

# Hands-on pdb4dna

<http://pdb4dna.in2p3.fr>

Emmanuel Delage

delage@clermont.in2p3.fr

Yann Perrot

perrot@clermont.in2p3.fr

Laboratoire de Physique Corpusculaire de Clermont Ferrand

Pôle Physique pour la Santé et l'Environnement

CNRS/IN2P3 – Université Blaise Pascal



# Hands-on pdb4dna

## 1

### Scope of the application

- PDB files
- Integration into Geant4
- Algorithm to find the closest atom in DNA
- Strand breaks

## 2

### Application

- Directory content
- Build the example
- Preparation before run
- Run with GUI
- Run in batch mode

## 3

### ROOT analysis

# Hands-on pdb4dna

## 1

### Scope of the application

- PDB files
- Integration into Geant4
- Algorithm to find the closest atom in DNA
- Strand breaks

## 2

### Application

- Directory content
- Build the example
- Preparation before run
- Run with GUI
- Run in batch mode

## 3

### ROOT analysis

# PDB files - atomistic representation

- Protein Data Bank files :
  - A way to **store** and **exchange** polyatomic structures (proteins, **DNA**)
  - File describing the **3D structural information** of molecules
  - A **worldwide free access** to files obtained experimentally by crystallography techniques or computed with geometry optimization codes

<http://www.rcsb.org/>

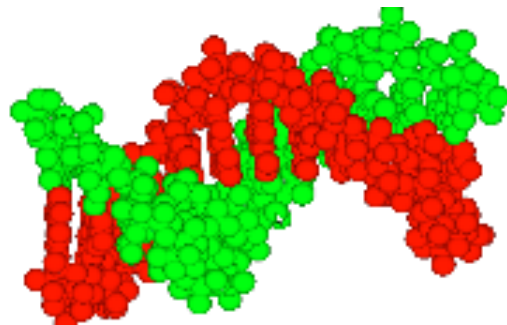
- File format :

- ASCII file ‘.pdb’
- We mainly extract information from **ATOM** keyword

```
HEADER                                TRANSFERASE 19-APR-13 4BJP
TITLE  CRYSTAL STRUCTURE OF E. COLI PENICILLIN BINDING PROTEIN 3
...
JRNL  DOI 10.1371/JOURNAL.PONE.0098042
...
REMARK 200 EXPERIMENT TYPE : X-RAY DIFFRACTION
...
ATOM   265  C  ASP  A   149  -38.902  78.078  31.778  C
ATOM   266  O  ASP  A   149  -39.587  78.804  32.500  O
ATOM   267  CB ASP  A   149  -36.407  78.326  31.747  C
...
TER   3106 ALA  A 567
...
END
```

# PDB files – DNA example 1/3

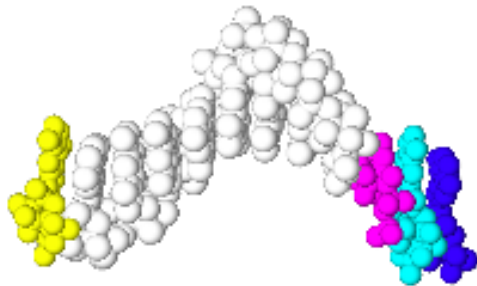
- 10 base pairs example – Molecule level



HEADER	DNA		10BP EXAMPLE	}	This file encodes DNA molecule
...					
MODEL	1			}	Start molecule
ATOM	1 O5'	A 1	9.256 -9.769 4.573	}	Atom coordinates – first strand (A)
ATOM	2 C5'	A 1	10.679 -9.579 4.526		
...					
ATOM	33 OP1	A 2	12.795 -8.381 9.736		
...					
ATOM	65 P	A 3	11.850 -2.418 12.300		
...					
ATOM	445 O5'	A 10	1.100 5.570 32.583		
TER	445	A 10			
ATOM	446 O5'	B 14	-9.356 10.980 33.794		
...					
ATOM	688 C4	B 20	2.805 3.343 8.223		
TER	688	B 20			
ENDMDL				}	End molecule

# PDB files – DNA example 2/3

- 10 base pairs example – nucleotide level



```
HEADER  DNA                               10BP EXAMPLE
...

MODEL  1

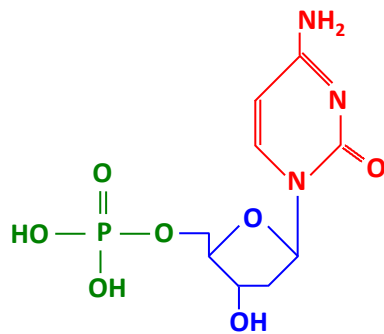
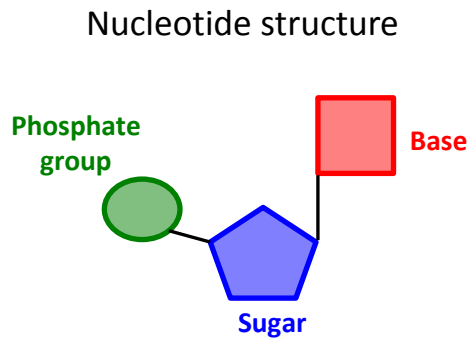
ATOM   1  O5'  A  1  9.256 -9.769  4.573  O
ATOM   2  C5'  A  1 10.679 -9.579  4.526  C
...
ATOM  33  OP1  A  2 12.795 -8.381  9.736  O
...
ATOM   65  P   A  3 11.850 -2.418 12.300  P
...
ATOM  445  O5'  A 10  1.100  5.570 32.583  O
TER    445      A 10

ATOM  446  O5'  B 14 -9.356 10.980 33.794  O
...
ATOM  688  C4   B 20  2.805  3.343  8.223  C
TER    688      B 20

ENDMDL
```

