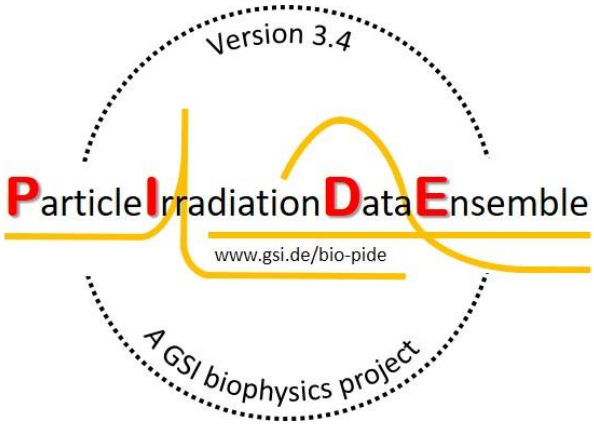

Particle Irradiation Data Ensemble: Information sheet

This document is a brief introduction to the particle irradiation data ensemble (PIDE) provided by GSI. In particular, it refers to the updated version PIDE3.x, while for the previous version PIDE2.0 a separate document is available (the running number x counts the technical updates, while the database structure is preserved). All users of PIDE should read this document before working with the database, as some important notes are given here for its usage in radiobiological research.

PIDE contains results of in-vitro cell survival experiments, gathered in a literature survey, which were performed pairwise after irradiation with photons and some ion species. The dose response curves are expressed by the linear-quadratic (LQ) parameters. In contrast to the originally released version PIDE2.0, the updated version PIDE3.x provides also raw data of experiments (doses and survival levels). Note that a number of bugs in PIDE2.0 were fixed in the updated version 3.x (including wrong entries) – a list of changes is available upon request by the developers (see Contact). The following publications cover the original setup of PIDE and its extension towards versions PIDE3.x, respectively:

Friedrich T, Scholz U, Elsässer T, Durante M, Scholz M. *Systematic analysis of RBE and related quantities using a database of cell survival experiments with ion beam irradiation.* J Radiat Res. 2013 May;54(3):494-514. doi: 10.1093/jrr/rrs114.

Friedrich T, Pfuhl T, Scholz M. *Update of the particle irradiation data ensemble (PIDE) for cell survival.* J Radiat Res. 2021 Jul 10;62(4):645-655. doi: 10.1093/jrr/rrab034.

Content

1. Structure of PIDE

The large collection of data in PIDE allows investigating specific questions regarding RBE with a better accuracy than when looking at one experiment only. However, at the same time the pooling may hide trends in the data, if experiments of different conditions are combined. To distinguish the different experimental conditions a number of additional parameters are hence listed in the PIDE for each experiment, which allows to work with restricted subsets.

GSI made PIDE accessible to the research community. The PIDE- package contains the following files:

i) Information_sheet.pdf
This document.

ii) PIDE3.x.xlsx
This Excel file contains all the data of PIDE except the raw survival data. Each pair of survival experiments (for ions and photons) is listed in one line. If you cannot read the .xlsx file, an ASCII file is available on request (see Contact).

In PIDE 3.x, the columns contain the data as described in the following table. Entries containing "N/A" (not applicable) indicate that the corresponding information is not available.

Column	Quality	Content
1 (A)	#ExpID	Running number labelling the database entry
2 (B)	#Publication	Running number labelling the publication
3 (C)	PublictionName	Short name of the publication, containing first author and year of publication
4 (D)	#IonExp	Running number labelling the ion irradiation experiments within a publication
5 (E)	Cells	Name of cell line
6 (F)	CellClass	Tumor cells (t) or normal cells (n)
7 (G)	CellOrigin	Human cells (h) or rodent cells (r)
8 (H)	CellCycle	Cell cycle phase (phases are given explicitly, or 'a' for 'asynchronous')
9 (I)	DNAcontent	Genomic length of diploid cells (in 10^9 bp, 5.6 for rodent and 6 for human cells)
10 (J)	PhotonRadiation	Photon radiation type: kVp energy for X-ray tubes, acceleration voltage for linear accelerators, isotope for gamma radiation
11 (K)	#PhotonExp	Running number labelling the corresponding photon dose response curve within a publication
12 (L)	Ion	Ion species
13 (M)	Charge	Charge of ions

