided below.

$$\text{LET}_{\text{e-}} = \int_0^{r_{\max}} 2\pi r \text{Dose}(r) \text{d}r, \tag{3}$$

where $r_{\max}$ is the maximum distance from the ion track where energy deposition occurs, r is the radius of an infinitely thin cylindrical layer, Dose(r) is the dependence of dose on the radius of the cylindrical layer.

The graphs obtained below allow for the validation of some Geant4 11.3 (ref08) physics lists based on experimental data taken from works [6, 12].

[2] As of version 11.3 (ref08), the *dynamic charge* concept has only been implemented for protons and alpha particles

Each graph is a composition of two graphs: the upper graph shows the absolute dose value depending on the distance to the track axis, and the lower graph shows the relative deviation of simulated data from experimental values[3]. The root mean square error (RMSE), calculated by formula (4), serves as a comparative metric.

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\eta_i^2}, \tag{4}$$

where $\eta_i = (1 - \text{Geant4}/\text{Experiment})$ is the relative deviation of simulated data from experimental values, N is the number of experimental values.

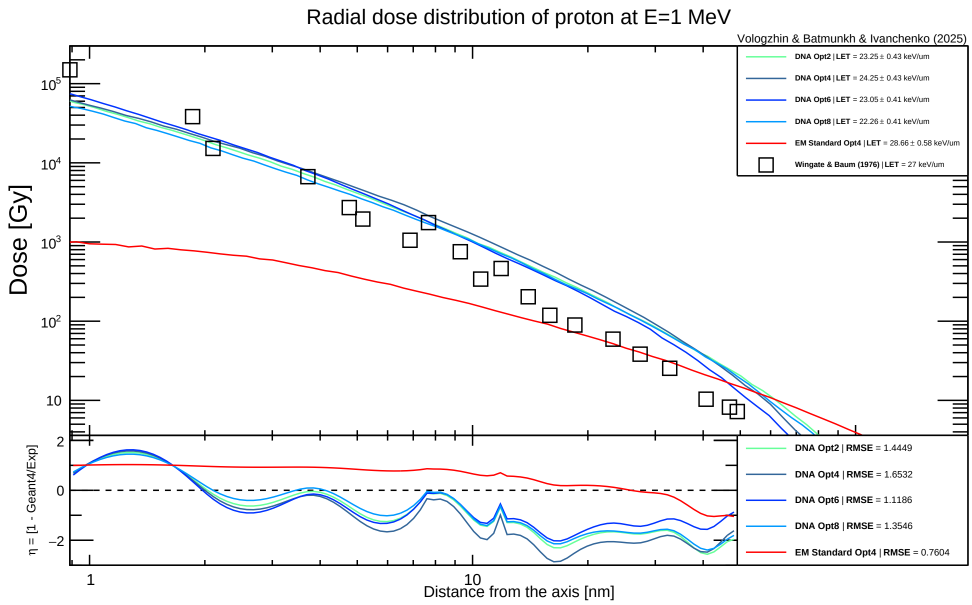

Figure 1 – Radial dose distribution for *proton* at energy $E = 1$ MeV

The first example considered is a proton with an energy $E = 1$ MeV, whose dose-radius dependence can be seen in Figure 1.

At first glance, it may seem that most of the experimental points lie near the curves characterizing DNA Physics Lists, but it is important to remember that logarithmic axes are used for the main graph. This serves as an additional reason for using a numerical metric, which in our case is RMSE.

Comparing LET values, it can be seen that the LET obtained using the Em Standard Physics List is closer to the experimental value, and also that the root mean square error (RMSE) has a smaller value in this case. This character of the curves is natural for the following reasons:

1. A significant part of the experimental points is located at a comparatively large distance, where DNA Physics Lists lose their advantage.

2. Em Standard Physics Lists use theoretical models developed specifically for protons, in particular, the Bragg Model[4].

[3]The deviation graph was obtained by cubic spline interpolation for greater clarity.
[4]In the case of a proton, there is no need to apply the *effective charge* concept, which introduces potential deviation in the case of comparatively heavier ions.

It follows that it makes sense to refine the experiment with a larger number of points at small distances to obtain a more accurate picture for comparing Em Standard and DNA Physics Lists. Nevertheless, Em Standard Option 4 yields good results, both in terms of LET and Radial Dose Distribution.

We also note that among the DNA Physics Lists, DNA Option 4 gives the LET value closest to the experimental one, however, the smallest deviation in the Radial Dose Distribution is observed for DNA Option 6, which, as a reminder, was specifically developed for electrons based on CPA-100 models.

