GEANT4-DNA GEOMETRY AND CHEMISTRY

Nathanael Lampe and Sebastien Incerti

Belgrade, November 2016

Structure

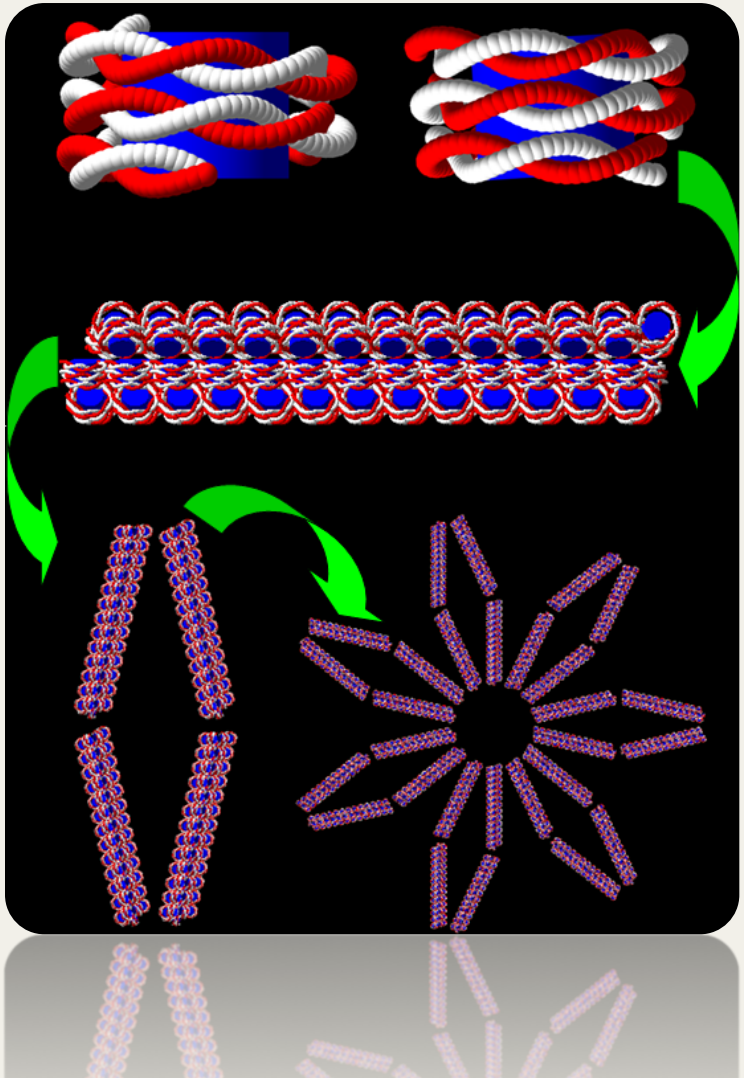
- Introduction to geometries for DNA-level simulations
  - *Hands on example: PDB4DNA*
- Introduction to the Geant4-DNA Chemistry module
  - *Hands on examples: chem1 and chem3*

Existing examples

- Whole Nuclear DNA
- PDB4DNA

- Coming Soon: MolecularDNA

« wholeNuclearDNA » extended example



Nucleosome

- *200 bp / nucleosome*
- *DNA diameter = 2.16 nm*
- *Histone = cylinder of 6.5 nm in diameter and 5.7 nm in height*

Chromatin fiber

- *90 nucleosomes / fiber*
- *7 nucleosomes / turn*
- *D = 31 nm*
- *L = 161 nm*

Chromatin fiber loop

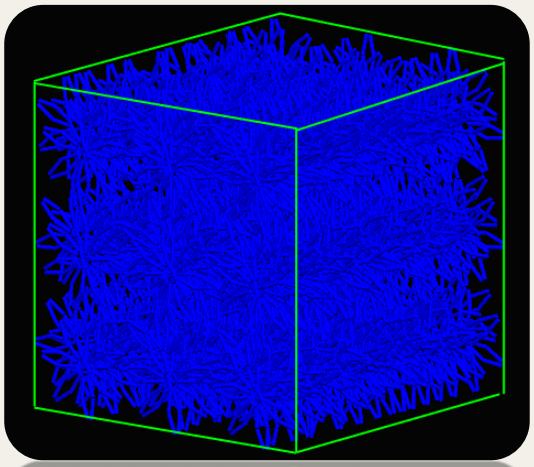
- *4 fibers / loop assembled in a diamond shape*
- *7 loops to form a "flower"**

* W. Friedland et al, Rad. Res 59 (2003), 401-410

« wholeNuclearDNA » extended example

« DetectorConstruction » class: implementation of an elliptical cell nucleus with similar dimensions of fibroblast grown on a microscopic plate at confluence

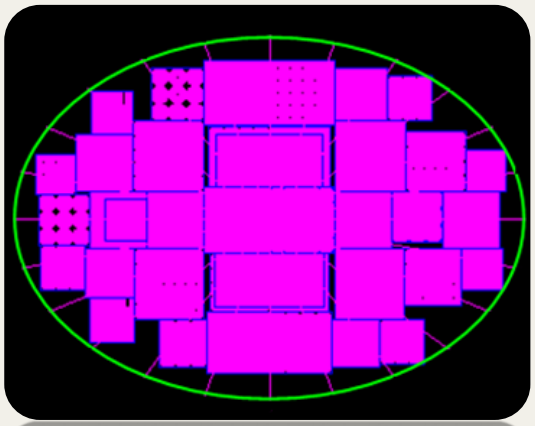
Chromosome domain example



Per nucleus
- 23 pairs of chromosomes
- 11875 flowers or 83125 loops
- 332 500 chromatin fibers
- 29 925 000 nucleosomes
- ~ 6 Gbp

Output: a ROOT file containing an n-tuple with the following values only for energy transfer points located in the backbone region:
-**Particle type** at the origin of the energy deposition
-**Process type (ionization, excitation)**
-Information on the **DNA strand** (flag 1 / 2)
-**Coordinates of the energy deposition** (x,y,z)
-**Energy deposition** amount

« Fibroblast » cell nucleus



- Nucleus-shape: ellipsoid
- Dimensions: 19.7 * 14.2 * 5 μm^3
- V = 732 μm^3
- 0.42 % of DNA / nucleus

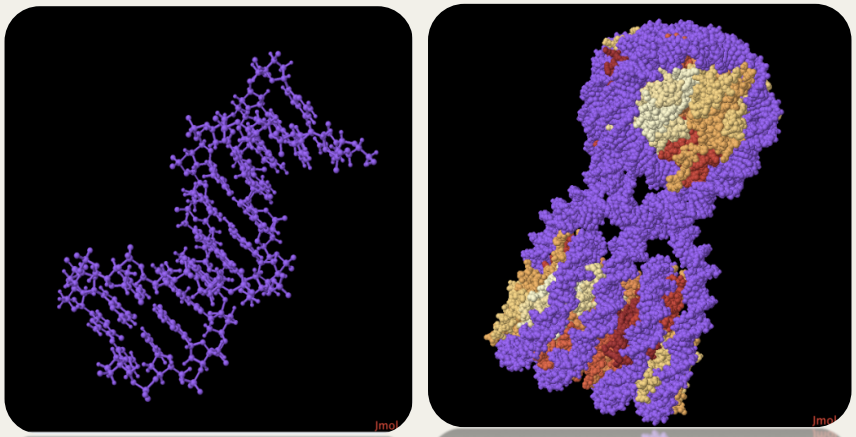
Nucl. Instrum. and Meth. B 298 (2013) 47 (link)

http://pdb4dna.in2p3.fr
http://geant4-dna.org

A new interface to describe geometries in Geant4-DNA

- PDB : Protein Data Bank
  http://www.rcsb.org/pdb/
  - *3D structure of molecules*
  - *Proteins*
  - *Nucleic acids*
- Description of DNA molecules
  - *1FZX.pdb*
  - Dodecamer
  - 12 DNA base pairs
  - (2,8 x 2,3 x 4,01 nm^3)
  - *1ZBB.pdb*
  - Tetranucleosome
  - 2 nucloosomes : 347 pairs of bases
  - (9,5 x 15,0 x 25,1 nm^3)

```
HEADER    STRUCTURAL PROTEIN/DNA                08-APR-05   1ZBB
TITLE     STRUCTURE OF THE 4_601_167 TETRANUCLEOSOME
...
ATOM      1  O5'  DA I   1      70.094  16.969 123.433  0.50238.00      0
ATOM      2  C5'  DA I   1      70.682  18.216 123.054  0.50238.00      C
ATOM      3  C4'  DA I   1      69.655  19.289 122.776  0.50238.00      C
...
TER   14223      DT J 347
...
HELIX    1   1 GLY A   44  SER A   57  1                                14
HELIX    2   2 ARG A   63  ASP A   77  1                                15
...
SHEET    1   A 2 ARG A  83  PHE A  84  0
SHEET    2   A 2 THR B  80  VAL B  81  1  O  VAL B  81   N  ARG A  83
```