14 (N)	IrradiationConditions	Irradiation modalities: monoenergetic (m) or spread out Bragg peak (s)
15 (O)	LET	Linear energy transfer in water (in keV/μm, for irradiation in spread out Bragg peak dose mean or track averaged LET)*
16 (P)	E	Specific energy of ions (in MeV/u), evaluated at the target*
17 (Q)	$\alpha_{X,\ Paper}$	Linear coefficient of LQ model (in Gy^{-1}) for response to photon reference radiation as given in the publication**
18 (R)	$\beta_{X,\ Paper}$	Quadratic coefficient of LQ model (in Gy^{-2}) for response to photon reference radiation as given in the publication**
19 (S)	$\alpha_{I,\ Paper}$	Linear coefficient of LQ model (in Gy^{-1}) for response to ion radiation as given in the publication**
20 (T)	$\beta_{I,\ Paper}$	Quadratic coefficient of LQ model (in Gy^{-2}) for response to ion radiation as given in the publication**
21 (U)	$\alpha_{X,\ Fit}$	Linear coefficient of LQ model (in Gy^{-1}) for response to photon reference radiation from fit to raw data in PIDE***
22 (V)	$\beta_{X,\ Fit}$	Quadratic coefficient of LQ model (in Gy^{-2}) for response to photon reference radiation from fit to raw data in PIDE***
23 (W)	$\alpha_{I,\ Fit}$	Linear coefficient of LQ model (in Gy^{-1}) for response to ion radiation from fit to raw data in PIDE***
24 (X)	$\beta_{I,\ Fit}$	Quadratic coefficient of LQ model (in Gy^{-2}) for response to ion radiation from fit to raw data in PIDE***

* Usually either the energy or LET were published. If both are given, they have been included in PIDE. If either E or LET are given, the counterpart was included in PIDE as calculated by stopping power computer codes like the GSI code ATIMA. In rare cases both LET and specific energy have been calculated for a given remaining range. For SOBPs authors give often either the track or dose mean LET, as indicated in the overview table. There, the corresponding energy was calculated by stopping power codes from the LET value and reflects the monoenergetic beam energy that would result in the same LET as the more complex SOBP radiation field.

**The LQ parameters are listed if they were published either numerically, shown in figures, or converted from given values of other parameterizations which can uniquely be converted into LQ values (e.g. from D_{10}, or D_0 or inactivation cross sections for exponential survival curves).

***The second set of LQ parameters reflects fits to the raw data contained in PIDE, which has been performed to all available datasets in the same way (automatized). There, three fit variants have been performed: A LQ, purely linear and purely quadratic fit to the raw data was evaluated without considering uncertainties of individual survival levels. The fits considered an offset term *c*, allowing for corrections of the plating efficiency based on all available survival information: For e.g. in the LQ fit, a model of type $\alpha D + \beta D^2 + c$ was fitted to the negative logarithm of the raw survival data. Hence the LQ, purely linear and purely quadratic fits were realized as 3-, 2- and 2-parameter fits, respectively. Finally, the best fit has been determined by the lowest reduced χ^2 value. Performing fits to raw data in exactly the same way justifies comparing the best fit results.

iii) References.pdf
This document contains all references from which the PIDE data originate. The reference equals the publication ID in the table above (i.e, [13] will label the 13th publication in PIDE, for example), and additional references related to data presented in one publication are labeled with attached letters to that ID starting from b (e.g. [13b]).

iv) Overview_table.pdf
This file contains a table which gives an overview over all publications respected in PIDE. It is also useful for finding publications covering specific experimental situations (e.g. experiments with neon, high LET, within SOBP…). Furthermore, the table contains some additional remarks on the way the data were retrieved from the original publication.

v) PIDE3.x_PhotonRawData.dat
This file contains raw data of photon dose response curves. Each dose response curve is reflected within a line in the document. A four column size header contains in ascending order the running number of the entry, the running number of the publication referred to (#Publication), the name of the publication (PublicationName), and the running number of photon experiments within that publication (#PhotonExp). The header is followed by the survival curve information, where pairs of doses and corresponding survival levels are simply appended alternatingly. If no raw data could be retrieved, the entries of the first pair of dose and survival were set to “N/A”.

vi) PIDE3.x_IonRawData.dat
This file contains raw data of ion dose response curves. Each dose response curve is reflected within a line in the document. A five column size header contains in ascending order the running number of the entry (#ExpID), the running number of the publication referred to (#Publication), the name of the publication (PublicationName), the running number of ion experiments within that publication (#IonExp), and the running number of photon experiments within that publication (#PhotonExp) that were used as reference experiments. The header is followed by the survival curve information, where pairs of doses and corresponding survival levels are simply appended alternatingly. If no raw data could be retrieved, the entries of the first pair of dose and survival were set to “N/A”.