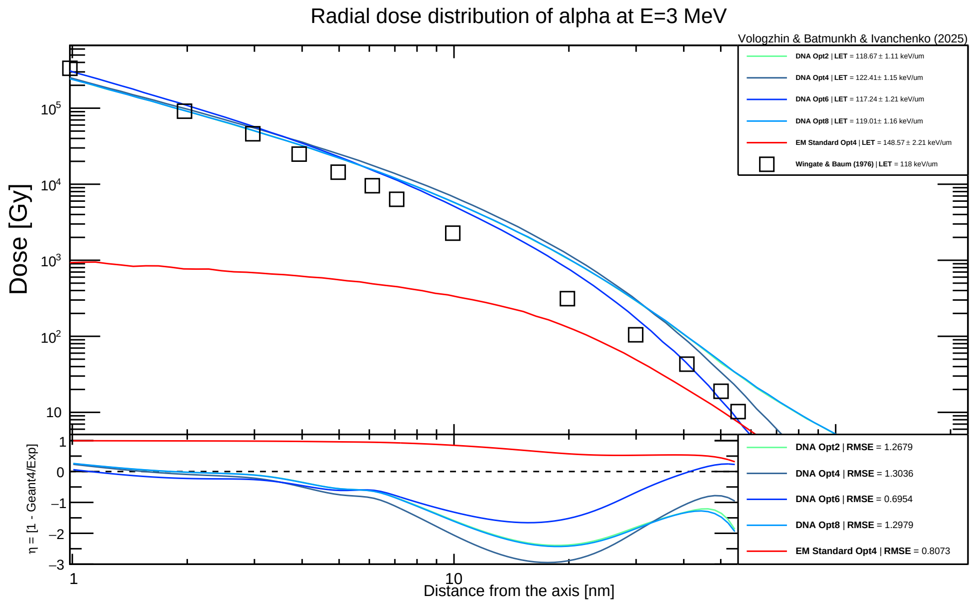

Figure 2 – Radial dose distribution for *alpha-particle* at energy $E = 3$ MeV

Em Standard Option 4 begins to lose its advantage starting with a heavier ion – an alpha particle, as seen in Figure 2. Despite the atomic number changing by only one unit compared to a proton (hydrogen atom nucleus), the absolute deviation of LET from the experimental value is as much as 20 units, which is significant when compared to the deviation of DNA Physics Lists, where the absolute deviation from the experimental value does not exceed 5 units. Moreover, if RMSE is considered, DNA Option 6 shows the best result in modeling the radial dose distribution.

In the case of carbon ions[5], presented in Figure 3, each of the physics lists shows a significant deviation from the experimental value, but Em Standard Option 4 still has the largest absolute deviation. Regarding RMSE, the situation is comparatively worse than in previous cases: now Em Standard Option 4 models only better than DNA Option 4, falling short of the DNA Physics Lists configurations. DNA Option 6 remains the best.

The last case is a ^{20}Ne, shown in Figure 4. The relative deviation of LET for the Standard Physics List is smaller than for the DNA Physics Lists, but RMSE ultimately falls short of all DNA Physics List configurations, and DNA Option 6 is no longer the best, yielding leadership to DNA Option 8.

Thus, we have shown that EM Standard Physics Option 4 can be used in exceptional cases: Radial dose distribution for protons and LET for heavy ions; in other cases, for more accurate modeling, DNA

[5]For the relative deviation plot in the case of ^{12}C and ^{20}Ne, linear interpolation is used, as cubic interpolation significantly alters the appearance of the picture, which may hinder clear representation.

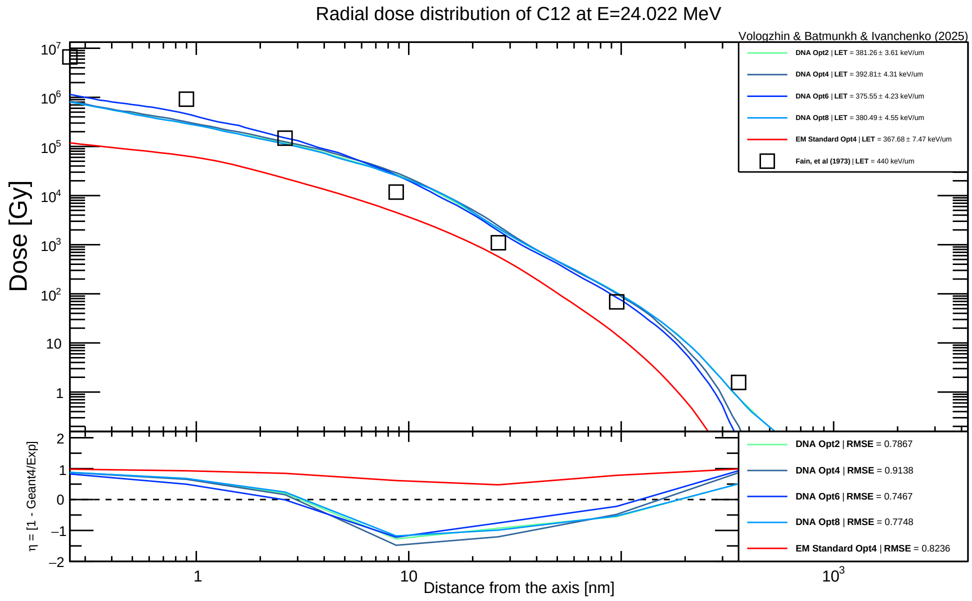

Figure 3 – Radial dose distribution for *carbon* (^{12}C) at energy $E = 24.022$ MeV