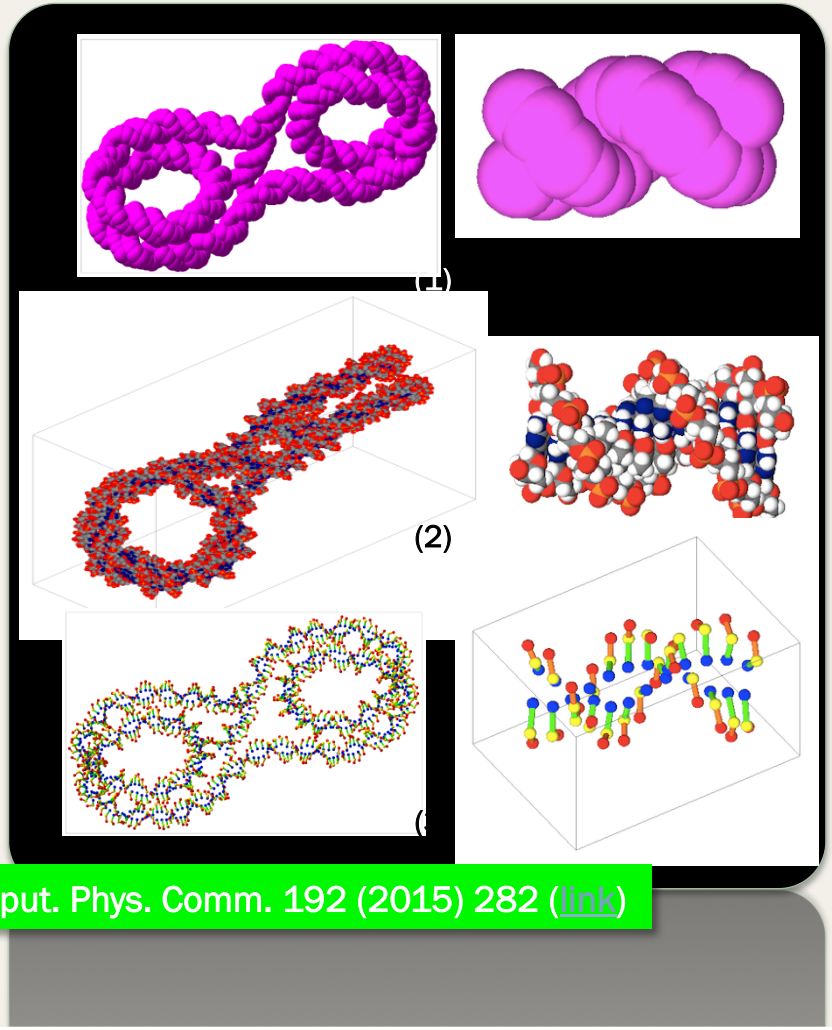
RCSB PDB PROTEIN DATA BANK

http://pdb4dna.in2p3.fr
http://geant4-dna.org

« PDB4DNA » suite

Click

- 1) A C++ library
  - *Reading of PDB files*
  - *Build bounding boxes from atom coordinates*
  - *Search for closest atom from a given point*
  - *Geometry and visualization : 3 granularities*
    - (1) Barycenter of nucleotides
    - (2) Atomistic
    - (3) Barycenter of nucleotide components
- 2) A Geant4-DNA example
  - *Water box surrounding the molecule*
  - *The output results consists in a ROOT file, containing for each event:*
    - energy deposit in bounding boxes
    - number of single strand breaks (SSB)
    - number of double strand breaks (DSB)
- Available on-line under Geant4 license



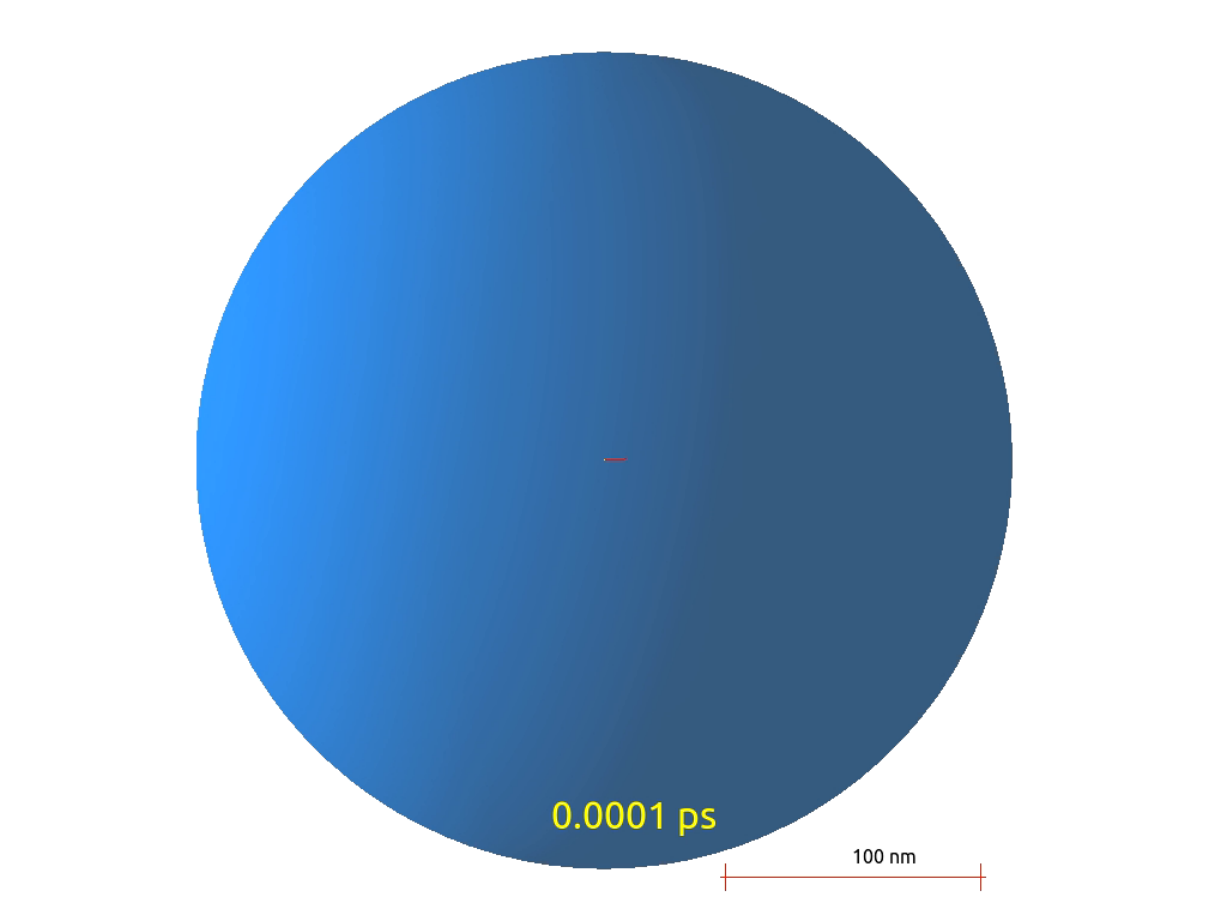
Comput. Phys. Comm. 192 (2015) 282 (link)

CHEM3

Geant4-DNA generated track of 1 keV electron, followed up to 100 ns during chemical stage. Using pre-release Geant4 10.1 code.

First, the physical stage is modeled, using Geant4-DNA physics: the 1 keV electron track is fully slowed down and thermalized in a 150 nm sphere of liquid water. Yellow points correspond to interaction events in liquid water.

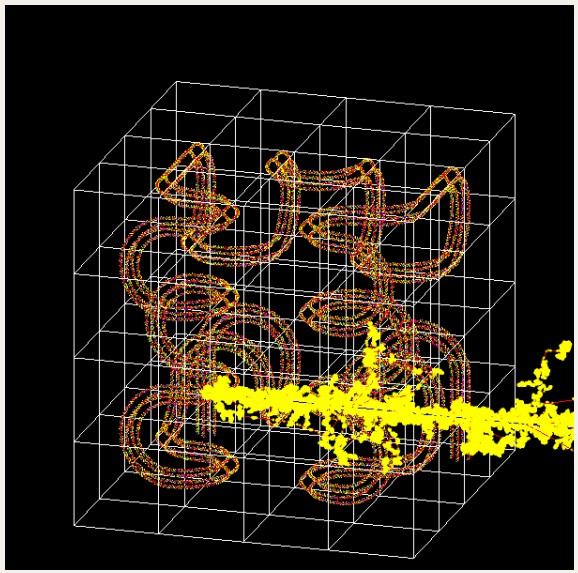
Next, the physico-chemical / chemical stage is simulated for the first 100 ns. Individual radical species diffusing in space and reacting with each other are shown as colour trails, the color coding chemical species type.



http://dx.doi.org/10.6084/m9.figshare.978887

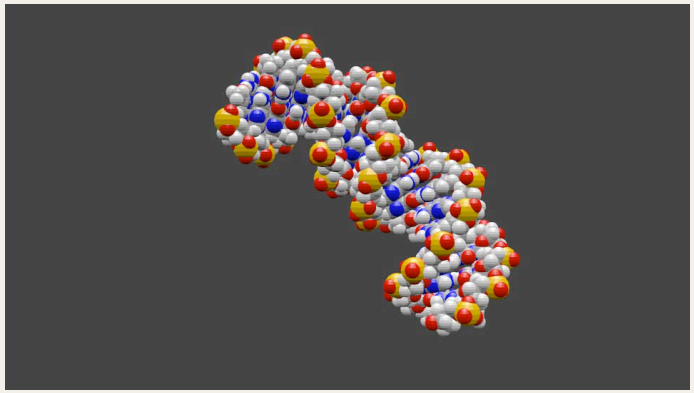
Coming Soon: MolecularDNA

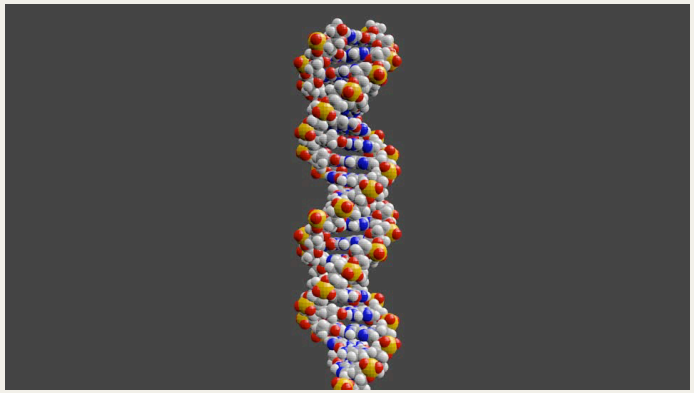
- A complete approach to model physics and chemistry with a simple specification of DNA geometries
- The idea is to quantify direct DNA damage (from physics) and indirect damage (from radicals)