# PDB files – DNA example 3/3

- 10 base pairs example – atomic level

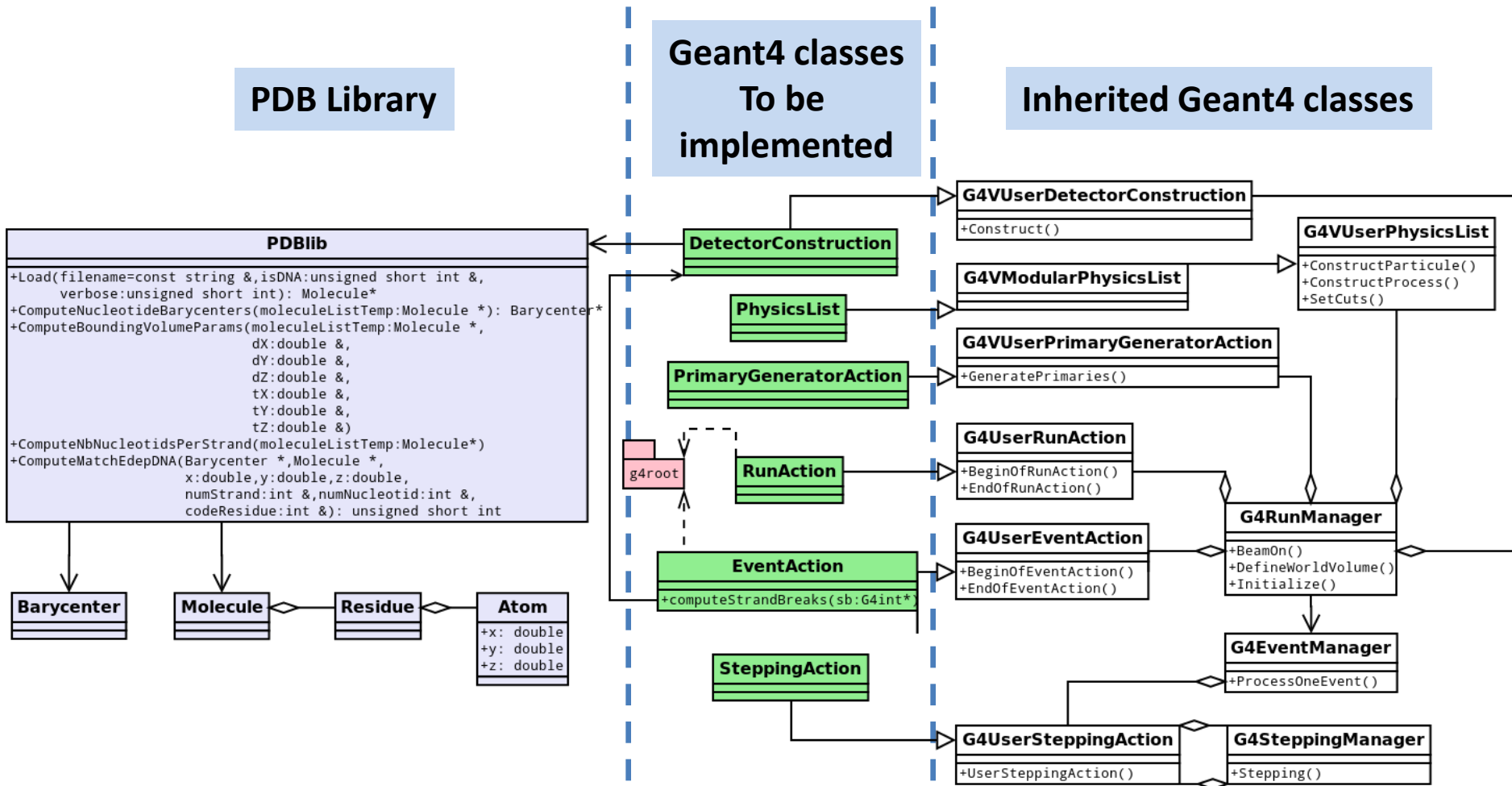


HEADER	DNA	10BP EXAMPLE	
...			
ATOM	65 P DCA 3	11.850 -2.418 12.300	P O O O } Phosphate group
ATOM	66 OP1 DCA 3	15.242 -2.080 12.670	
ATOM	67 OP2 DCA 3	11.107 -3.437 11.073	
ATOM	68 O5' DCA 3	12.981 -1.059 12.299	
ATOM	69 C5' DCA 3	11.563 0.189 11.896	C C O C O C } Sugar
ATOM	70 C4' DCA 3	12.527 1.314 11.875	
ATOM	71 O4' DCA 3	11.455 0.915 11.029	
ATOM	72 C3' DCA 3	11.954 1.427 11.278	
ATOM	73 O3' DCA 3	12.405 2.645 11.894	
ATOM	74 C2' DCA 3	10.453 1.552 11.042	
ATOM	75 C1' DCA 3	10.249 1.499 11.527	N C O N C C } Base
ATOM	76 N1 DCA 3	9.058 0.693 11.174	
ATOM	77 C2 DCA 3	7.998 1.341 10.557	
ATOM	78 O2 DCA 3	8.060 2.547 10.316	
ATOM	79 N3 DCA 3	6.897 0.610 10.230	
ATOM	80 C4 DCA 3	6.836 -0.699 10.496	
ATOM	81 N4 DCA 3	5.745 -1.381 10.158	
ATOM	82 C5 DCA 3	7.924 -1.372 11.112	
ATOM	83 C6 DCA 3	9.008 -0.643 11.454	

...

