NAFlex: A web server for the study of Nucleic Acids Flexibility

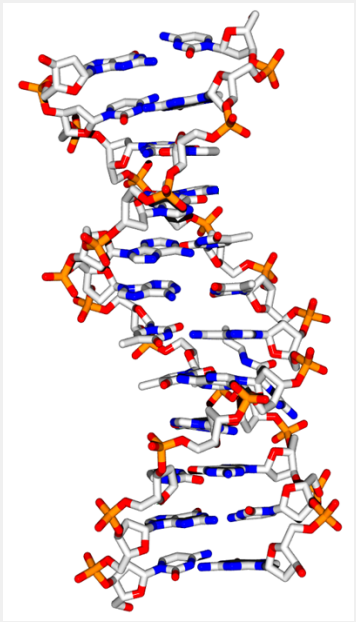
BioExcel Webinar, 20/07/2017

Federica Battistini, IRB-Barcelona

Nucleic Acids SEQUENCE

Nucleic Acids
SEQUENCE

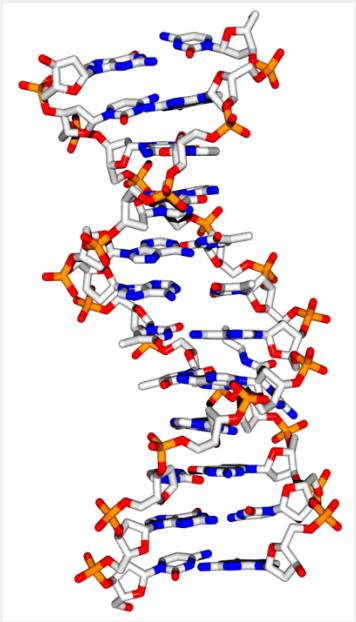
STRUCTURE



Structure depends on the sequence, not ideal B-DNA
Each base pair characterized by flexibility and geometry

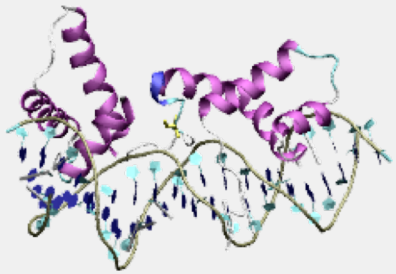
NA SEQUENCE

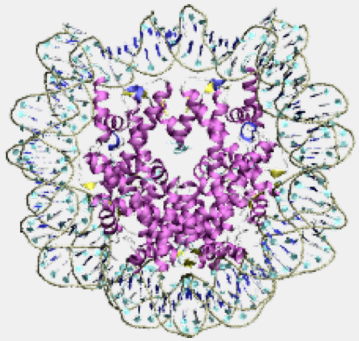
STRUCTURE



FUNCTION

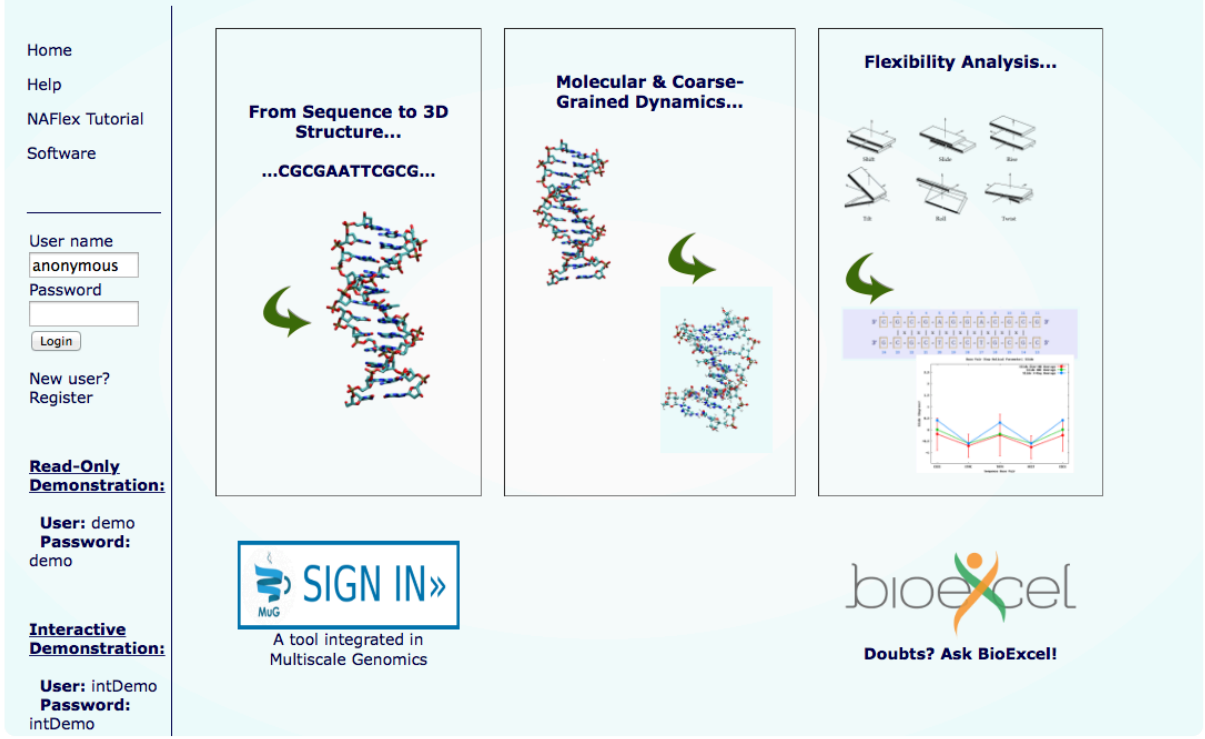
Protein Recognition
Protein-DNA binding
Genome organization
Expression control
.....