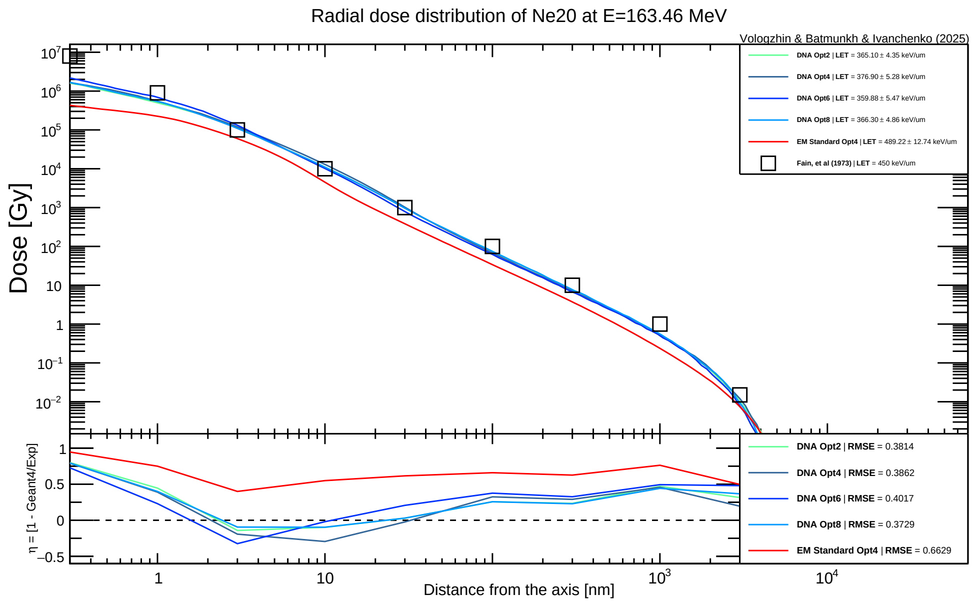

Figure 4 – Radial dose distribution for *Neon* (^{20}Ne) at energy $E = 163.46$ MeV

Physics Lists should be used, with DNA Option 6 being preferred in the context of delta-electrons. However, it is too early to draw conclusions about which physics list to use, as there is an opportunity to compare which physics list offers advantages in calculation speed, which can play a key role in some tasks.

We also note that an attempt was made to validate ^{20}Ne at an energy of $E = 377$ MeV/amu, whose experimental data could potentially be taken from work [10], but due to the length of calculations and time limitations, it was not completed, therefore, validation of ^{20}Ne was performed at a comparatively lower energy – $E = 163.43$ MeV.

Run Characteristics

The price of using the accuracy of DNA Physics Lists is calculation speed, which is due to the chosen approach – the Detailed Track Structure approach. For this reason, summary tables of runs are provided in this work, in which the following quantities can be seen:

1. Cylinder's sizes
2. Number of ions launched
3. Maximum energy of secondary electrons
4. Calculation time per particle
5. LET
6. Processes that occurred during the simulation

As well as other quantities, which the reader can familiarize themselves with in more detail independently.

Table 3 – Simulation data for **protons**

Physics List	Initial Energy	Cyl. Length	Cyl. Radius	Ions Num.	Fleet Out Ions	Mean Ion Steps	Mean e- Num.	Mean 2'ry e- Num.	Max 2'ry e- Energy	Fleet Out e-	Gamma Num.	Elapsed Time/particle (s)	LET (keV/μm)	Processes List
EM Standard Opt4	1 MeV	10 μm	5 mm	10000	10000	197	203	162	2.17484 keV	1967	24	0.001895	28.6598	Transportation eBrem eIoni hIoni msc
DNA Opt2				100	100	4523	15401	4527	2.14003 keV	1540151	4	0.061172	23.2363	Transportation e_G4DNAExcitation e_G4DNAIonisation proton_G4DNAIonisation
DNA Opt4						4554	12777	4558	2.15023 keV	1277789	4	0.107833	24.2239	Transportation e_G4DNAExcitation e_G4DNAIonisation proton_G4DNAIonisation
DNA Opt6						4541	15223	4546	2.16887 keV	1522305	9	0.081007	23.0523	Transportation e_G4DNAIonisation e_G4DNAExcitation e_G4DNAIonisation
DNA Opt8						4881	14878	4883	2.14728 keV	1487823	3	0.056844	22.2686	Transportation e_G4DNAExcitation e_G4DNAIonisation proton_G4DNAIonisation

Table 4 – Simulation data for **alpha particles**

Physics List	Initial Energy	Cyl. Length	Cyl. Radius	Ions Num.	Fleet Out Ions	Mean Ion Steps	Mean e- Num.	Mean 2'ry e- Num.	Max 2'ry e- Energy	Fleet Out e-	Gamma Num.	Elapsed Time/particle (s)	LET (keV/μm)	Processes List
EM Standard Opt4	3 MeV	10 μm	5 mm	10000	10000	55	84	49	1.64471 keV	164	29	0.000569	148.566	Transportation eBrem eIoni ionIoni msc
DNA Opt2				100	100	26031	79593	26047	1.6109 keV	7959380	13	0.396983	118.635	Transportation alpha_G4DNAIonisation e_G4DNAExcitation e_G4DNAIonisation
DNA Opt4						26040	66647	26055	1.63177 keV	6664707	8	0.582651	122.386	Transportation alpha_G4DNAIonisation e_G4DNAExcitation e_G4DNAIonisation
DNA Opt6						26020	78008	26037	1.61464 keV	7800872	13	0.331357	117.182	Transportation alpha_G4DNAIonisation e_G4DNAExcitation e_G4DNAIonisation
DNA Opt8						26019	79783	26034	1.61095 keV	7978306	13	0.381380	118.976	Transportation alpha_G4DNAIonisation e_G4DNAExcitation e_G4DNAIonisation

In each case, it is evident that the EM Standard Physics List has a significant advantage in calculation speed – almost 40 times compared to any DNA Physics List, but among them, DNA Option 6 holds the leadership. Such an advantage is characteristic only for this task: in previous work by M. Vologzhin & V. Ivanchenko, whose results were presented at the Geant4-DNA collaboration meeting and at the XXII Lomonosov Conference[6], it was shown that in the task of calculating Range & Detour factor, DNA Option 8, based on the aforementioned *Dynamic charge* concept, has the shortest calculation time.

[6]The results will be published in the journal *Moscow University Physics Bulletin* in 2026.