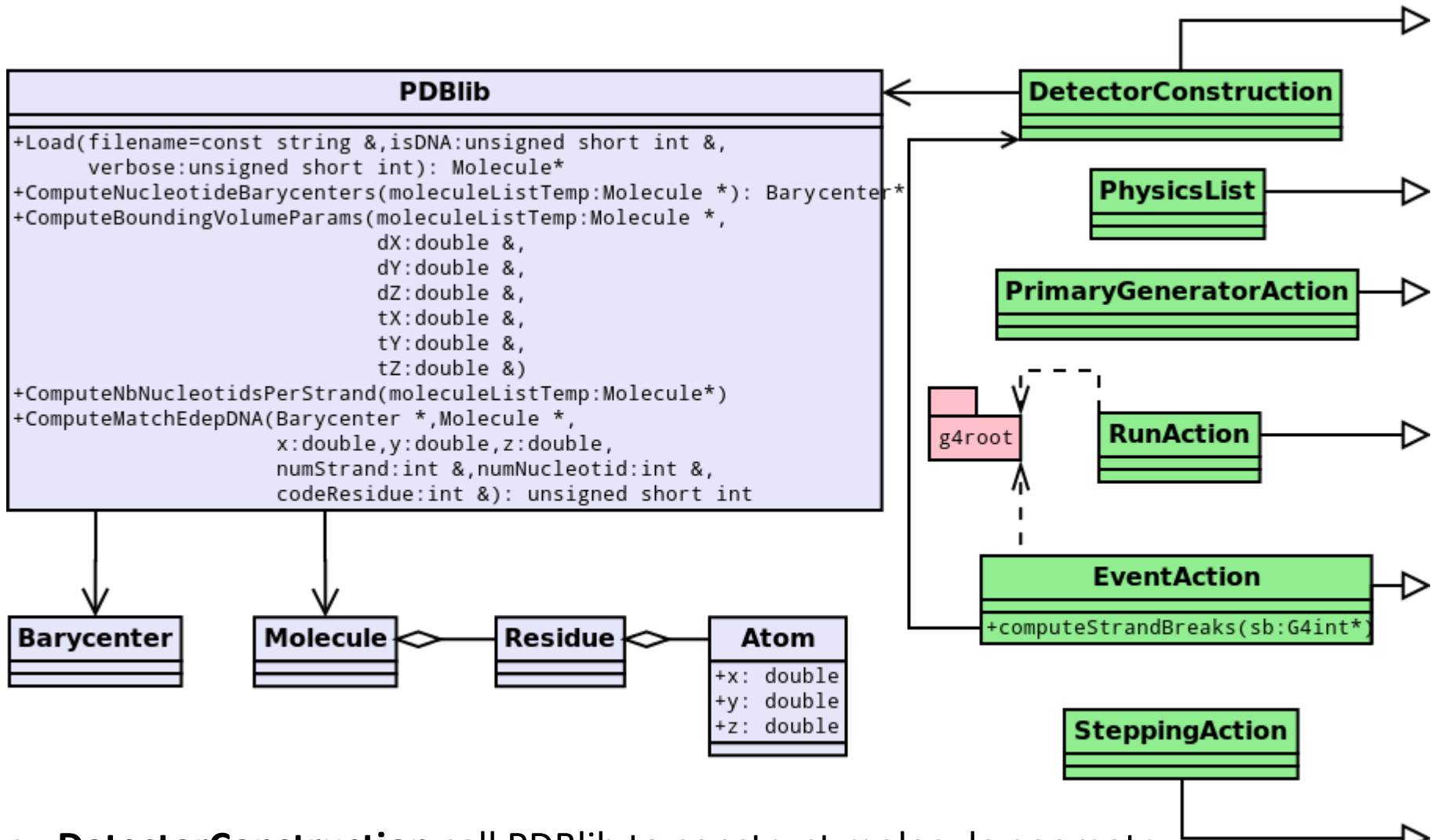
# Integration into Geant4 (1/2)

- Pdb4dna example will be provided in the next Geant4 release 10.1





# Integration into Geant4 (2/2)

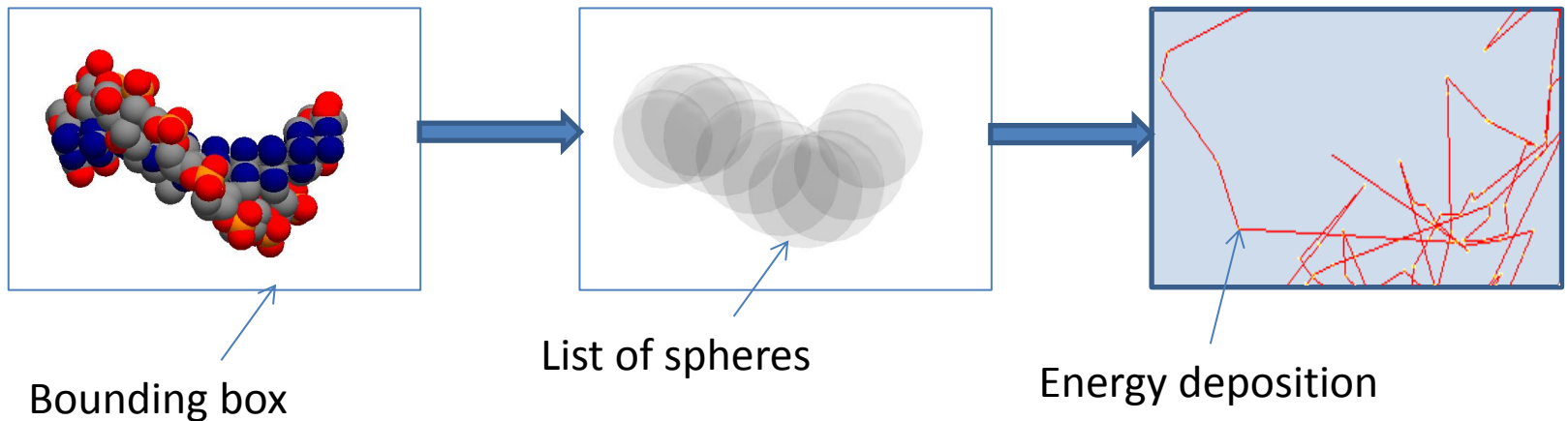


- **DetectorConstruction** call PDBlib to construct molecule geometry
- **EventAction** compute match between Edep and DNA via Detector construction