Structure depends on the sequence, not ideal B-DNA
Each base pair characterized by flexibility and geometry

Nucleic Acids Flexibility



mmb.irbbarcelona.org/NAFlex

Nucleic Acids Flexibility

NA flex

Input

From Sequence

From Structure

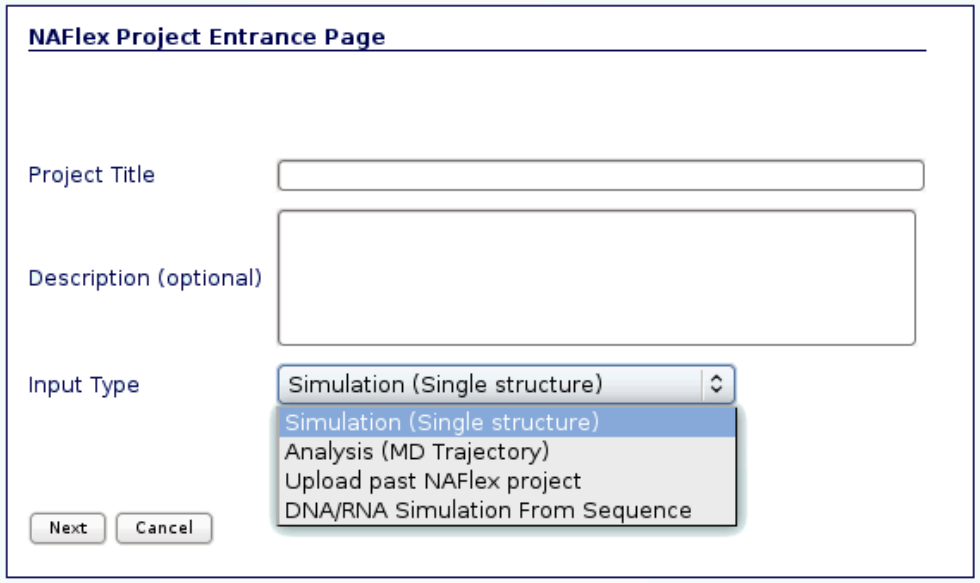
From Trajectory

From Saved Project

NAFlex Input Options

NAFlex offers four main entry options:

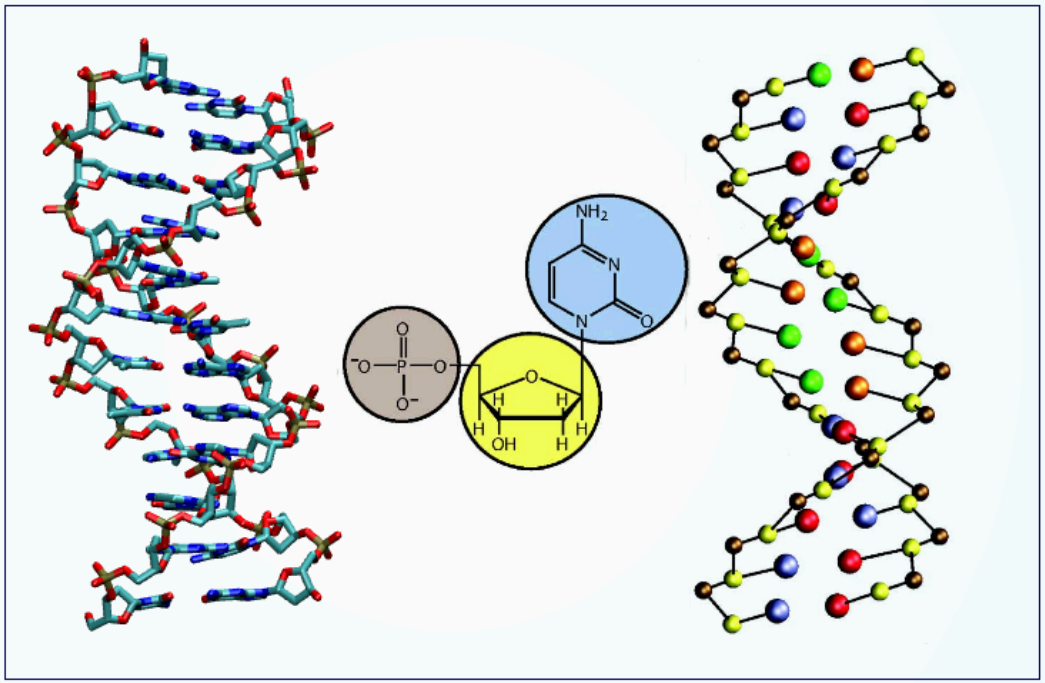
- **Simulation (Single Structure)**:
  Starts from a structure (PDB format), setup and run a MD simulation.
- **Analysis (MD Trajectory)**:
  Starts from a trajectory (usual MD trajectory formats), analyse, get information or convert between MD trajectory formats.
- **DNA/RNA Simulation from Sequence**:
  Starts from a sequence, obtain a 3D structure from a nucleic acid sequence.
- **Upload past NAFlex Project**:
  Upload a previously stored NAFlex project.



NAFlex Simulation Engines

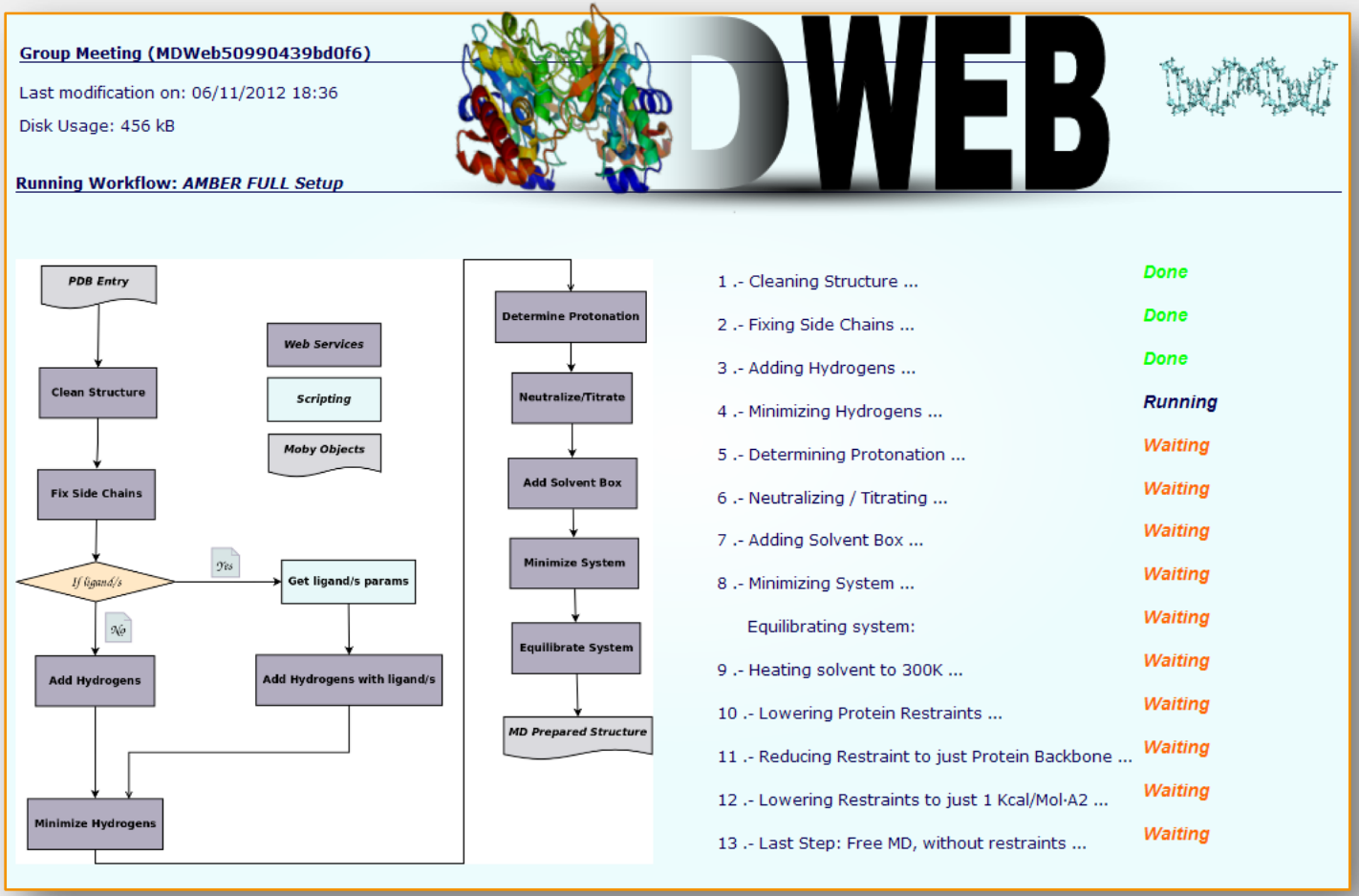
NAFlex offers a variety of simulation tools obtain dynamic information of nucleic acids. Basically the server has two main kind of **Simulation Engines**:

- **Atomistic Molecular Dynamics Simulations**:
  Molecular Dynamics Simulations with fully atomistic representation, in which all atoms (usually including the solvent's) are considered explicitly.
- **Coarse-Grained Simulations**:
  Coarse-Grained Simulations, where collections of atoms (number of them depending on the desired resolution) are represented by a few spherical beads connected by springs. **NAFlex** implements a couple of Coarse-grained algorithms, one at Nucleotide-Base Level (***Mesoscopic Elastic Model***) and another one at M x Base-Pair Level (***Worm-Like Chain Model***).



Atomistic Molecular vs Coarse-Grained Representation (J.Chem.Phys. 126, 2007)

Nucleic Acids MD (by MDWeb)

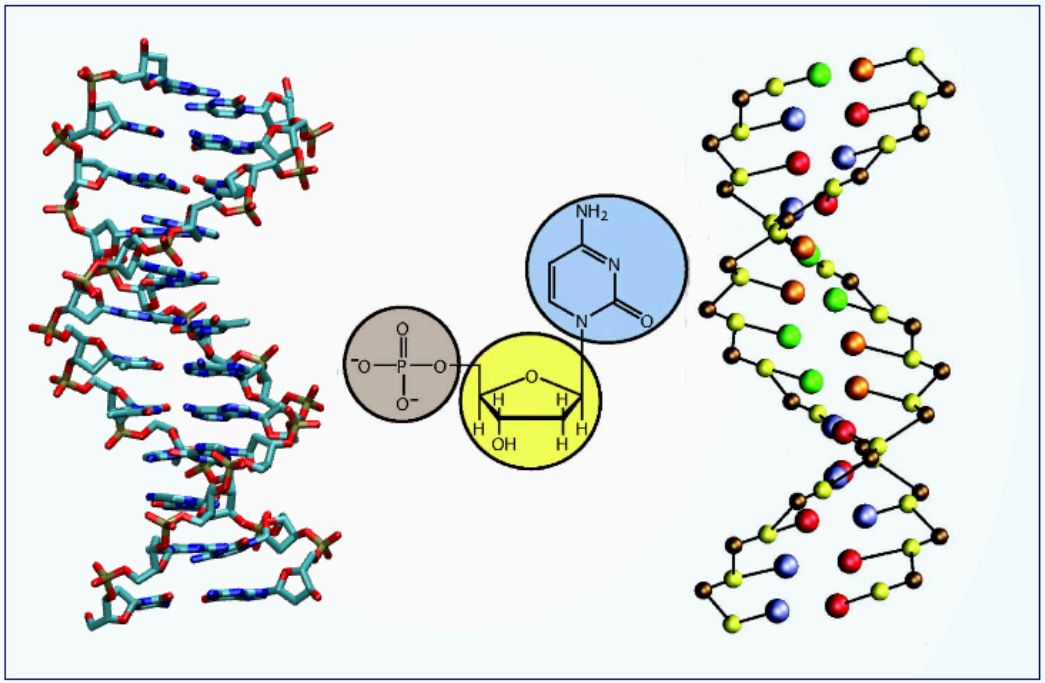


Hospital et al, Bioinformatics 2012, 28(9): 1278-1279

NAFlex Simulation Engines

NAFlex offers a variety of simulation tools obtain dynamic information of nucleic acids. Basically the server has two main kind of **Simulation Engines**:

- **Atomistic Molecular Dynamics Simulations**:
  Molecular Dynamics Simulations with fully atomistic representation, in which all atoms (usually including the solvent's) are considered explicitly.
- **Coarse-Grained Simulations**:
  Coarse-Grained Simulations, where collections of atoms (number of them depending on the desired resolution) are represented by a few spherical beads connected by springs. **NAFlex** implements a couple of Coarse-grained algorithms, one at Nucleotide-Base Level (***Mesoscopic Elastic Model***) and another one at M x Base-Pair Level (***Worm-Like Chain Model***).