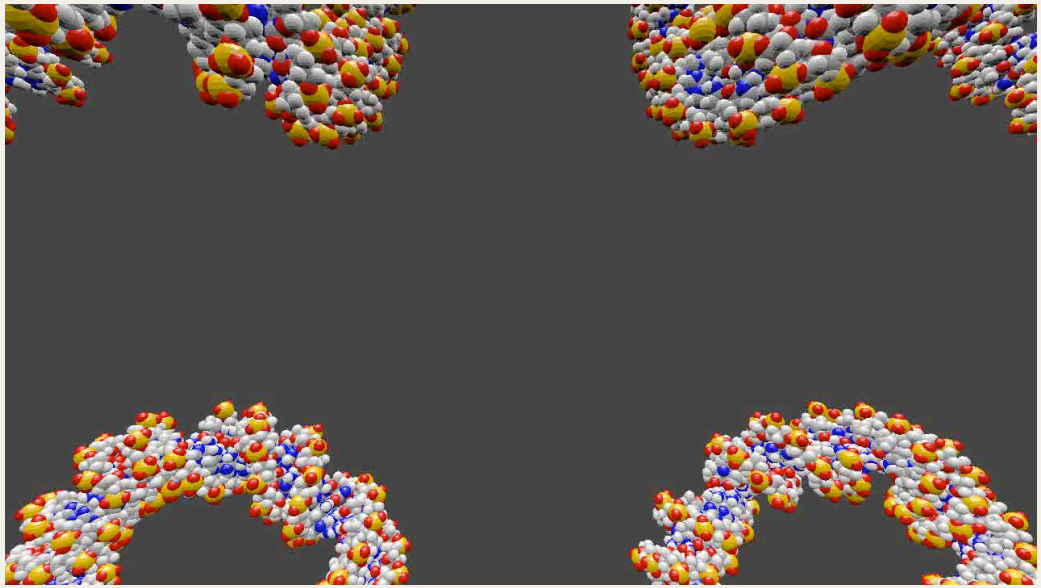


Geometry Definition

- Repeating geometries are used to describe the DNA macro-molecule.



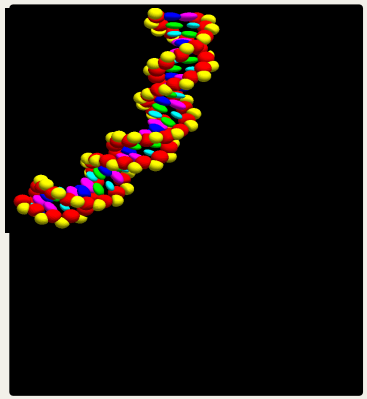




One continuous chain can be made from these two repeating units. Here the DNA is represented atomistically (and turns are exaggerated for clarity)

Geometry Definition

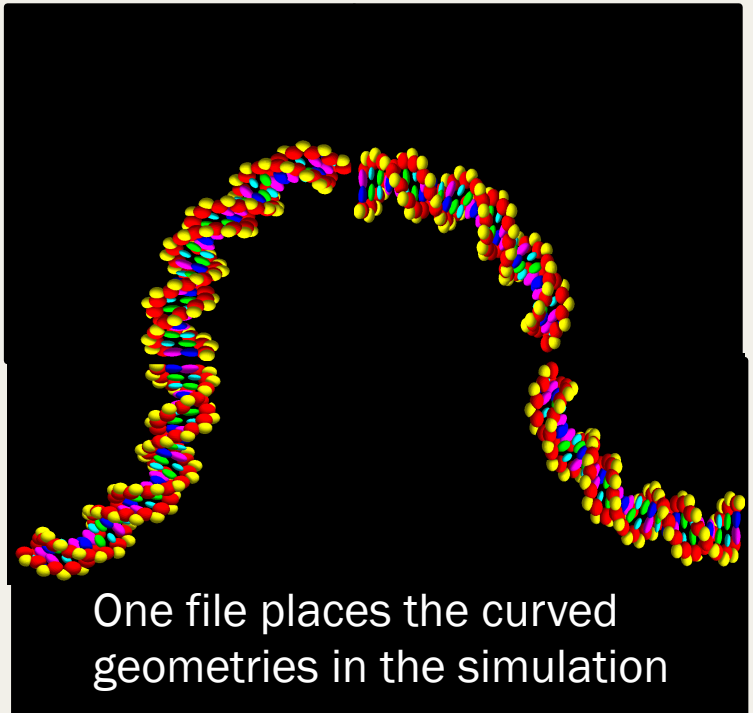
- Two types of files permit the definition of arbitrary geometries.
  - *One file to specify the position of molecules within each placement volume*
  - *One file to specify the position and rotation of each placement volume*
  - *DNA is built based on a molecular representation of the chain*
- One base pair = 2 sugars, 2 phosphates, 2 bases



One file specifies this curved geometry

180° rotation

No rotation



Implementation

```
34  MolecularDetectorConstruction::MolecularDetectorConstruction():
35      G4VUserDetectorConstruction()
36  {
37      fpDNAGeometry = new MolecularDNAGeometry();
38  }
39
```

All DNA geometry classes are contained in and initialised in MolecularDNAGeometry.
Importantly, this activates a number of G4Messengers

```
17  /dnageom/placementSize 50 nm
18  /dnageom/definitionFile geometries/bacteria-XFXFXFX-4.txt
19  /dnageom/placementVolume turn geometries/fourstrands_50nm_turn.txt
20  /dnageom/placementVolume turntwist geometries/fourstrands_50nm_turn.txt true
21  /dnageom/placementVolume straight geometries/fourstrands_50nm_straight.txt
```

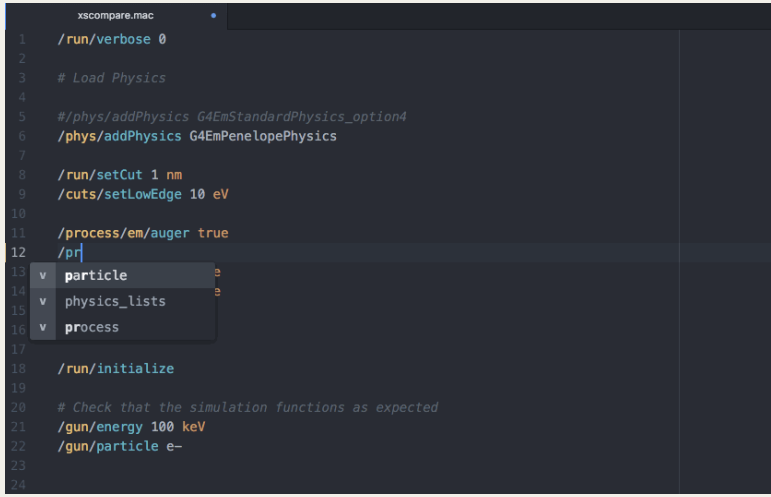
These messengers set the data files for when the DNA is constructed

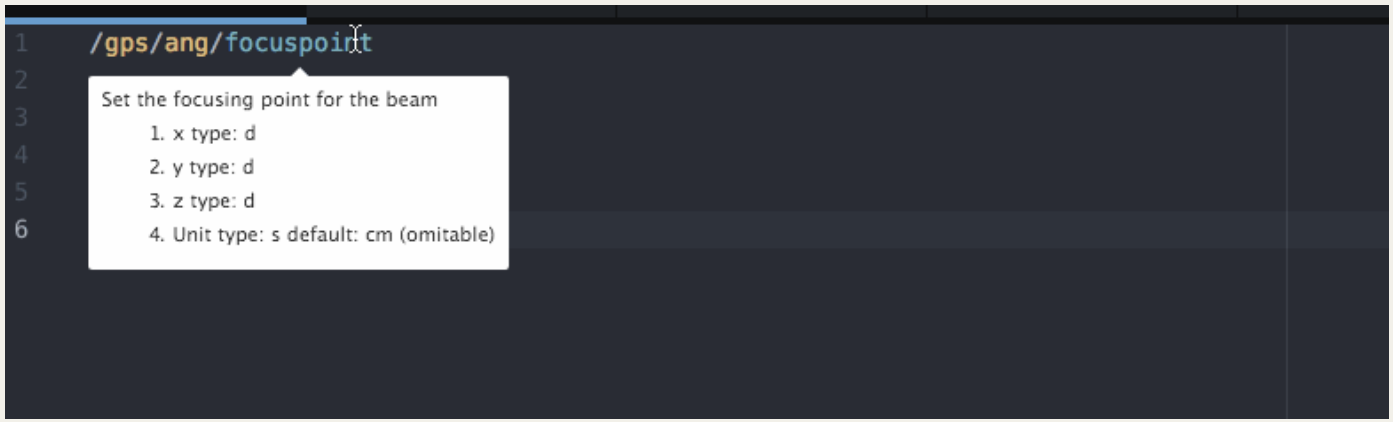
```
62  G4VPhysicalVolume* MolecularDetectorConstruction::ConstructDetector()
63  {
64      // Define world logical and physical volume
65
66      this->GetDNAGeometry()->BuildDNA(worldLogical);
67      return world;
68  }
```

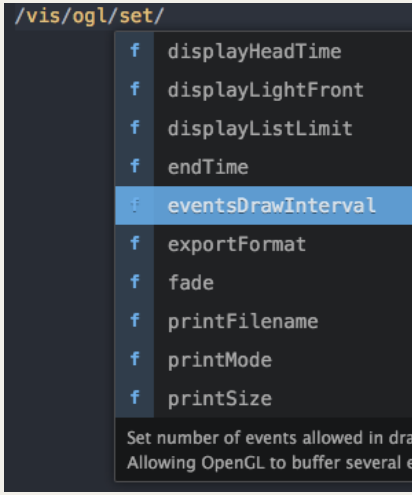
The DNA is constructed in a logical volume which contains it entirely.