Table 5 – Simulation data for ^{12}**C** particles

Physics List	Initial Energy	Cyl. Length	Cyl. Radius	Ions Num.	Fleet Out Ions	Mean Ion Steps	Mean e- Num.	Mean 2'ry e- Num.	Max 2'ry e- Energy	Fleet Out e-	Gamma Num.	Elapsed Time/particle (s)	LET (keV/μm)	Processes List
EM Standard Opt4	24.022 MeV	10 μm	5 mm	10000	10000	2989	3042	2645	4.39489 keV	17893	566	0.023963	367.675	Transportation eBrem eIoni ionIoni msc
DNA Opt2				100	100	80873	253174	81068	4.3523 keV	25317438	99	1.392444	381.126	GenericIon_G4DNAIonisation Transportation e_G4DNAExcitation e_G4DNAIonisation
DNA Opt4						80828	213010	81020	4.33517 keV	21301029	128	2.212260	392.662	GenericIon_G4DNAIonisation Transportation e_G4DNAExcitation e_G4DNAIonisation
DNA Opt6						80881	247760	81073	4.32766 keV	24776089	114	1.364502	375.464	GenericIon_G4DNAIonisation Transportation e_G4DNAExcitation e_G4DNAIonisation
DNA Opt8						80837	252734	81028	4.357 keV	25273486	127	1.413996	380.466	GenericIon_G4DNAIonisation Transportation e_G4DNAExcitation e_G4DNAIonisation

Table 6 – Simulation data for ^{20}**Ne** particles

Physics List	Initial Energy	Cyl. Length	Cyl. Radius	Ions Num.	Fleet Out Ions	Mean Ion Steps	Mean e- Num.	Mean 2'ry e- Num.	Max 2'ry e- Energy	Fleet Out e-	Gamma Num.	Elapsed Time/particle (s)	LET (keV/μm)	Processes List
EM Standard Opt4	163.46 MeV	10 μm	5 mm	10000	10000	5771	5836	4633	18.0161 keV	106318	3378	0.048847	489.224	Rayl Transportation compt eBrem eIoni ionIoni msc
DNA Opt2				100	100	73078	241559	73397	17.736 keV	24155902	224	1.219458	365.191	GenericIon_G4DNAIonisation Transportation e_G4DNAExcitation e_G4DNAIonisation
DNA Opt4						73021	203914	73339	17.8985 keV	20391450	212	1.885818	376.818	GenericIon_G4DNAIonisation Transportation e_G4DNAExcitation e_G4DNAIonisation
DNA Opt6						73015	236506	73333	17.8451 keV	23650674	166	1.181640	360.037	GenericIon_G4DNAIonisation Transportation e_G4DNAExcitation e_G4DNAIonisation
DNA Opt8						73048	242311	73368	17.8348 keV	24231189	246	1.284894	366.383	GenericIon_G4DNAIonisation Transportation e_G4DNAExcitation e_G4DNAIonisation

Run Parameters

The authors of this work support the principle of result reproducibility, so we provide the run parameters that were used in our Geant4 application, so that every Geant4 user can replicate the results of our work. However, please note that some commands may not work in your application, for example, */physics/setPhysics G4EmDNAPhysics_option2*, as they are custom for our application, but examples of implementing such macro commands can be found in many examples provided by Geant4 itself. For instance, the "radial" example, which is located in the Geant4 installation directory at the following path:

.../install_folder/share/Geant4/examples/extended/medical/dna/radial/

Table 7 – Geant4 macro command configurations for various physics options

DNA Opt 2

```
# Cuts
/run/setCut 1 nm
/cuts/setLowEdge 100 eV
/process/em/lowestElectronEnergy 100 eV

# Ion parameters
/process/em/lowestMuHadEnergy 20 eV
/process/em/lowestElectronEnergy 20 eV
/process/eLoss/StepFunctionIons 0.1 10 nm

# Activate some process
/process/em/fluo true
/process/em/auger true
/process/em/augerCascade true
/process/em/deexcitationIgnoreCut true

#### Choosing physics ####
/physics/setPhysics G4EmDNAPhysics_option2

#### Run ####
/run/initialize

#### Simplifications ####
/process/inactivate e-_G4DNAElastic
/process/inactivate e-_G4DNAVibExcitation
/process/inactivate e-_G4DNAAttachment
/process/inactivate e-_G4DNAElectronSolvation
/process/inactivate alpha_G4DNAExcitation
/process/inactivate alpha_G4DNAElastic
/process/inactivate alpha_G4DNAChargeDecrease
/process/inactivate proton_G4DNAExcitation
/process/inactivate proton_G4DNAElastic
/process/inactivate proton_G4DNAChargeDecrease
```

DNA Opt 4

```
# Cuts
/run/setCut 1 nm
/cuts/setLowEdge 100 eV
/process/em/lowestElectronEnergy 100 eV

# Ion parameters
/process/em/lowestMuHadEnergy 20 eV
/process/em/lowestElectronEnergy 20 eV
/process/eLoss/StepFunctionIons 0.1 10 nm

# Activate some process
/process/em/fluo true
/process/em/auger true
/process/em/augerCascade true
/process/em/deexcitationIgnoreCut true

#### Choosing physics ####
/physics/setPhysics G4EmDNAPhysics_option4

#### Run ####
/run/initialize

#### Simplifications ####
/process/inactivate e-_G4DNAElastic
/process/inactivate e-_G4DNAElectronSolvation
/process/inactivate alpha_G4DNAExcitation
/process/inactivate alpha_G4DNAElastic
/process/inactivate alpha_G4DNAChargeDecrease
/process/inactivate proton_G4DNAExcitation
/process/inactivate proton_G4DNAElastic
/process/inactivate proton_G4DNAChargeDecrease
```