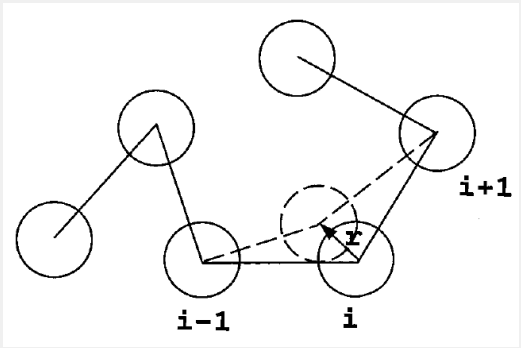


Atomistic Molecular vs Coarse-Grained Representation (J.Chem.Phys. 126, 2007)

NAFlex Coarse-Grained Dynamics

A Metropolis Montecarlo algorithm associated to DNA helical parameters is used to obtain coarse-grained DNA dynamics.

Goñi et al, Bioinformatics 2008, 24: 1731-1732



Worm-Like Chain (WLC), Montecarlo Algorithm.

1 bead for each 4 base-pair steps.

Debye Hückel Equation to take into account implicit solvent and ionic concentration (Salt-dependent).

Uniform charge of DNA.

DNA stretching, bending and torsion.

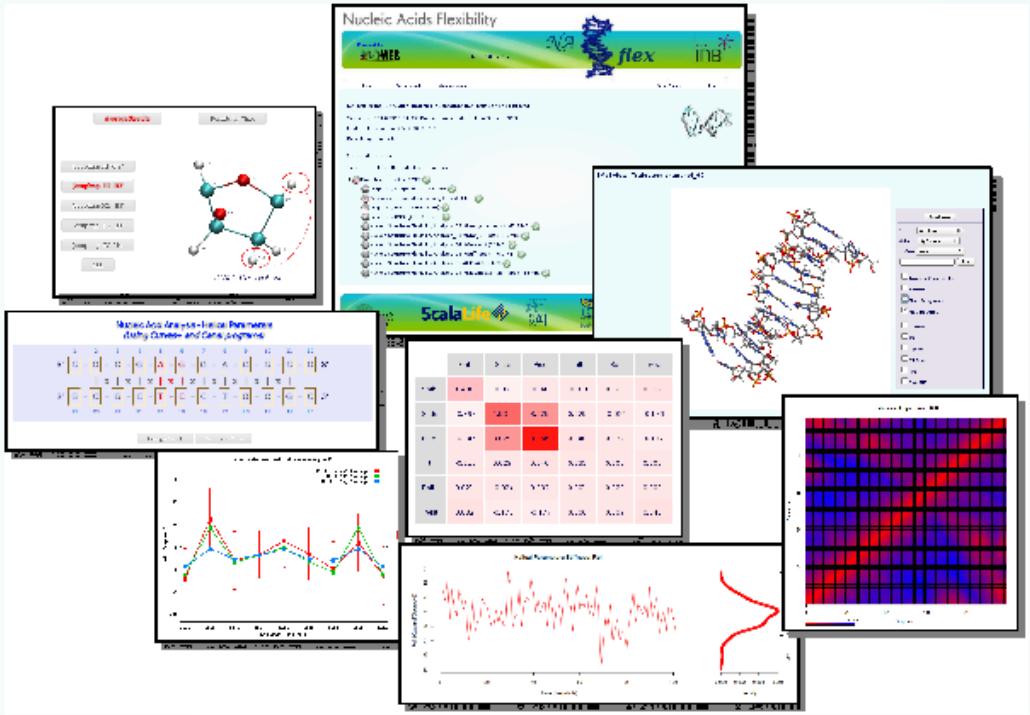
Jian et al, JCPhysics 1997, 136: 168-179

NA SEQUENCE-DEPENDENT PHYSICAL PROPERTIES

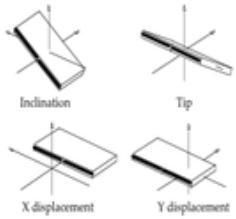
NAFlex Analysis Tools

NAFlex offers a variety of analysis packages for mining nucleic acids trajectories. They can be grouped in two main families:

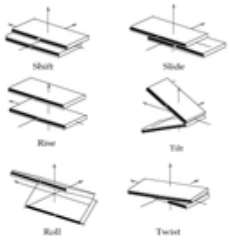
- **Standard Cartesian Analysis**:
  Basic trajectory cartesian analysis such as Root Mean Square Deviation (RMSd), RMSd per Nucleotide, Radius of Gyration (RGyr), Bfactors, etc.
- **Nucleic Acids Flexibility**:
  Set of different Flexibility analysis of Nucleic Acids: Helical Parameters, Principal Components (PCA), HB/Stacking energies, NMR-observables, etc.