Specialised Run/Event/Stepping Actions handle analysis

- *Users of the Atom text editor can access syntax highlighting and auto-completion for G4 macros.*
- *Want to find out more: http://goodfoodgoodcode.com contains some more explanations of what I do.*
- *Interested in modelling plasmids/human DNA/other DNA using this approach? Come speak to me.*







http://atom.io/packages/language-geant4-m

Follow me on Twitter: @natlampe

Geant4 release 10.2 Beta

PDB4DNA

Sebastien Incerti (CNRS) and Nathanael Lampe (LPC Clermont/CENBG)
based on material provided by
E. Delage and Y. Perrot (LPC Clermont)

Protein Data Bank

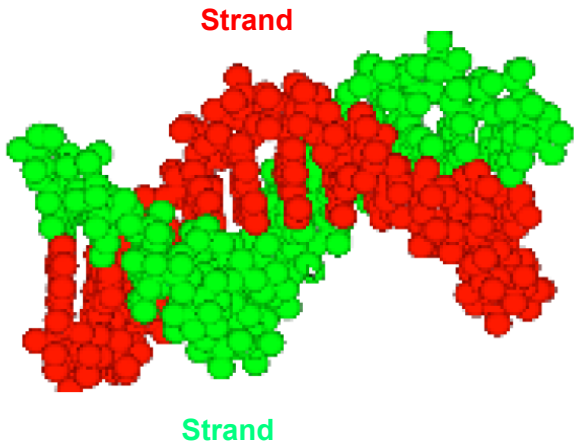
http://www.rcsb.org/

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

- 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
```

10 base pairs: two whole DNA strands




```
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
```

This file encodes DNA molecule

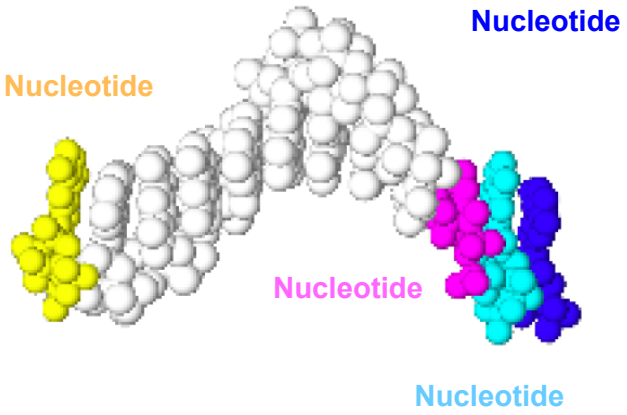
Start molecule

Atom coordinates – first strand (A)

Atom coordinates – second strand (B)

End molecule

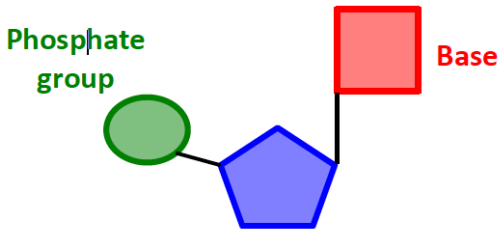
10 base pairs: individual nucleotides

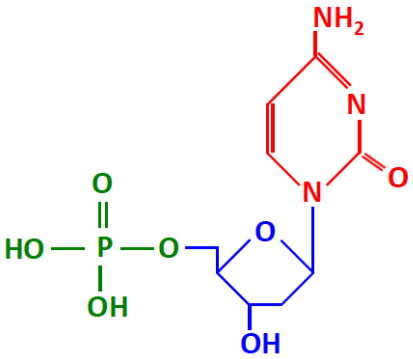



```
HEADER    DNA                          10BP EXAMPLE
...