DNA Opt 6

```
# Cuts
/run/setCut 1 nm
/cuts/setLowEdge 100 eV
/process/em/lowestElectronEnergy 100 eV

# Ion parameters
/process/em/lowestMuHadEnergy 20 eV
/process/em/lowestElectronEnergy 20 eV
/process/eLoss/StepFunctionIons 0.1 10 nm

# Activate some process
/process/em/fluo true
/process/em/auger true
/process/em/augerCascade true
/process/em/deexcitationIgnoreCut true

#### Choosing physics ####
/physics/setPhysics G4EmDNAPhysics_option6

#### Run ####
/run/initialize

#### Simplifications ####
/process/inactivate e-_G4DNAElastic
/process/inactivate e-_G4DNAElectronSolvation
/process/inactivate alpha_G4DNAExcitation
/process/inactivate alpha_G4DNAElastic
/process/inactivate alpha_G4DNAChargeDecrease
/process/inactivate proton_G4DNAExcitation
/process/inactivate proton_G4DNAElastic
/process/inactivate proton_G4DNAChargeDecrease
```

DNA Opt 8

```
# Cuts
/run/setCut 1 nm
/cuts/setLowEdge 100 eV
/process/em/lowestElectronEnergy 100 eV

# Ion parameters
/process/em/lowestMuHadEnergy 20 eV
/process/em/lowestElectronEnergy 20 eV
/process/eLoss/StepFunctionIons 0.1 10 nm

# Activate some process
/process/em/fluo true
/process/em/auger true
/process/em/augerCascade true
/process/em/deexcitationIgnoreCut true

#### Choosing physics ####
/physics/setPhysics G4EmDNAPhysics_option8

#### Run ####
/run/initialize

#### Simplifications ####
/process/inactivate e-_G4DNAElastic
/process/inactivate e-_G4DNAVibExcitation
/process/inactivate e-_G4DNAAttachment
/process/inactivate e-_G4DNAElectronSolvation
/process/inactivate alpha_G4DNAExcitation
/process/inactivate alpha_G4DNAElastic
/process/inactivate alpha_G4DNAChargeDecrease
/process/inactivate proton_G4DNAExcitation
/process/inactivate proton_G4DNAElastic
/process/inactivate proton_G4DNAChargeDecrease
```

Em Standard Opt 4

```
# Cuts
/run/setCut 1 nm
/cuts/setLowEdge 100 eV
/process/em/lowestElectronEnergy 100 eV

# Ion parameters
/process/em/lowestMuHadEnergy 20 eV
/process/em/lowestElectronEnergy 20 eV
/process/eLoss/StepFunctionIons 0.1 10 nm

# Activate some process
/process/em/fluo true
/process/em/auger true
/process/em/augerCascade true
/process/em/deexcitationIgnoreCut true

#### Choosing physics ####
/physics/setPhysics G4EmStandardPhysics_option4

#### Run ####
/run/initialize
```

II Clustering algorithms

Clustering is a tool that links Geant4 simulations with experimental damage data. Clustering algorithms can be used for two purposes:

1. Grouping energy deposition events into clusters to estimate the spatial distribution of potential DNA damage foci

2. Grouping Single strand breaks (SSB) to classify damage into types such as: SSB (a cluster of one element), DSB (a cluster of two elements, each lying on different strands of the DNA molecule no further than 10 nm[7]), cDSB (complex breaks, i.e., clusters consisting of more than two SSBs within the same 10 nm).

Different clustering algorithms can be used to solve these problems, each having an advantage in one task or another. However, in this work, the following will be considered: K-Means (Classic), K-Means (Genetic algorithm), DBSCAN.

K-Means (Classic)

The K-Means algorithm is a single-parameter algorithm where the parameter is the number of clusters K. Such an algorithm is less efficient among those presented due to the following facts:

1. There is no specific physical justification for choosing the number of clusters: for K, one can choose the number of ion steps, the number of secondary electrons, or the total number of electrons formed during the simulation.

2. The convergence of the algorithm to a global extremum is not guaranteed: the result of the algorithm strongly depends on the choice of initial centroids.

Despite this, the algorithm is quite fast, which makes it an excellent educational example of a clustering algorithm.

For an example of using this algorithm, one can refer to Figure 5. Block diagrams and the algorithm itself are not provided in this work for the sake of compactness; however, they can be found in works [3,8,9].

K-Means (Genetic algorithm)

The problem mentioned in the previous subsection regarding finding a global extremum depending on the initial position of centroids can be partially mitigated using a Genetic algorithm – we modify the K-Means algorithm using a Genetic algorithm.

[7] Pitch of the primary DNA structure helix.