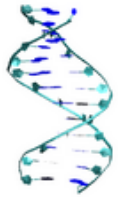
Trajectory Analyses >> *(Click to expand/shrink)*



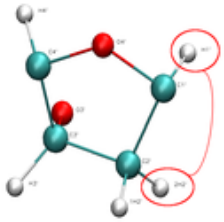
Curves Analysis



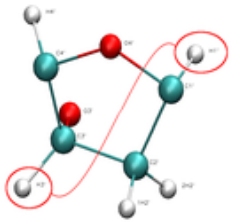
Stiffness Analysis



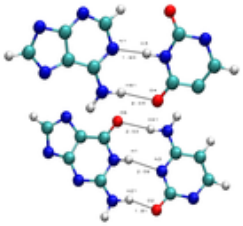
PCAzip Analysis



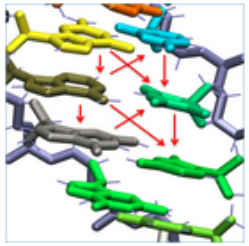
NMR_JC Analysis



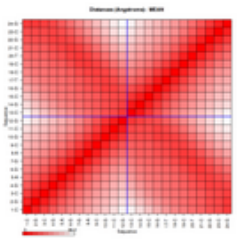
NMR NOEs Analysis



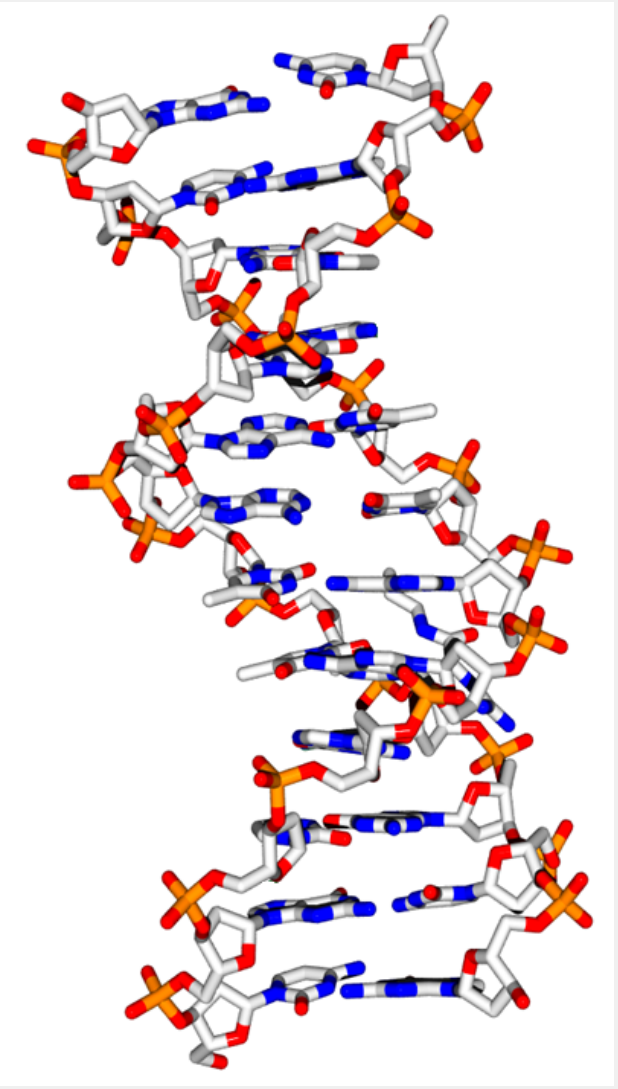
HBs Analysis



Stacking Analysis



Contacts Analysis