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
```

10 base pairs: atomistic content of nucleotides

Nucleotide structure





Cytosine nucleotide

HEADER DNA 10BP EXAMPLE

...

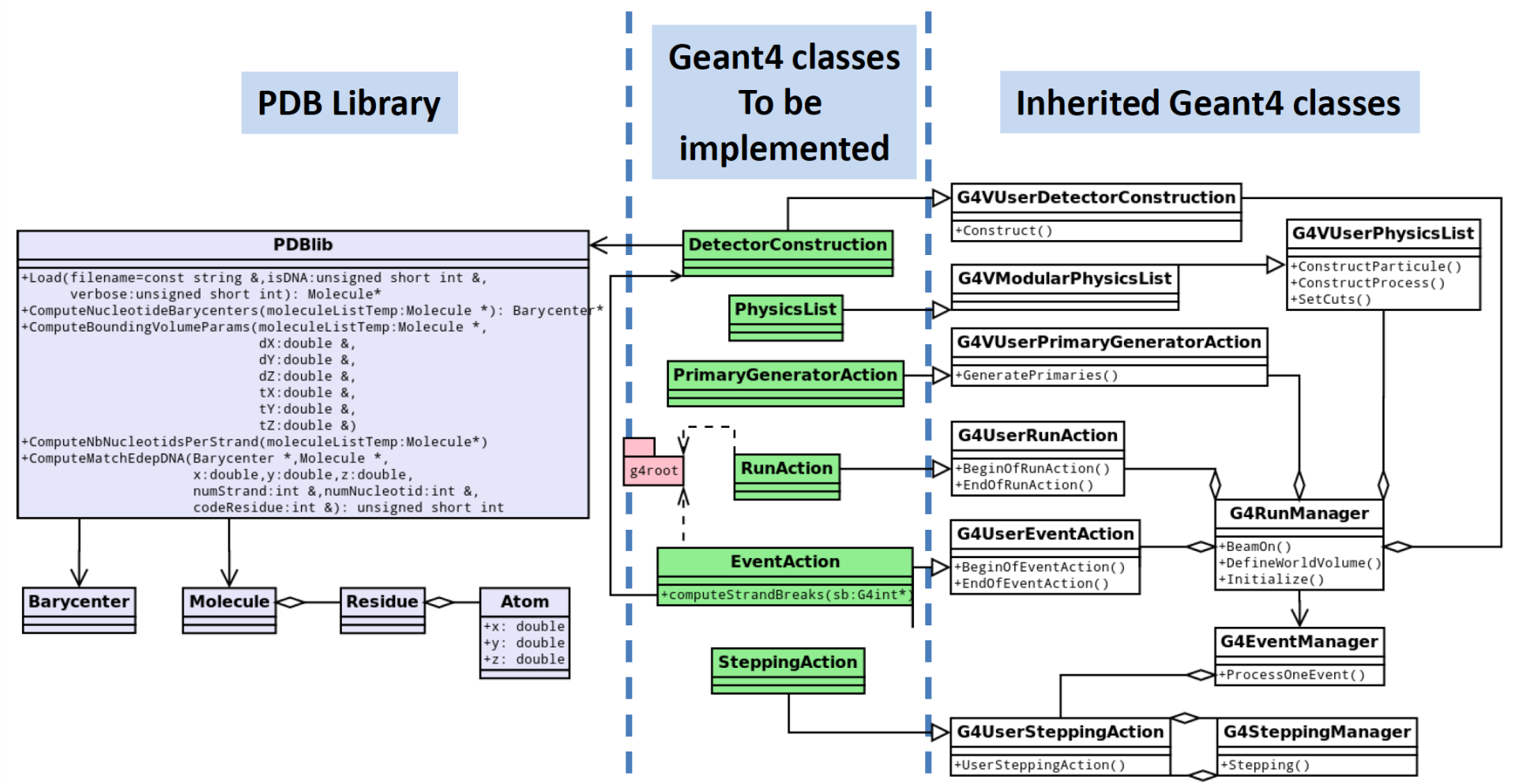
ATOM	65	P	DC	A	3	11.850	-2.418	12.300	P
ATOM	66	OP1	DC	A	3	15.242	-2.080	12.670	O
ATOM	67	OP2	DC	A	3	11.107	-3.437	11.073	O
ATOM	68	O5'	DC	A	3	12.981	-1.059	12.299	O
ATOM	69	C5'	DC	A	3	11.563	0.189	11.896	C
ATOM	70	C4'	DC	A	3	12.527	1.314	11.875	C
ATOM	71	O4'	DC	A	3	11.455	0.915	11.029	O
ATOM	72	C3'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

Phosphate group

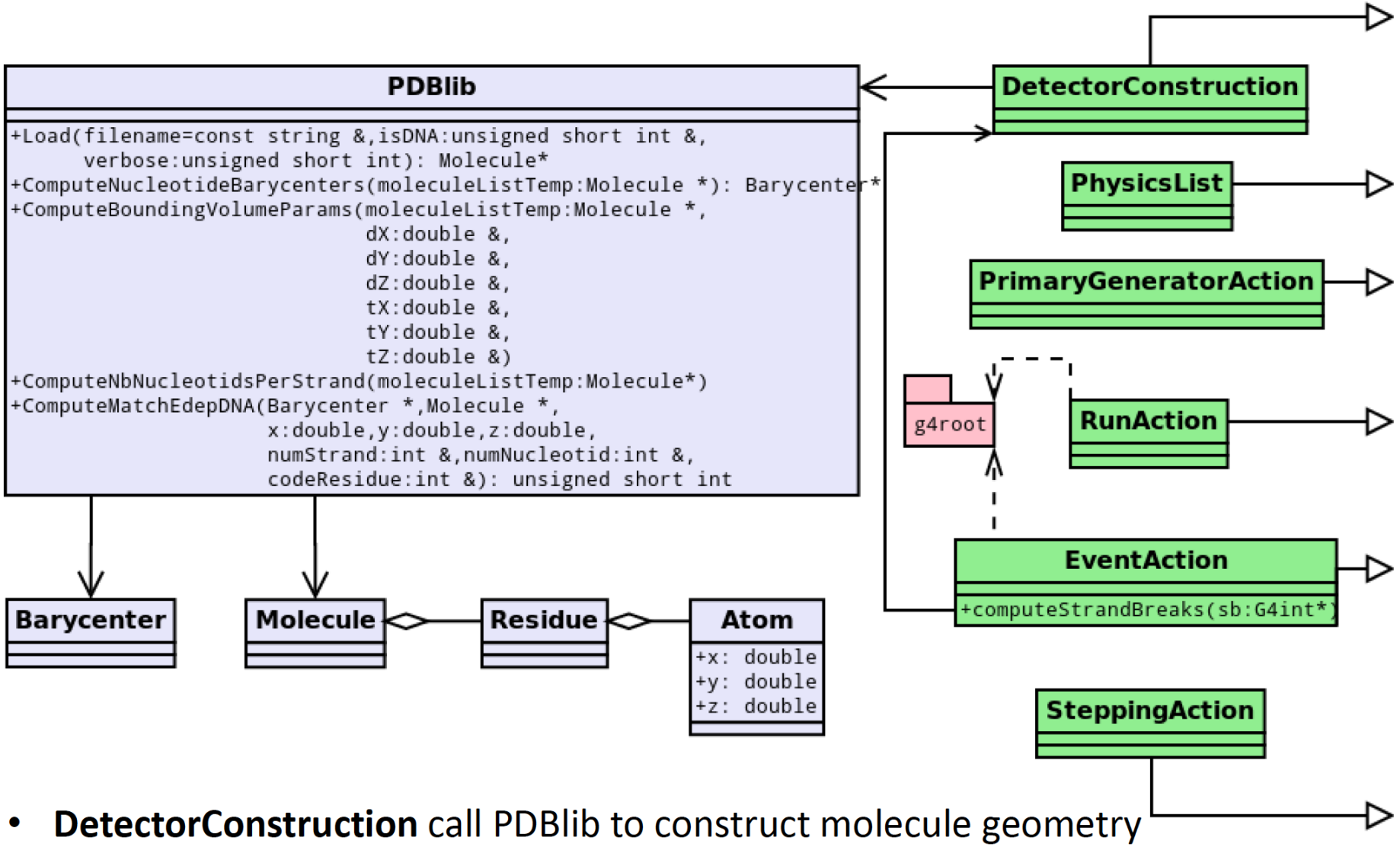
Sugar

Base

Integration into Geant4 (1/2)



Integration into Geant4 (2/2)

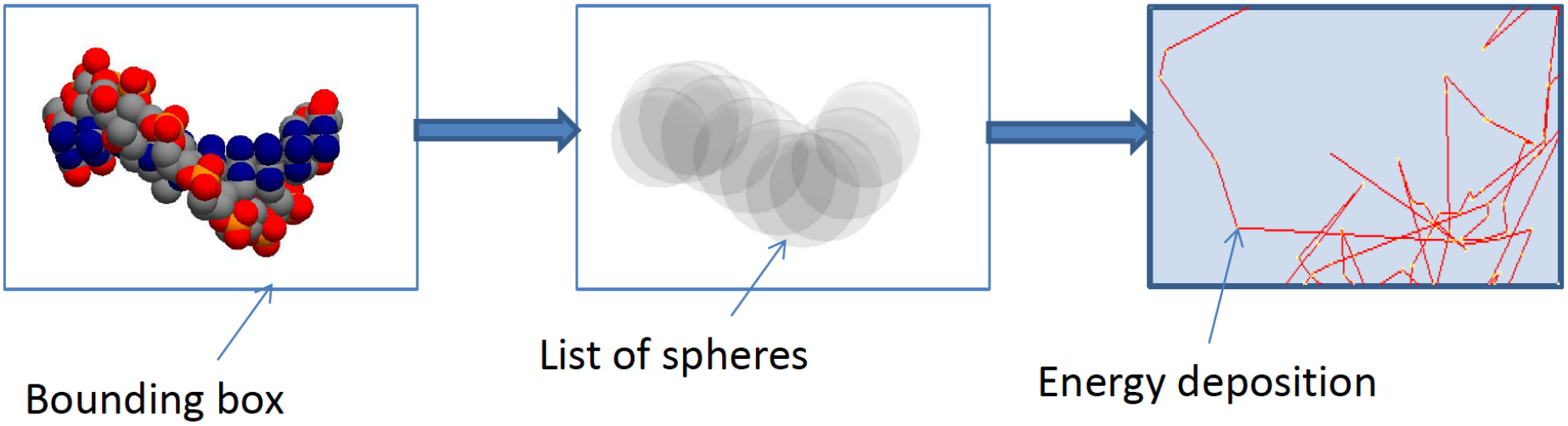


- **DetectorConstruction** call PDBlib to construct molecule geometry
- **EventAction** compute match between Edep and DNA via Detector construction

Algorithm to find the closest atom (1/2)

GOAL : allocate energy depositions to a geometrical 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

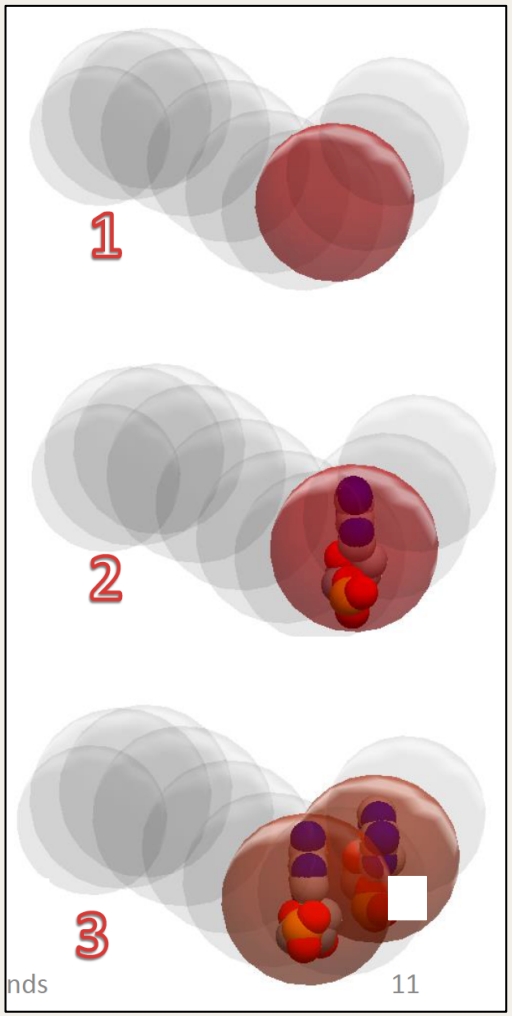


Algorithm to find the closest atom (2/2)

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:

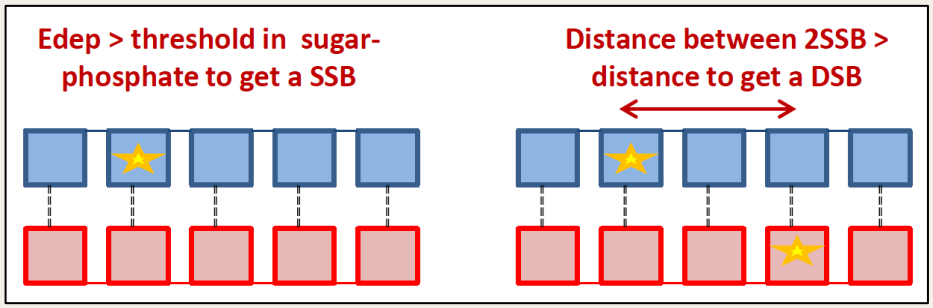
– *No DNA hit*

– *or 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*



« PDB4DNA » HANDS-ON

Directory content

```
ll $G4EXAMPLES/extended/medical/dna/pdb4dna/
```

```
2266137 Jun 26 12:11 1ZBB.pdb
   1380 Jun 26 12:11 analysis.C
   2222 Jun 26 12:11 CMakeLists.txt
    420 Jun 26 12:11 GNUmakefile
   1819 Jun 26 12:11 gui.mac
   2197 Jun 26 12:11 History
   4096 Jun 29 09:16 include/
    201 Jun 26 12:11 init.mac
    197 Jun 26 12:11 init_vis.mac
   8202 Jun 26 12:11 pdb4dna.cc
    341 Jun 26 12:11 pdb4dna.in
  11786 Jun 26 12:11 pdb4dna.out
   4744 Jun 26 12:11 README
    274 Jun 26 12:11 runInGUI.mac
   4096 Jun 29 09:16 src/
   2722 Jun 26 12:11 vis.mac
```

PDB file

ROOT macro file

Macro to run in batch mode