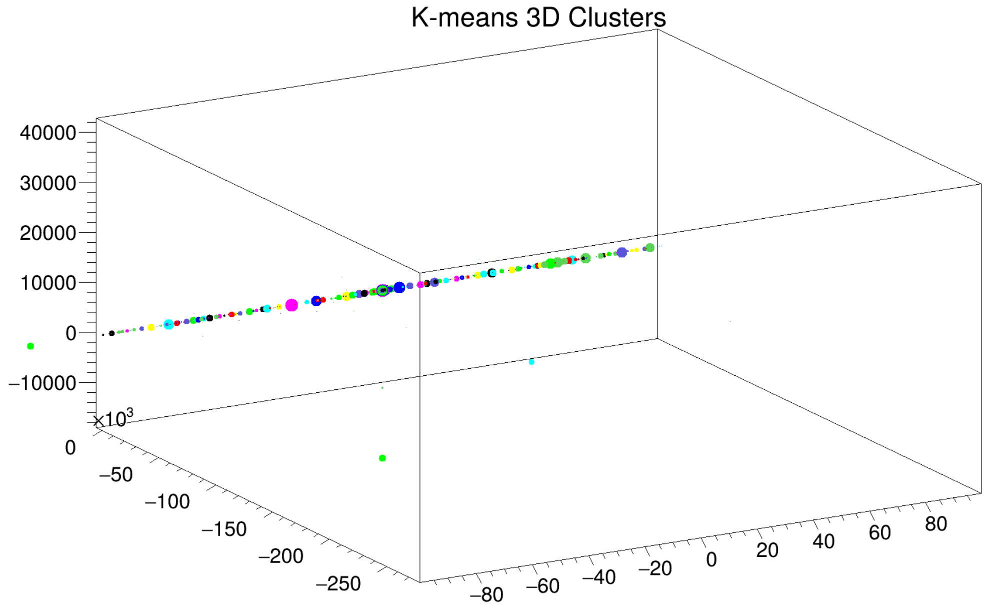

Figure 5 – Clustering of an alpha particle track with $K = 400$, determined by the number of δ-electrons ionized during passage through a cylinder of thickness $h = 100$ nm

The essence of the genetic algorithm lies in using the following biological concepts: selection, crossover, and mutation. Thanks to them, it becomes possible to reach a global extremum by repeatedly applying the genetic algorithm in each "generation". This algorithm can be studied in more detail in work [8].

The uniqueness of this algorithm lies in its greater parametric freedom to avoid getting stuck in local extrema due to the mentioned selection, crossover, and mutation procedures.

One example of applying the genetic algorithm[8] is shown in Figure 6. In our case, the implementation for K-Means modified by Genetic algorithm was as follows:

1. Shuffling all data from Ntuple

2. Create I individuals in initial generation.

3. **Selection**: We evaluate the width of each cluster in each individual using the fitness function F:

$$F = \sum_{k=1}^{K} \sum_{n=1}^{N_k} |\vec{c}_k - \vec{r}_{nk}| \, ,$$

   where N_k is the number of points in the k-th cluster, K is the number of clusters fixed at the beginning of the program, $\vec{c}_k$ is the k-th centroid, $\vec{r}_{nk}$ is the n-th point that belongs to k-th centroid.
   Individuals with the lowest fitness function value are discarded from the algorithm by a threshold value $p = 0.05$.

4. **Crossover**: If the data is represented as a linear chain, a pair of individuals exchange points up to a certain point, and after that, the data do not change. If an individual does not find a pair, it is

[8]Startup parameters are specified in the caption to Figure.

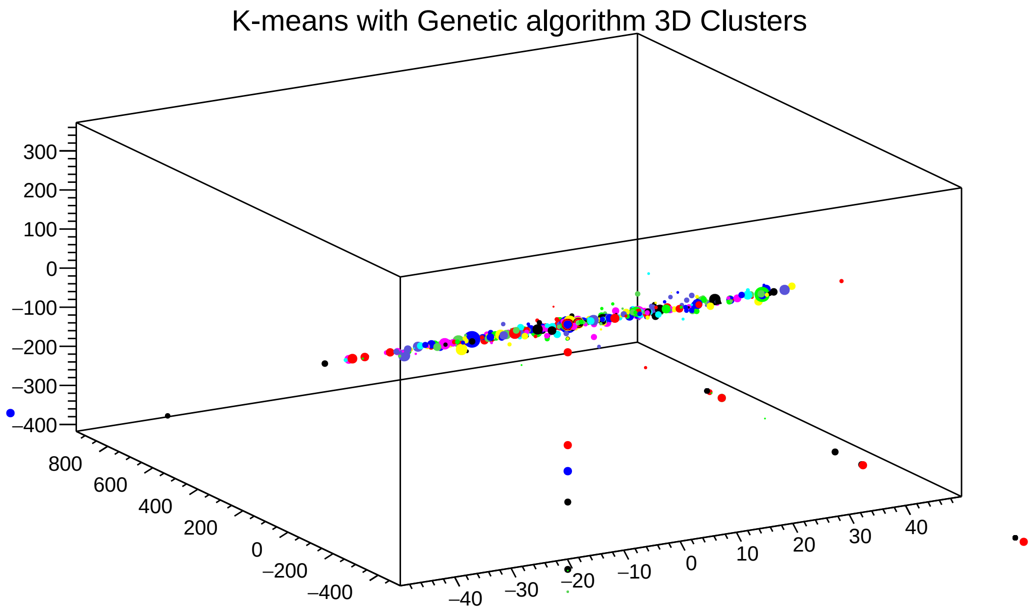

Figure 6 – Clustering of an alpha particle track using the KMGA method with $K = 400$ and $I = 10$, determined by the number of δ-electrons ionized during passage through a cylinder of thickness $h = 100$ nm

discarded.

5. **Mutation**: A random point changes its affiliation from one cluster to another, number of mutations $M = 1$.

6. The algorithm repeats from step 2 until only one individual or a pair of individuals remains, among which the one with the highest fitness value is chosen.