- Helical Parameters
- Stiffness Constants

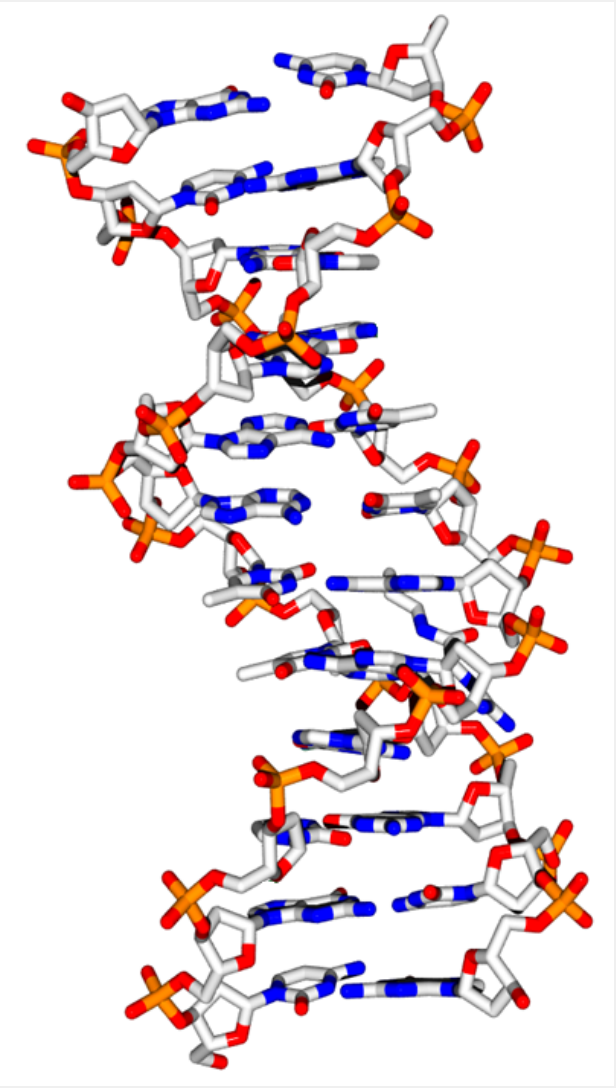
- Principal Component (PCA)

- Nuclear Magnetic Resonance Observables (J-Couplings, NOE's)

- Canonical Hydrogen Bond (HB) Analysis
- Atom Pairs Distances
- Backbone Analysis

- HB/Stacking Energies

- Distance Contact maps



- **Helical Parameters**
- **Stiffness Constants**

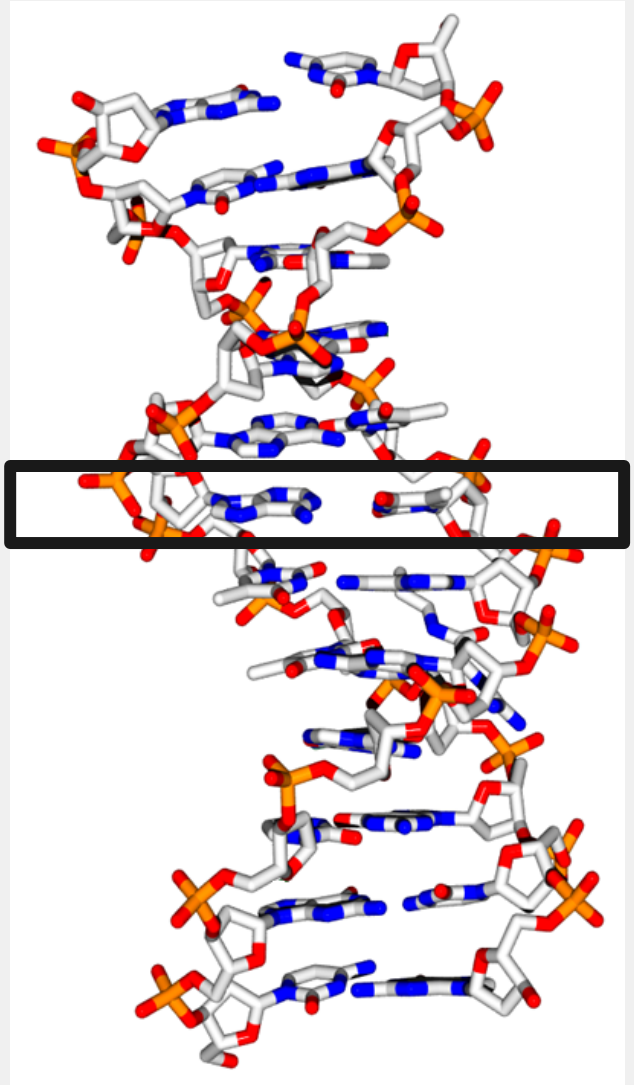
- Principal Component (PCA)

- Nuclear Magnetic Resonance Observables (J-Couplings, NOE's)