```
Jun 26 12:11 ActionInitialization.hh
Jun 26 12:11 Analysis.hh
Jun 26 12:11 CommandLineParser.hh
Jun 26 12:11 DetectorConstruction.hh
Jun 26 12:11 DetectorMessenger.hh
Jun 26 12:11 EventAction.hh
Jun 26 12:11 EventActionMessenger.hh
Jun 26 12:11 PDBatom.hh
Jun 26 12:11 PDBbarycenter.hh
Jun 26 12:11 PDBlib.hh
Jun 26 12:11 PDBmolecule.hh
Jun 26 12:11 PDBresidue.hh
Jun 26 12:11 PhysicsList.hh
Jun 26 12:11 PrimaryGeneratorAction.hh
Jun 26 12:11 RunAction.hh
Jun 26 12:11 RunInitObserver.hh
Jun 26 12:11 SteppingAction.hh
```

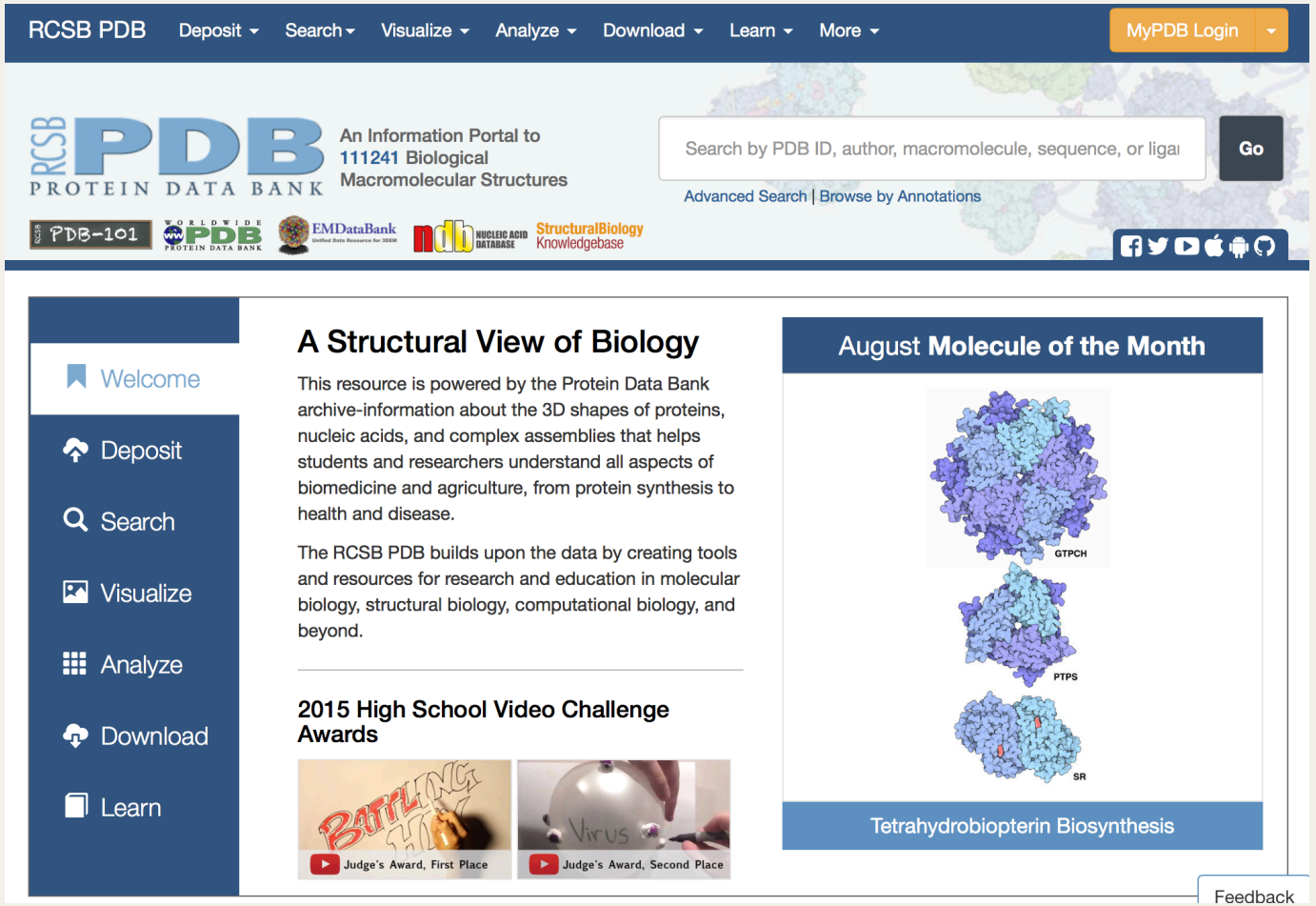
Compile and link the application

```
cd
cp -R $G4EXAMPLES/extended/medical/dna/pdb4dna/ .
mkdir build-pdb4dna
cd build-pdb4dna
cmake ../pdb4dna
make –j2
```

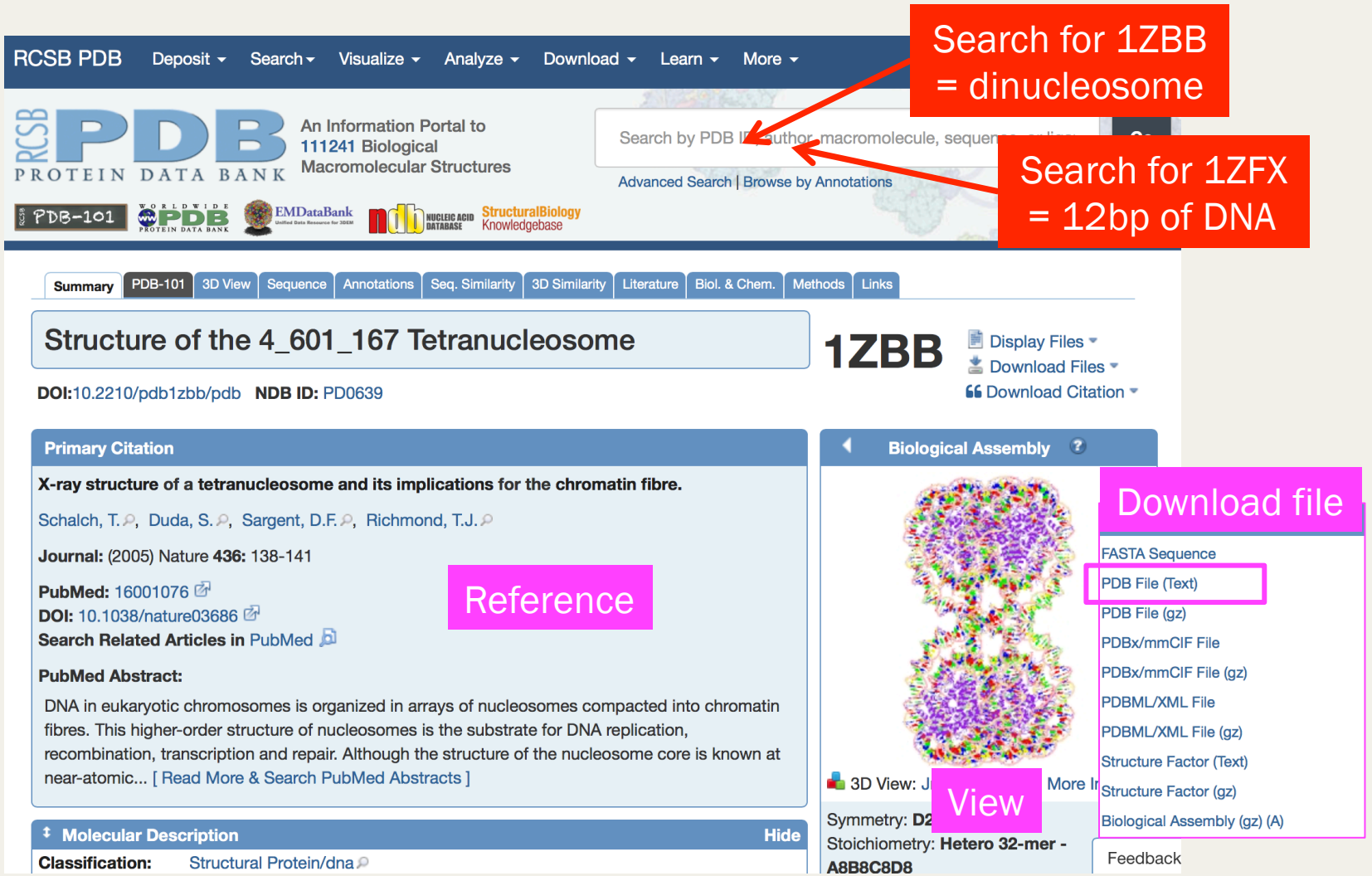
```
localhost.localdomain:/build-pdb4dna < 482 >make -j2
Scanning dependencies of target pdb4dna
[  5%] [ 11%] Building CXX object CMakeFiles/pdb4dna.dir/pdb4dna.cc.o
Building CXX object CMakeFiles/pdb4dna.dir/src/RunInitObserver.cc.o
[ 17%] Building CXX object CMakeFiles/pdb4dna.dir/src/ActionInitialization.cc.o
[ 23%] Building CXX object CMakeFiles/pdb4dna.dir/src/PDBmolecule.cc.o
[ 29%] Building CXX object CMakeFiles/pdb4dna.dir/src/PDBatom.cc.o
[ 35%] Building CXX object CMakeFiles/pdb4dna.dir/src/CommandLineParser.cc.o
[ 41%] Building CXX object CMakeFiles/pdb4dna.dir/src/PrimaryGeneratorAction.cc.o
[ 47%] Building CXX object CMakeFiles/pdb4dna.dir/src/DetectorMessenger.cc.o
[ 52%] Building CXX object CMakeFiles/pdb4dna.dir/src/SteppingAction.cc.o
[ 58%] Building CXX object CMakeFiles/pdb4dna.dir/src/PhysicsList.cc.o
[ 64%] Building CXX object CMakeFiles/pdb4dna.dir/src/PDBresidue.cc.o
[ 70%] Building CXX object CMakeFiles/pdb4dna.dir/src/RunAction.cc.o
[ 76%] Building CXX object CMakeFiles/pdb4dna.dir/src/PDBbarycenter.cc.o
[ 82%] Building CXX object CMakeFiles/pdb4dna.dir/src/DetectorConstruction.cc.o
[ 88%] Building CXX object CMakeFiles/pdb4dna.dir/src/EventActionMessenger.cc.o
[ 94%] Building CXX object CMakeFiles/pdb4dna.dir/src/EventAction.cc.o
[100%] Building CXX object CMakeFiles/pdb4dna.dir/src/PDBlib.cc.o
Linking CXX executable pdb4dna
[100%] Built target pdb4dna
localhost.localdomain:/build-pdb4dna < 483 >
```

Preparation before run (1/2)

Visit PDB web site **http://www.rcsb.org**



Preparation before run (2/2)



Running with GUI

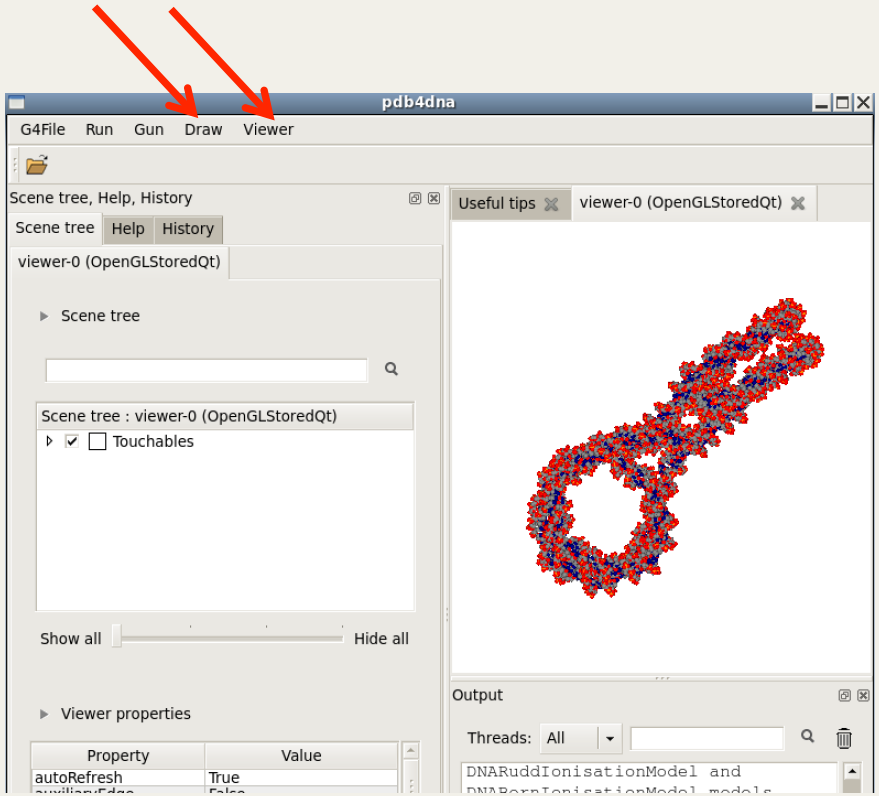
- Rename your PDB file (extension « pdb ») as a macro file (« mac » extension)