# Algorithm to find the closest atom 1/2

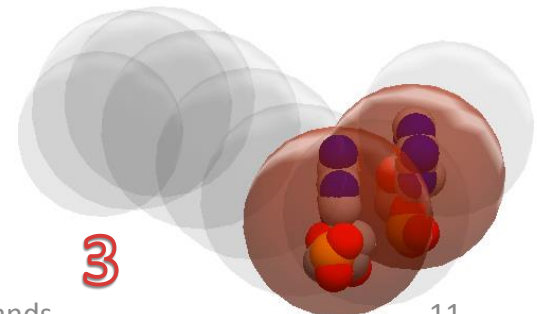
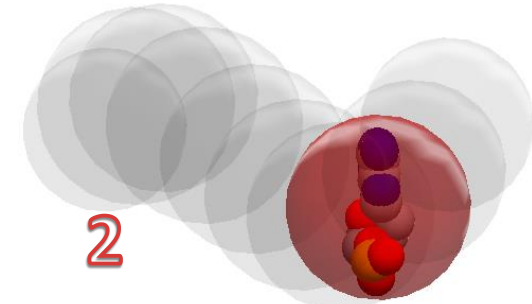
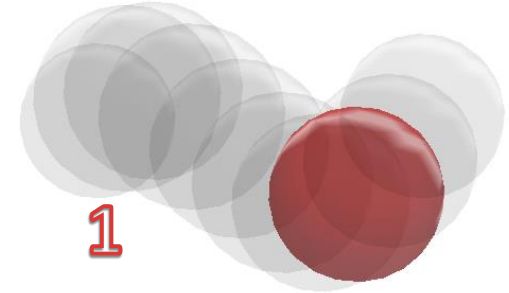
**Goal** = allocate **energy depositions** to an **element** [sugar, phosphate, base] of nucleotides and then deduce SSB and DSB

- A **bounding box** is calculated with atoms coordinates
- No other Geant4 solid is needed for simulation
- We consider that a sphere is a good approximation to englobe a nucleotide. A **list of spheres** representing nucleotides is calculated
- DNA specific



# Algorithm to find the closest atom 2/2

- **1**: Find the closest **nucleotide** from the energy deposition inside the two strands
- **2**: Find the closest **atom** from the energy deposition
  - atom by atom inside the selected nucleotide
  - considering Van der Waals radii
- **3**: Due to sphere overlapping, find a better match in the next two nucleotides in the list
- **4**: Return algorithm response:
  - No DNA hit
  - DNA hit, return **nucleotide ID, DNA strand, group type** (sugar, phosphate or base)

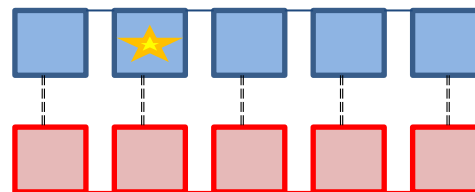


# Strand breaks

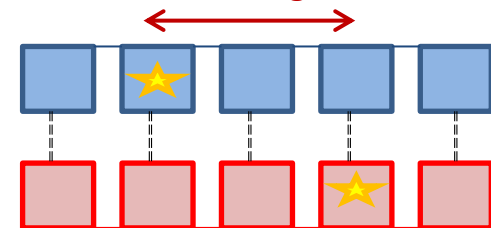
- Begin of event:
  - A **map** for each strand is created to **store** nucleotide ID and associated energy deposition
- For each step:
  - If the step is in the bounding volume, increment energy deposition per event
  - Ask to PDBlib to **check** the step **position**
    - If the step is in a sugar or a phosphate: get nucleotide ID, strand number, energy deposit, **update** the **map** (ID, Edep+=StepEdep)
- End of event:

- **Compute** and store **strand breaks**
- Store energy deposit in the bounding volume

Edep > threshold in sugar-phosphate to get a SSB



Distance between 2SSB > distance to get a DSB



# Hands-on pdb4dna

1

## Scope of the application

- PDB files
- Integration into Geant4
- Algorithm to find the closest atom in DNA
- Strand breaks

2

## Application

- Directory content
- Build the example
- Preparation before run
- Run with GUI
- Run in batch mode

3

## ROOT analysis

# Directory content

```
$ ll /mnt/g4tuto/geant4-10.01-install/share/Geant4-10.1.0/examples/extended/medical/dna/pdb4dna
```

```
09 ./
09 ../
08 1FZX.pdb
08 1ZBB.pdb
08 analysis.C
08 CMakeLists.txt
08 gui.mac
08 History
09 include/
08 init.mac
08 init_vis.mac
08 pdb4dna.cc
08 pdb4dna.in
08 README
08 README
08 runInGUI.mac
09 src/
08 vis.mac
```

→ PDB files

→ Root macro file

→ Macro to run  
in batch mode

→ Macro to run  
with GUI

```
./
../
ActionInitialization.hh
Analysis.hh
DetectorConstruction.hh
DetectorMessenger.hh
EventAction.hh
EventActionMessenger.hh
PDBatom.hh
PDBbarycenter.hh
PDBlib.hh
PDBmolecule.hh
PDBresidue.hh
PhysicsList.hh
PrimaryGeneratorAction.hh
RunAction.hh
RunInitObserver.hh
SteppingAction.hh
```

# Build the example (1/2)

```
$ module load geant4/10.01-mt
$ cd tutorial
$ cp -r /mnt/g4tuto/geant4-10.01-install/share/Geant4-10.1.0/examples/extended/medical/dna/pdb4dna .
$ mkdir pdb4dna-build
$ cd pdb4dna-build
$ cmake ../pdb4dna
```

```
user@user-VirtualBox:~/tutorial/pdb4dna-build$ cmake ../pdb4dna
-- The C compiler identification is GNU 4.8.2
-- The CXX compiler identification is GNU 4.8.2
-- Check for working C compiler: /usr/bin/cc
-- Check for working C compiler: /usr/bin/cc -- works
-- Detecting C compiler ABI info
-- Detecting C compiler ABI info - done
-- Check for working CXX compiler: /usr/bin/c++
-- Check for working CXX compiler: /usr/bin/c++ -- works
-- Detecting CXX compiler ABI info
-- Detecting CXX compiler ABI info - done
-- Configuring done
-- Generating done
-- Build files have been written to: /home/user/tutorial/pdb4dna-build
user@user-VirtualBox:~/tutorial/pdb4dna-build$
```