2. Some words of caution

This section is to clearly clarify the possibility of errors or inconsistencies in the PIDE. On the one hand this is attributed to the collectors: Copying / deriving many numbers is to some extent error prone. Another reason lies in the experiments and in the different methods to obtain the experimental results. All users are therefore strongly suggested to cross-check their work with the original publications instead of using PIDE only. In case any errors or inconsistencies in PIDE are detected, please contact GSI to contribute to the PIDE project (see: contact). Below some aspects which might induce uncertainties and inaccuracies are listed.

i) Energy and LET

- When leaving the Bethe-Bloch region, uncertainties of stopping power computer codes like ATIMA etc. start to increase.
- Some authors also published energy and LET. As they typically used different codes, or measured the quantities, differences might occur due to different methods.
- In mixed radiation fields (such as within an extended Bragg peak) there is no unique choice of LET. Indeed most researchers evaluate the dose averaged LET while others give track averaged LET values (see overview table).
- Keep in mind that 'monoenergetic' beams are never exactly monoenergetic, but always subject to straggling, scattering and fragmentation. Hence, LET is always some averaged quantity, although the way how its value is obtained is not always clear.

ii) Linear quadratic (LQ) parameters

- For RBE typically photon radiation is used as a reference. However, it is known that X-rays from X-ray tubes are slightly more effective than those from clinical accelerators or γ-rays. The users should check for their work, if this issue must be taken into account or not. In PIDE3.x a corresponding column is adapted to reflect the used photon radiation quality.
- For calculating survival levels from colony formation assays, the plating efficiency needs to be known. Some authors estimate that from controls only, others use a three parameter fit to reveal α, β and the plating efficiency in one step.
- Some authors perform an error weighted fit, others an unweighted fit to dose response data. So again, methods may differ between the publications.
- In all fits performed within PIDE no weights have been applied to consider specific uncertainties of data points. Note that this method is applied to all available raw data sets in the same way. The resulting LQ parameters differ in some cases considerably from those given in the publications (if available), mostly because another fit approach has been chosen (e.g. purely linear instead of LQ) or because in the PIDE fits the plating efficiency was always redetermined within the fit, which is rarely the case in the publications.

iii) Errors and uncertainties

- The level of accuracy differs between the publications. Some authors presented many dose points and / or many platings or independent experiments within a survival curve, others less, depending on the particular aspect of research followed.
- Some authors provide a thorough error analysis of their data, others don't. However, there is no unique way of calculating uncertainty, hence the uncertainty estimates differ, again.
- The parameters α and β are correlated in usual fit methods. This leads to non-trivial corresponding uncertainty limits. This fact is known, but usually ignored.
- High-LET survival curves are often linear. Hence it's quite natural to investigate them with a lower number of samples, i.e. with fewer dose points than for photon dose response curves. On

the other hand, for this case often only a one parameter fit is performed, while the corresponding fit of the reference curve respects two parameters. So it's not straight forward to evaluate and compare the accuracy even for a pair of photon and ion dose response curve.

- Some practicable approximate error estimate is currently being developed for PIDE at GSI.

Again, having in mind these aspects, all users are strongly advised not to use PIDE as a 'black box', but also to care about possible issues, which might give rise to inconsistent conclusions. Although the developers of PIDE believe that statistics partially overcomes most of these issues, here a general warning shall be spoken out very clearly.

3. Dos and don'ts

All users of the PIDE are kindly asked to participate: If you

- find errors in the provided material
- are aware of any other suitable data which should be part of PIDE
- have any further ideas regarding the structure of PIDE
- if you have done research with PIDE which might be of interest for all PIDE users

you are invited to contact the developers at GSI (see: contact).

All users are further kindly requested not to distribute this material: All registered PIDE users will be informed about updates and any detected issues via e-mail in order to prevent misleading conclusions and to keep up correctness of the material provided as good as possible. So, instead of distributing the files, please distribute the project page:

http://www.gsi.de/bio-pide

Please mention the use of PIDE in your publications by also referencing to the version number (PIDE3.x), which allows reproducing your published work.

4. Updates

Updates with fixed typos or other bugs will be released as necessary, updates containing new data in reasonable time intervals. All registered users will be informed via e-mail, if an update is available.

5. Contact

The PIDE has been developed by the modeling section of the GSI biophysics department. For any requests or concerns contact:

Thomas Friedrich
t.friedrich@gsi.de