```
cp 1ZBB.pdb 1ZBB.mac
```

- Open GUI

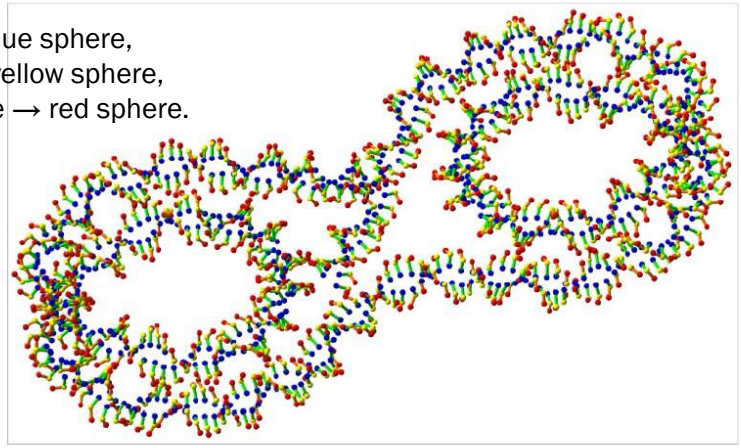
```
./pdb4dna -gui
```

- Open the 1ZBB.mac file
- Click on Viewer → Set style surface
- Click on Draw and try different options…

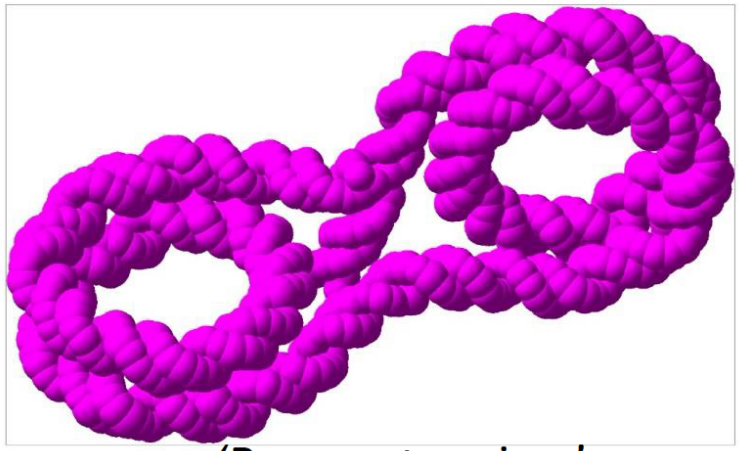


3 visualizations

- Base → blue sphere,
- Sugar → yellow sphere,
- Phosphate → red sphere.

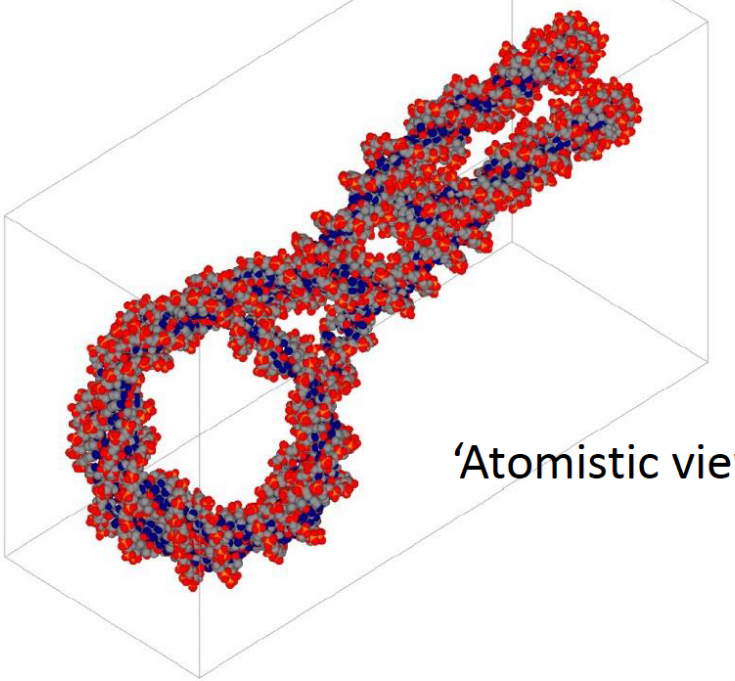


'Residue view'



'Barycenter view'

CPK coloring
- Hydrogen(H) → white sphere,
- Carbon(C) → gray sphere,
- Oxygen(O) → red sphere,
- Nitrogen(N) → dark blue sphere,
- Sulfur(S) → yellow sphere,
- Phosphorus(P) → orange sphere,
- others/undefined → pink sphere

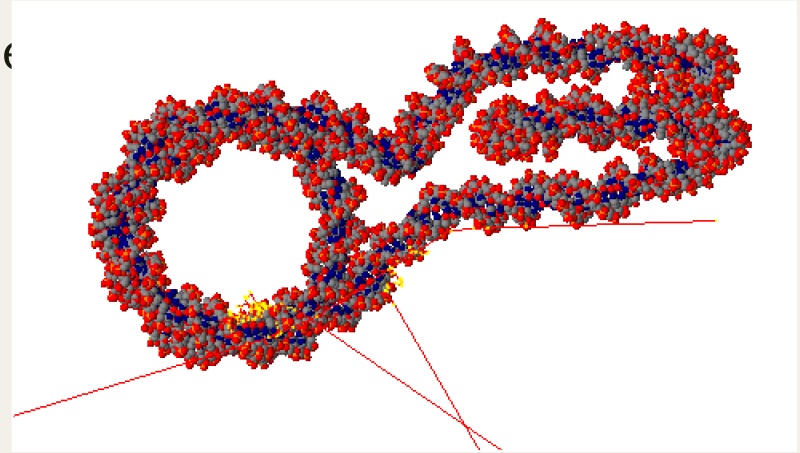


'Atomistic view'

Running with GUI

- Once you have loaded your PDB file, go to the