In addition, code was implemented to estimate the ratio of the number of clusters outside a certain effective radius to the number of clusters that are inside the effective radius, which is interpreted as the number of clusters corresponding to the number of secondary electron energy depositions, to the number of clusters corresponding to the number of energy depositions of the incident ion. The dependence of this ratio on the effective radius is shown in Figure 7. Upon obtaining the graphs, it was found that ROOT has some problem with the correct display of points on a three-dimensional graph. With a large number of densely packed points, some points are "thrown out" of the graph, which can be seen in Figure 6.

DBSCAN

The last algorithm considered, which is regularly used in the context of Geant4-DNA, is DBSCAN[9]. This is a single-parameter clustering algorithm that is based on forming clusters using a certain effective radius r_{eff}. It starts from a random point and checks for neighbors within the effective radius; if so, the algorithm continues within this cluster, and if not, it moves to the next random point that the algorithm has not yet visited.

As before, a detailed description of the algorithm can be found in work [7].

[9]Density-Based Spatial Clustering of Applications with Noise

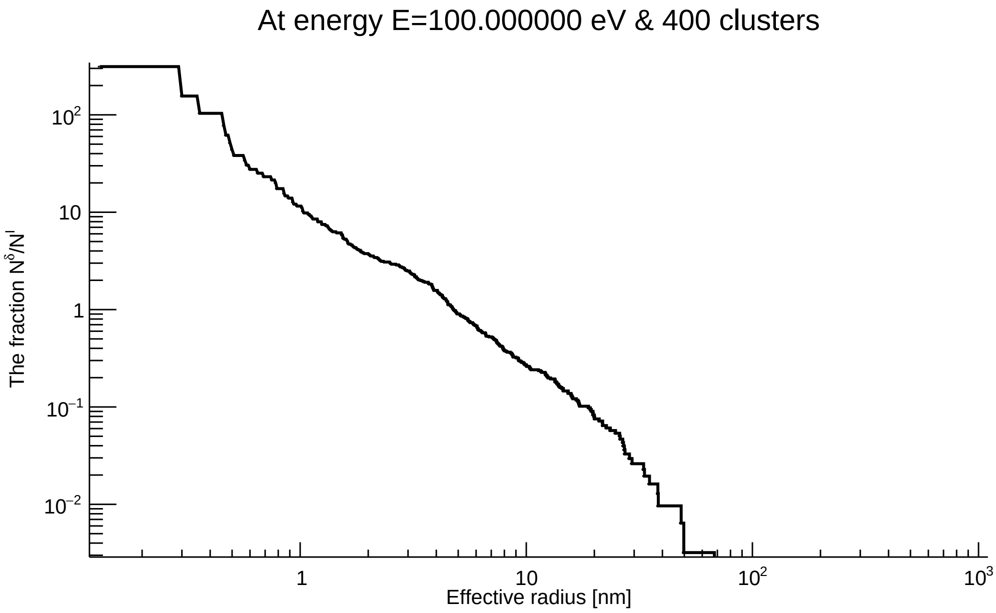

Figure 7 – Ratio of the number of clusters outside the effective radius to the number of clusters inside the effective radius as a function of effective radius for an energy cut of events $E_{\text{cut}} = 100$ eV in the case of $K = 400$ clusters

The results of the algorithm for $N = 10$ alpha particles with a cylinder length $L = 10\ \mu$m are presented in Figure 8: In Figure 8, we interpret black points as SSB (single strand breaks), red points as DSB (double strand breaks), and blue points as cDSB (complex double strand breaks).

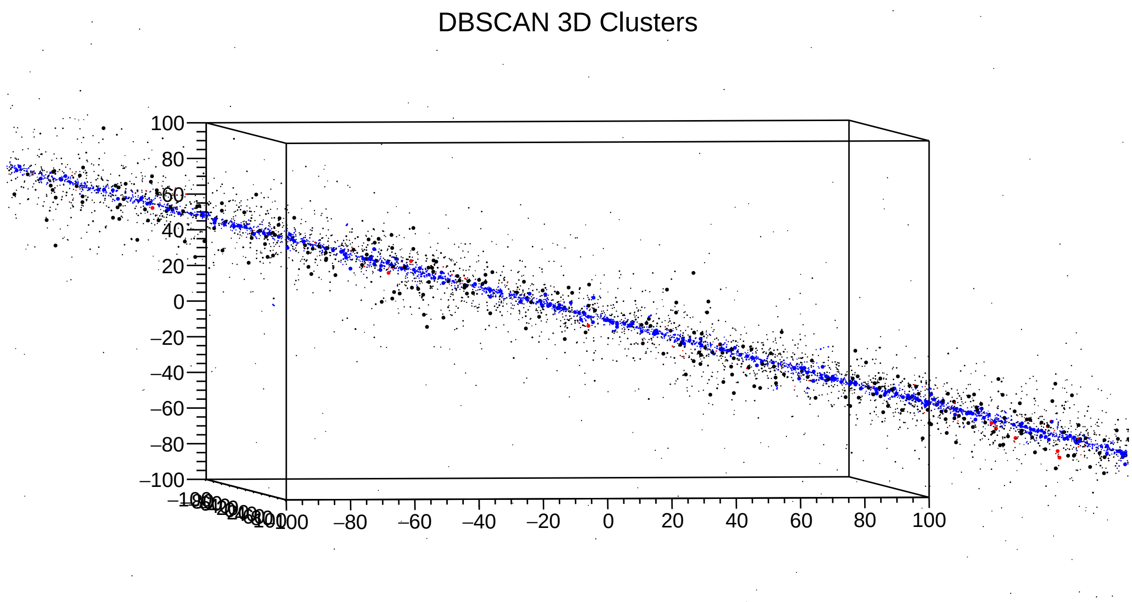

Figure 8 – Clustering of an alpha particle track using the DBSCAN method with $r_{\text{eff}} = 3.2$ nm – distance between a pair of neighboring nucleotides in a DNA molecule, $N_{\text{min}} = 2$ – minimum number of points in a cluster