# Build the example (1/2)

```
$ make -j2
```

```
user@user-VirtualBox:~/tutorial/pdb4dna-build$ make -j2
Scanning dependencies of target pdb4dna
[ 6%] [ 12%] Building CXX object CMakeFiles/pdb4dna.dir/pdb4dna.cc.o
Building CXX object CMakeFiles/pdb4dna.dir/src/PDBlib.cc.o
[ 18%] Building CXX object CMakeFiles/pdb4dna.dir/src/RunAction.cc.o
[ 25%] Building CXX object CMakeFiles/pdb4dna.dir/src/SteppingAction.cc.o
[ 31%] Building CXX object CMakeFiles/pdb4dna.dir/src/PDBresidue.cc.o
[ 37%] Building CXX object CMakeFiles/pdb4dna.dir/src/PDBbarycenter.cc.o
[ 43%] Building CXX object CMakeFiles/pdb4dna.dir/src/DetectorMessenger.cc.o
[ 50%] Building CXX object CMakeFiles/pdb4dna.dir/src/PDBmolecule.cc.o
[ 56%] Building CXX object CMakeFiles/pdb4dna.dir/src/PrimaryGeneratorAction.cc.o
o
[ 62%] Building CXX object CMakeFiles/pdb4dna.dir/src/RunInitObserver.cc.o
[ 68%] Building CXX object CMakeFiles/pdb4dna.dir/src/PhysicsList.cc.o
[ 75%] Building CXX object CMakeFiles/pdb4dna.dir/src/DetectorConstruction.cc.o
[ 81%] Building CXX object CMakeFiles/pdb4dna.dir/src/EventActionMessenger.cc.o
[ 87%] Building CXX object CMakeFiles/pdb4dna.dir/src/EventAction.cc.o
[ 93%] Building CXX object CMakeFiles/pdb4dna.dir/src/ActionInitialization.cc.o
[100%] Building CXX object CMakeFiles/pdb4dna.dir/src/PDBatom.cc.o
Linking CXX executable pdb4dna
[100%] Built target pdb4dna
user@user-VirtualBox:~/tutorial/pdb4dna-build$
```



# Preparation before run (1/3)

Visit Protein Data Bank website → <http://www.rcsb.org/>

The screenshot shows the RCSB Protein Data Bank website homepage. At the top, there is a navigation bar with links for Home, Deposition, Education, News, Tools, Help & Contact Us, and More. The main header features the PDB logo and the text "An Information Portal to Biological Macromolecular Structures". Below the header is a search bar with a dropdown menu showing "Everything", "Author", "Macromolecule", "Sequence", and "Ligand". The search bar contains the text "e.g., PDB ID, molecule name, author". Below the search bar is a section titled "Biological Macromolecular Resource" with a "Full Description" link. This section contains three featured articles: "Molecule of the Month: Ebola Virus Proteins", "Protein Structure Initiative Featured System: Deciphering Microbial DUFs", and "RCSB PDB Mobile Supported on iOS & Android Platforms". The mobile section includes logos for the App Store and Google Play. At the bottom, there is an "Explore Archive" section with filters for Organism, Taxonomy, Exp. Method, X-ray Resolution, Release Date, and Polymer Type. The right sidebar contains a "Customize This Page" section with a "New Features" link and a "Latest release: August 2014" announcement, and an "RCSB PDB News" section with a "Comparison Tool for Exploring Sequence and Structure Alignments" link.

# Preparation before run (2/3)

You can find a file by its name, type, sequence, etc...

Home Deposition Education News Tools Help & Contact Us More

RCSB PDB PROTEIN DATA BANK

RCSB PDB-101

An Information Portal  
As of Tuesday Oct 21, 2014 at 5 PM PDT there are

Search  
Advanced  
Browse

Everything Author Macromolecule Sequence Ligand

1FZX

PDB ID DB Reference

- 1FZX
- PDB ID: 1FZX (1FZX - NDB)

close

Biological Macromolecules  
Full Description

Learn: Featured Molecules

Structural View of Biology

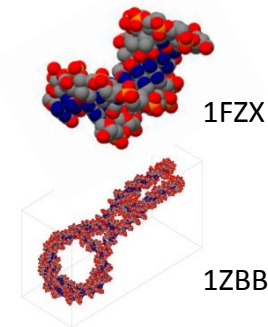
List View of Archive By: Title | Data

For the tutorial, we look at:

**1FZX** file : 12 bp DNA strands (run with GUI)

**1ZBB** file : dinucleosome (run in batch mode)

(Note that for the tutorial, those PDB files are provided in the pdb4dna directory)



« Do you know **1AOI.pdb**? »

# Preparation before run (3/3)

The screenshot shows the RCSB Protein Data Bank website for entry 1FZX. The page is titled "NMR SOLUTION STRUCTURE OF THE DNA DODECAMER GGCAAAAACGG". The entry ID is 1FZX. The primary citation is "Solution structure of an A-tract DNA bend." by MacDonald, D. P., Herbert, K. P., Zhang, X. P., Pologruto, T. P., Lu, P. P. (2001) J.Mol.Biol. 306: 1081-1098. The abstract describes the NMR structure of a DNA dodecamer d(GGCAAAAACGG)/d(CCGTTTTTGGCC) containing an A-tract. The structure is characterized by a propeller twist in the AT adenine bases and a negative inclination in the A-tract. The abstract also mentions the determination of the structure of the control sequence d(GGCAAGAAACGG)/d(CCGTTTCTGGCC). The page includes a search bar, navigation tabs, and a "Download Files" menu with options like "FASTA Sequence", "PDB File (Text)", "PDB File (gz)", "mmCIF File", "mmCIF File (gz)", "PDBML/XML File", "PDBML/XML File (gz)", "NMR Restraints (Text)", "NMR Restraints (gz)", and "V2 NMR Restraints (gz)". A "Structure Image" section shows a 3D model of the DNA dodecamer. The page also includes a "MyPDB Personal Annotations" section.