```
/gun/particle e-
/gun/energy 1 keV
/run/initialize
/run/beamOn 1
```



- Might be a bit slow with visualization
- Nice for visualization, but swtich to batch mode for energy deposition scoring…

Running in batch mode (1/2)

1. Prepare your macro file using an editor (nedit, geany)

– *You can alternatively use pdb4dna.in*

```
localhost.localdomain:/build-pdb4dna < 540 >more pdb4dna.in
#/control/execute vis.mac

/run/initialize

/tracking/verbose 0

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

/PDB4DNA/event/setEnergyThres 8.22 eV #default value: 8.22 eV
/PDB4DNA/event/setDistanceThres 10    #default value: 10

/gun/particle e-
/gun/energy 100 keV

/run/initialize

/run/printProgress 100

/run/beamOn 10000
localhost.localdomain:/build-pdb4dna < 541 >
```

Select your PDB file

Construct bounding volume

Set thresholds for strand breaks: energy, distance

Primaries

Run control

Running in batch mode (2/2)

2. Run the simulation

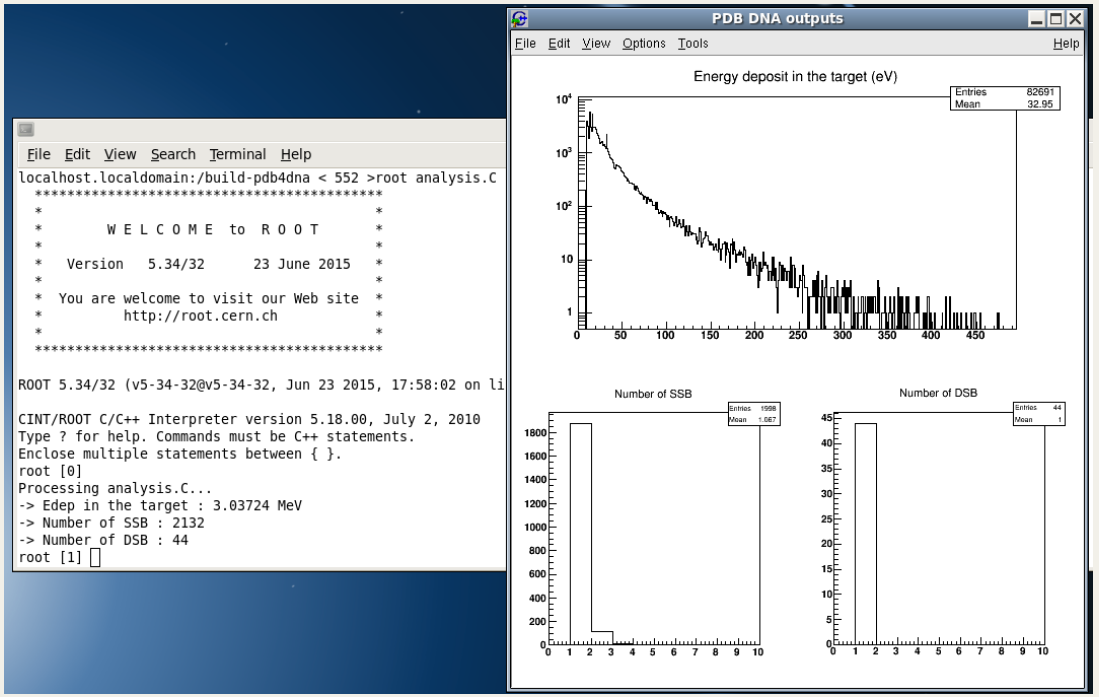
```
./pdb4dna -mac pdb4dna.in -mt 2
```

Macro file

Number of threads

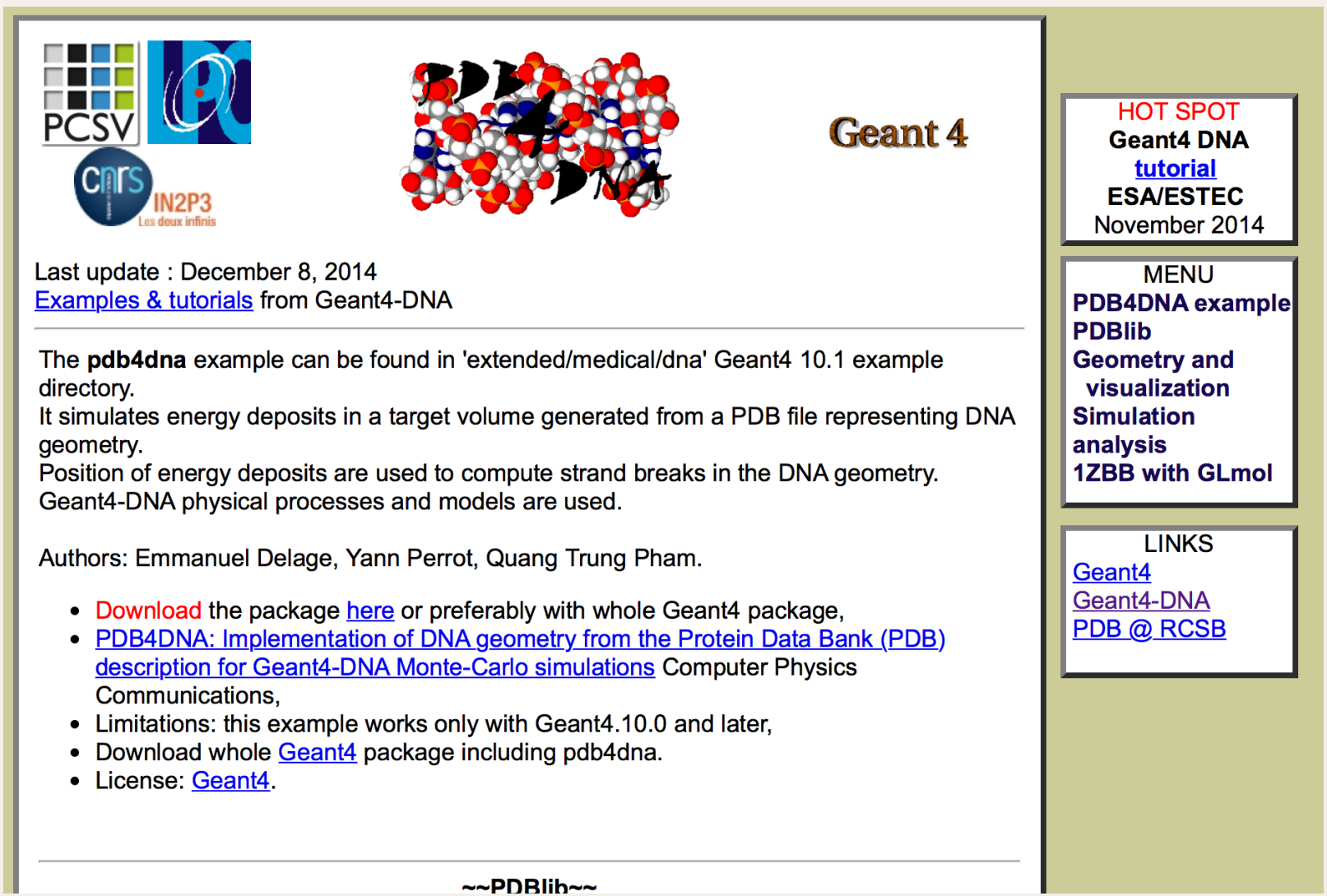
Analysis of results

- All results are stores in the file pdb4dna_output.root
- A ROOT macro file is provided for easy analysis: 3 histograms
  - *For each event*
- energy deposits in the bounding volume
- # SSB
- # DSB
- Quit Geant4
  `Idle> exit`
- Do
  `root analysis.C`



http://pdb4dna.in2p3.fr

Dedicated web site



PCSV

cnrs IN2P3 Les deux infinis

Geant 4

Last update : December 8, 2014
Examples & tutorials from Geant4-DNA

The **pdb4dna** example can be found in 'extended/medical/dna' Geant4 10.1 example directory.
It simulates energy deposits in a target volume generated from a PDB file representing DNA geometry.
Position of energy deposits are used to compute strand breaks in the DNA geometry.
Geant4-DNA physical processes and models are used.

Authors: Emmanuel Delage, Yann Perrot, Quang Trung Pham.

- Download the package here or preferably with whole Geant4 package,
- PDB4DNA: Implementation of DNA geometry from the Protein Data Bank (PDB) description for Geant4-DNA Monte-Carlo simulations Computer Physics Communications,
- Limitations: this example works only with Geant4.10.0 and later,
- Download whole Geant4 package including pdb4dna.
- License: Geant4.

~~PDBlib~~

HOT SPOT
Geant4 DNA
tutorial
ESA/ESTEC
November 2014

MENU
PDB4DNA example
PDBlib
Geometry and visualization
Simulation analysis
1ZBB with GLmol

LINKS
Geant4
Geant4-DNA
PDB @ RCSB