Conclusion

During the two-month START internship at the Laboratory of Radiation Biology, the following results were obtained:

1. Validation of Radial dose distribution in the context of Geant4 version 11.3 (ref08) for the following ions and corresponding energies: proton ($E = 1$ MeV), alpha-particle ($E = 3$ MeV), ^{12}C ($E = 24.022$ MeV), ^{20}Ne ($E = 163.46$ MeV). The physics lists considered were: Em Standard Option 4, DNA Option 2/4/6/8.

2. Validation of LET for the same list of ions and physics lists.

3. An evaluation of the calculation time for each of the considered physics lists was performed.

4. All run commands are provided for the reader to verify the results.

5. The following clustering algorithms were mastered: K-Means (Classic), K-Means (Genetic algorithm), DBSCAN.

6. The ratio of the number of clusters outside and inside the effective radius as a function of the effective radius was obtained for K-Means (Genetic algorithm).

The obtained results will be discussed at the annual meeting of the Geant4 collaboration and the quarterly meeting of the Geant4 electromagnetic group.

We note the importance of the completed internship, as it provided an opportunity for collaboration between employees of the Laboratory of Radiation Biology (LRB) JINR and the Laboratory of High Energy Physics Data Analysis (LHEPDA) TSU to study the effects of cosmic radiation on biological structures.

The intern plans to continue interaction with the presented scientific supervisor of the internship within the Interest project.

References

1. Batmunkh, M., Bayarchimeg, L., Bugay, A.N. Mathematical modeling of radiation-induced effects in structures of the central nervous system under the action of heavy charged particles. JINR Preprint P11-2024-49, 2024. Dubna: Joint Institute for Nuclear Research, 2024.

2. Batmunkh, M., Bayarchimeg, L., Bugay, A.N., Lkhagva, O. Computer simulation of radiation damage mechanisms in the structure of brain cells // PROCEEDINGS OF THE 24TH INTERNATIONAL SCIENTIFIC CONFERENCE OF YOUNG SCIENTISTS AND SPECIALISTS (AYSS-2020). Dubna, 2021. P. 050001. DOI: 10.1063/5.0063370.

3. Batmunkh, M., Bayarchimeg, L., Lkhagva, O., Belov, O. Cluster analysis of HZE particle tracks as applied to space radiobiology problems // Physics of Particles and Nuclei Letters. 2013. Vol. 10, No. 7, pp. 854–859. DOI: 10.1134/s1547477113070075.

4. Bulanova, T.S., Boreyko, A.V., Zadneprianetc, M.G., Krasavin, E.A., Kulikova, E.A. Formation of DNA Double-Strand Breaks in Rat Brain Neurons after Irradiation with Krypton Ions (78Kr) // Physics of Particles and Nuclei Letters. 2019. Vol. 16, No. 4, pp. 402–408. DOI: 10.1134/s1547477119040083.

5. Cucinotta, F.A., Wilson, J.W., Dubey, R.R., Katz, R. Radial dose distributions in the delta-ray theory of track structure // Two-center effects in ion–atom collisions: A symposium in honor of M. Eugene Rudd. Lincoln, NE (USA), 1996. (AIP Conference Proceedings. Vol. 362). Pp. 245–265. DOI: 10.1063/1.50083.

6. Fain, J., Monnin, M., Montret, M. Spatial Energy Distribution around Heavy-Ion Path // Radiation Research. 1974. Vol. 57, No. 3, pp. 379–389.

7. Francis, Z., Villagrasa, C., Clairand, I. Simulation of DNA damage clustering after proton irradiation using an adapted DBSCAN algorithm // Computer Methods and Programs in Biomedicine. 2011. Vol. 101, No. 3, pp. 265–270. DOI: 10.1016/j.cmpb.2010.12.012.

8. Malki, A.A., Rizk, M.M., El-Shorbagy, M.A., Mousa, A.A. Hybrid Genetic Algorithm with K-Means for Clustering Problems // Open Journal of Optimization. 2016. Vol. 5, No. 2, pp. 71–83. DOI: 10.4236/ojop.2016.52009.

9. Nikjoo, H., Uehara, S., Wilson, W.E., Hoshi, M., Goodhead, D.T. Track structure in radiation biology: theory and applications // International Journal of Radiation Biology. 1998. Vol. 73, No. 4, pp. 355–364.

10. Varma, M.N., Baum, J.W. Energy Deposition in Nanometer Regions by 377 MeV/Nucleon ^{20}Ne Ions // Radiation Research. 1980. Vol. 81, pp. 355–363.

11. Vasil'eva, M., Dushanov, E., Bugay, A. MODELING OF DNA DAMAGE REPAIR INDUCED BY HEAVY IONS IN MAMMALIAN CELLS // Russian Journal of Biological Physics and Chemistry. 2022. Vol. 7, No. 4, pp. 557–564. DOI: 10.29039/rusjbpc.2022.0560.

12. Waligorski, M.P.R., Hamm, R.N., Katz, R. The radial distribution of dose around the path of a heavy ion in liquid water // Nuclear Tracks and Radiation Measurements. 1986. Vol. 11, No. 6, pp. 309–319.