Description of the molecule  
References

Download the file

Online view  
(require Java)

# Run in GUI (1/4)

Due to GUI limitation, if you want to select a file into open file dialog you need to rename the extension from pdb to mac.

```
$ cp file.pdb file.mac
```

Run GUI

```
$ ./pdb4dna
```

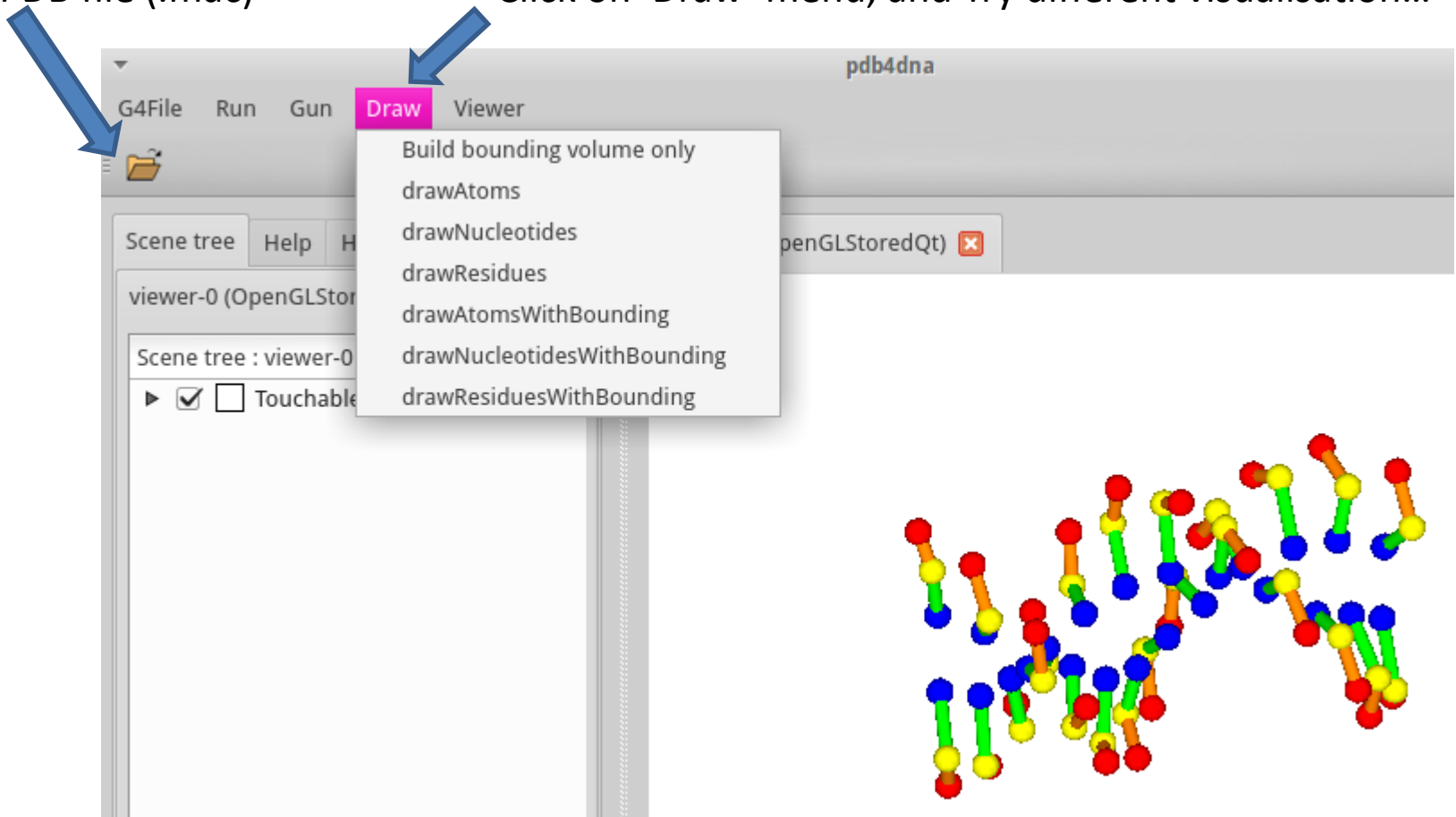
Be careful, the command will change in the next release to: `./pdb4dna -gui`



# Run in GUI (2/4)

Load PDB file (.mac)

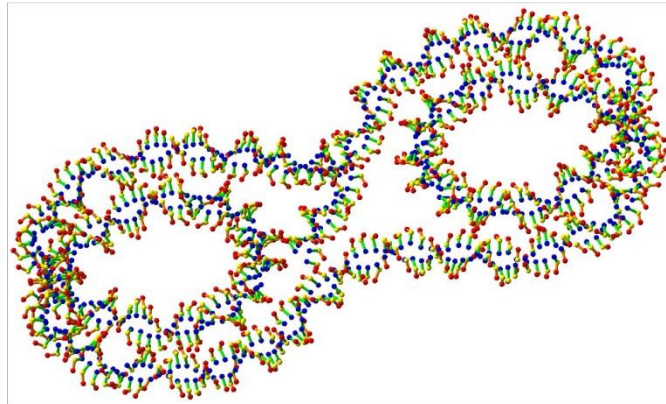
Click on 'Draw' menu, and Try different visualisation...



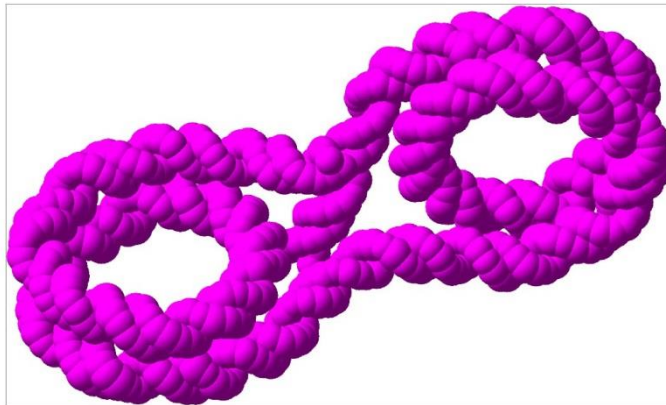
# Run in GUI (3/4)

Those screenshots correspond to 1ZBB.pdb (.mac)

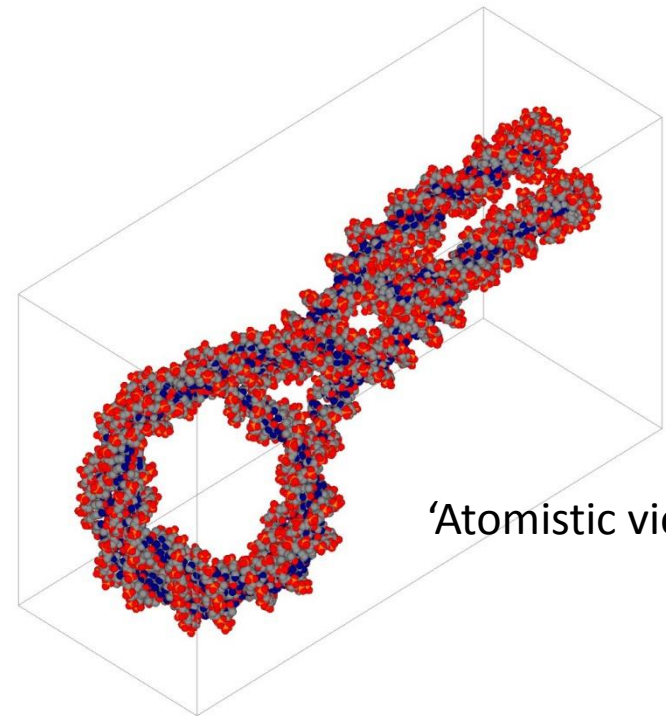
Visualization of a di-nucleosome (Asymmetric Unit of a tetra-nucleosome)



'Residue view'

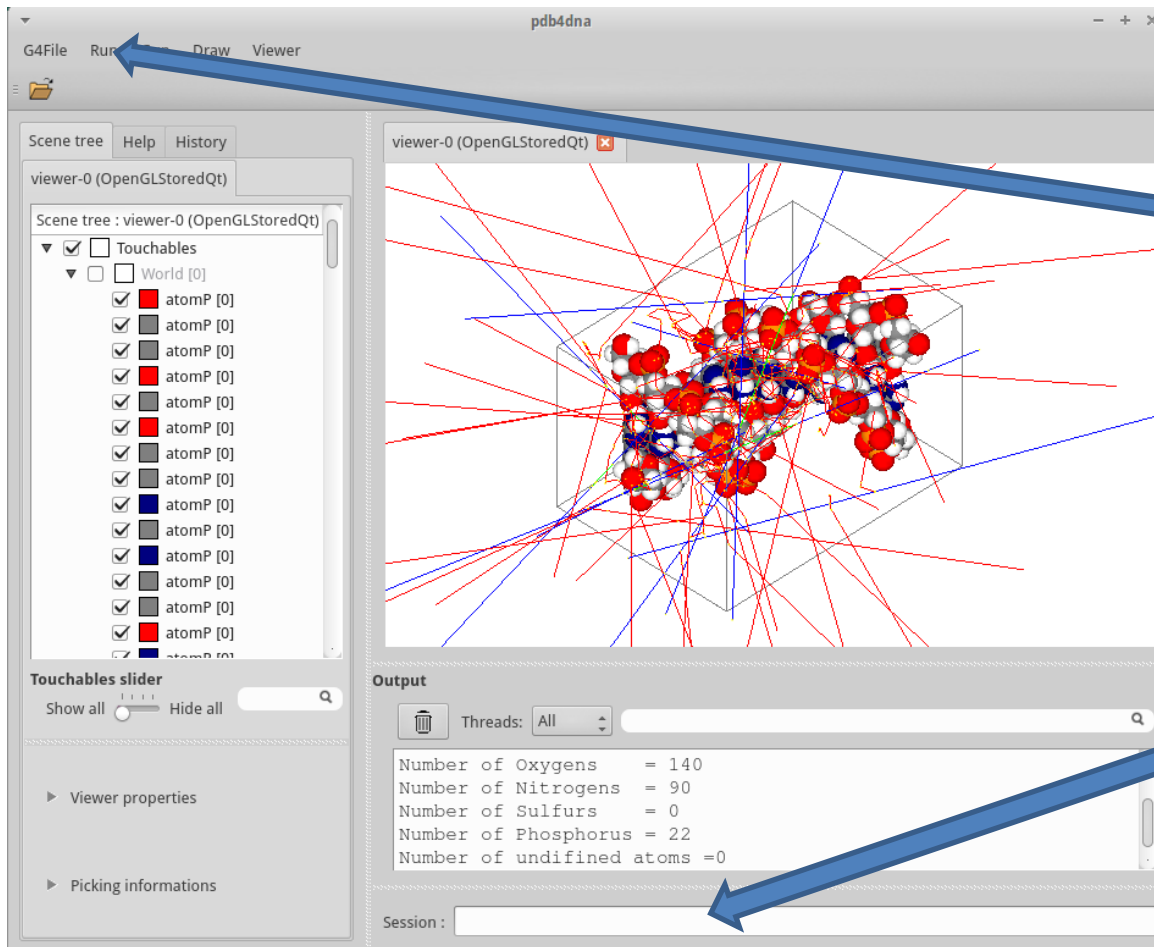


'Barycenter view'



'Atomistic view'

# Run in GUI (4/4)



Launch a simulation configured by 'runInGUI.mac' in the Run menu

Call run and gun commands in the session text box

```
/gun/particle e-
/gun/energy 10 keV
/run/beamOn 10
```

...



# Run in batch mode (1/2)

1. Prepare your macro file “tutorial.in” using your favorite editor (geany...)

You can start from the “pdb4dna.in” macro file

```
tutorial.in
/run/initialize

/PDB4DNA/det/loadPDB 1ZBB.pdb
/PDB4DNA/det/buildBoundingV

/PDB4DNA/event/setEnergyThres 8.22 eV
/PDB4DNA/event/setDistanceThres 10

/gun/particle proton
/gun/energy 1 keV

/run/initialize
/run/beamOn 10000
```

Select your PDB file

Construct the bounding volume

Set thresholds for strand breaks: energy, distance

Set characteristics of primaries

Initialize and run the simulation

# Run in batch mode (2/2)

## 2. Run the simulation

```
$ ./pdb4dna -m <your_macro_file> -t <number of threads>
```

Be careful, the command will change to:

```
./pdb4dna -mac yourmacrofile -mt numberofthread
```

```
$ ./pdb4dna -m tutorial.in -t 2
```

```
---> Begin of Event: 9500  
G4WT0 >  
---> Begin of Event: 9600  
G4WT0 >  
---> Begin of Event: 9700  
G4WT1 >  
---> Begin of Event: 9800  
G4WT0 >  
---> Begin of Event: 9900  
Graphics systems deleted.  
Visualization Manager deleting...  
G4WT1 > Total navigation history collections cleaned: 7  
G4WT0 > Total navigation history collections cleaned: 7  
Total navigation history collections cleaned: 3  
user@user-VirtualBox:~/tutorial/pdb4dna-build$
```

# Hands-on pdb4dna

## 1

### Scope of the application

- PDB files
- Integration into Geant4
- Algorithm to find the closest atom in DNA
- Strand breaks

## 2

### Application

- Directory content
- Build the example
- Preparation before run
- Run with GUI
- Run in batch mode

## 3

### ROOT analysis

# ROOT analysis

- A simulation will produce a root file “**pdb4dna\_output.root**”
- 3 histograms: **energy deposits** in the bounding volume per event, **SSB** and **DSB**

## \$ **root analysis.C**

```
*****  
*                                     *  
*      W E L C O M E  t o  R O O T    *  
*                                     *  
*  Version   5.32/04      13 July 2012 *  
*                                     *  
*  You are welcome to visit our web site *  
*      http://root.cern.ch             *  
*                                     *  
*****
```

ROOT 5.32/04 (tags/v5-32-04@45051, Jul 13 2012,  
15:55:10 on linuxx8664gcc)

CINT/ROOT C/C++ Interpreter version 5.18.00, July 2, 2010  
Type ? for help. Commands must be C++ statements.  
Enclose multiple statements between { }.

root [0]

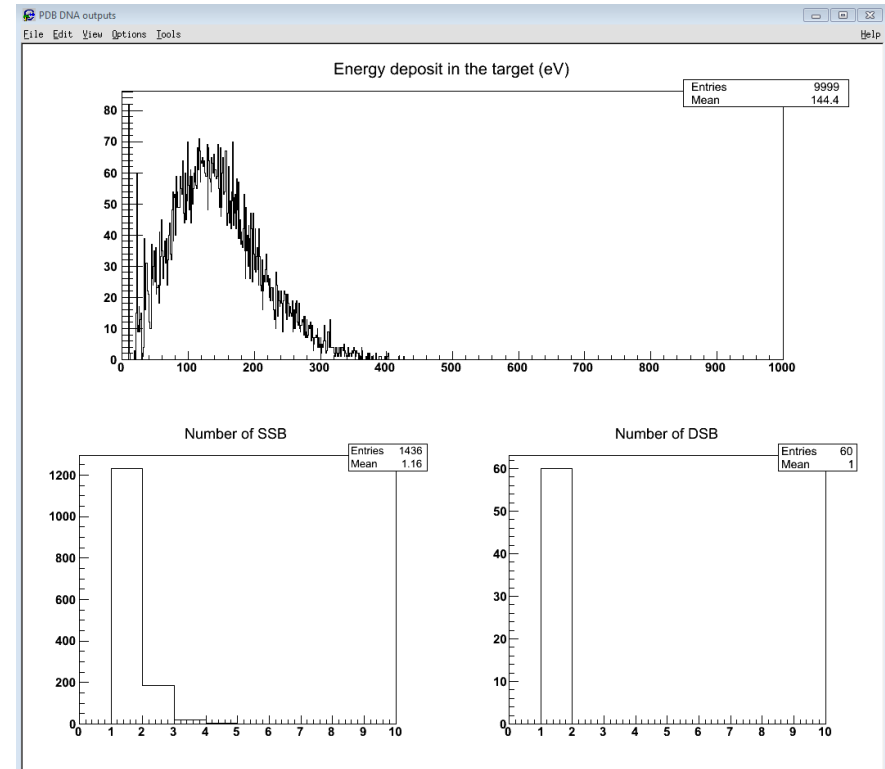
Processing analysis.C...

-> **Edep in the target : 1.44367 Mev**

-> **Number of SSB : 1666**

-> **Number of DSB : 60**

root [1]



# Go further

Now, you have an overview of pdb4dna.

Your feedback is important.

Contact us, we will provide you information concerning technical aspect, implementation, possible improvements...

[delage@clermont.in2p3.fr](mailto:delage@clermont.in2p3.fr)

[perrot@clermont.in2p3.fr](mailto:perrot@clermont.in2p3.fr)

Thanks a lot and thanks to the organizers!