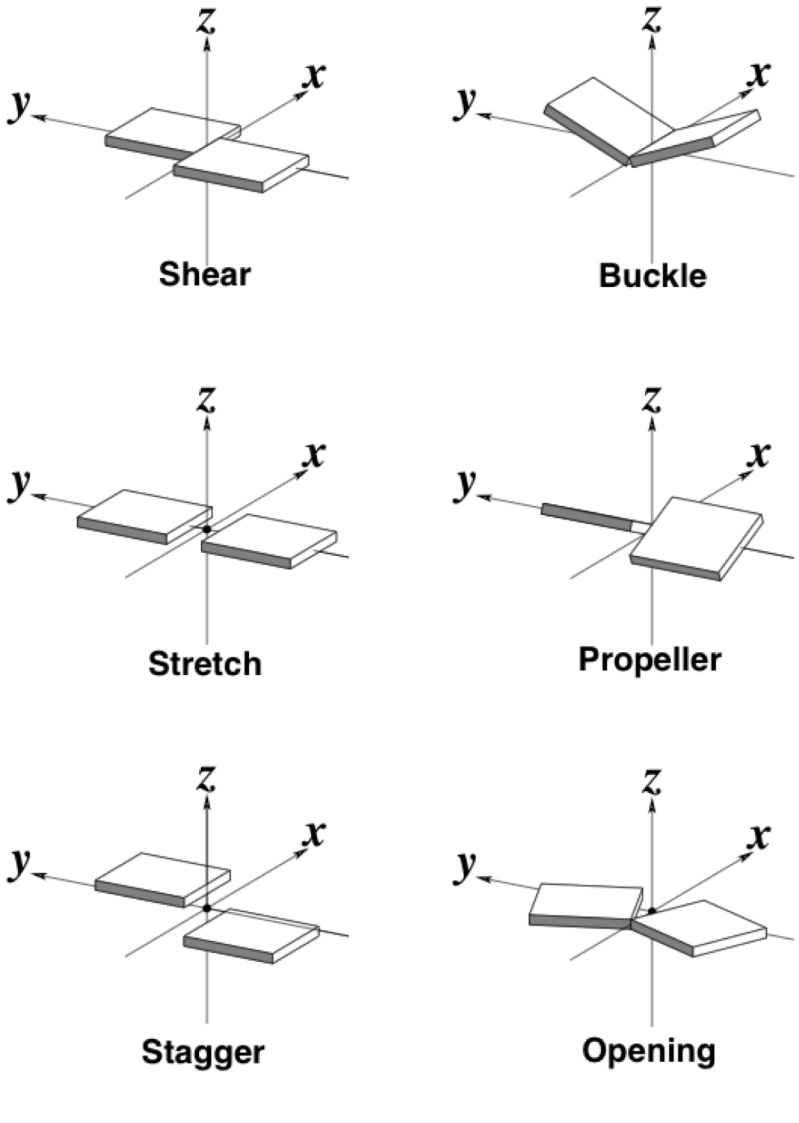
- Canonical Hydrogen Bond (HB) Analysis
- Atom Pairs Distances
- Backbone Analysis

- HB/Stacking Energies

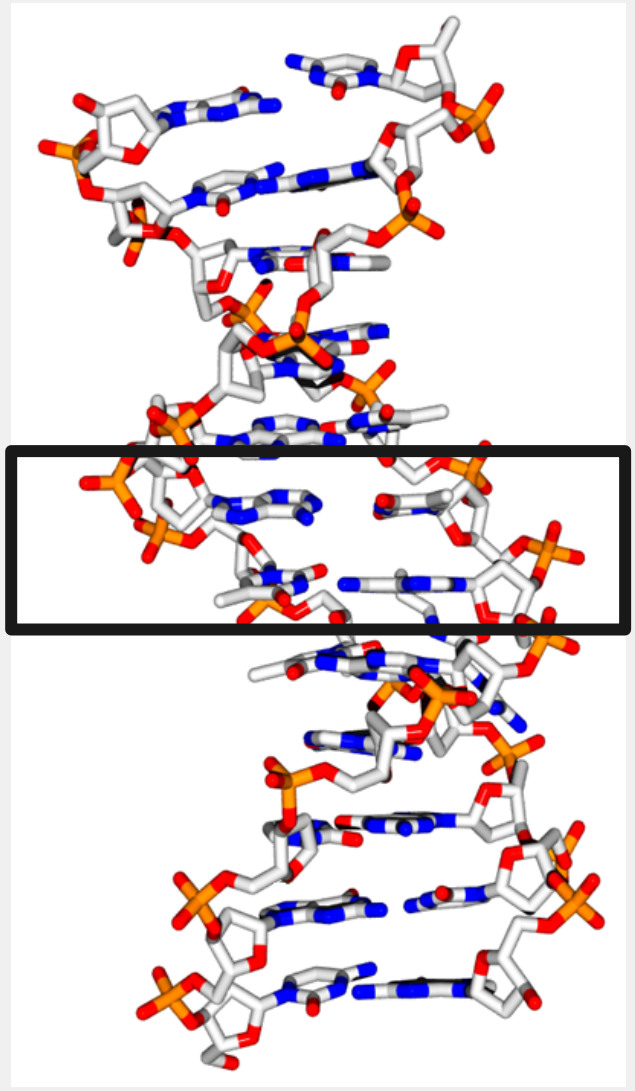
- Distance Contact maps

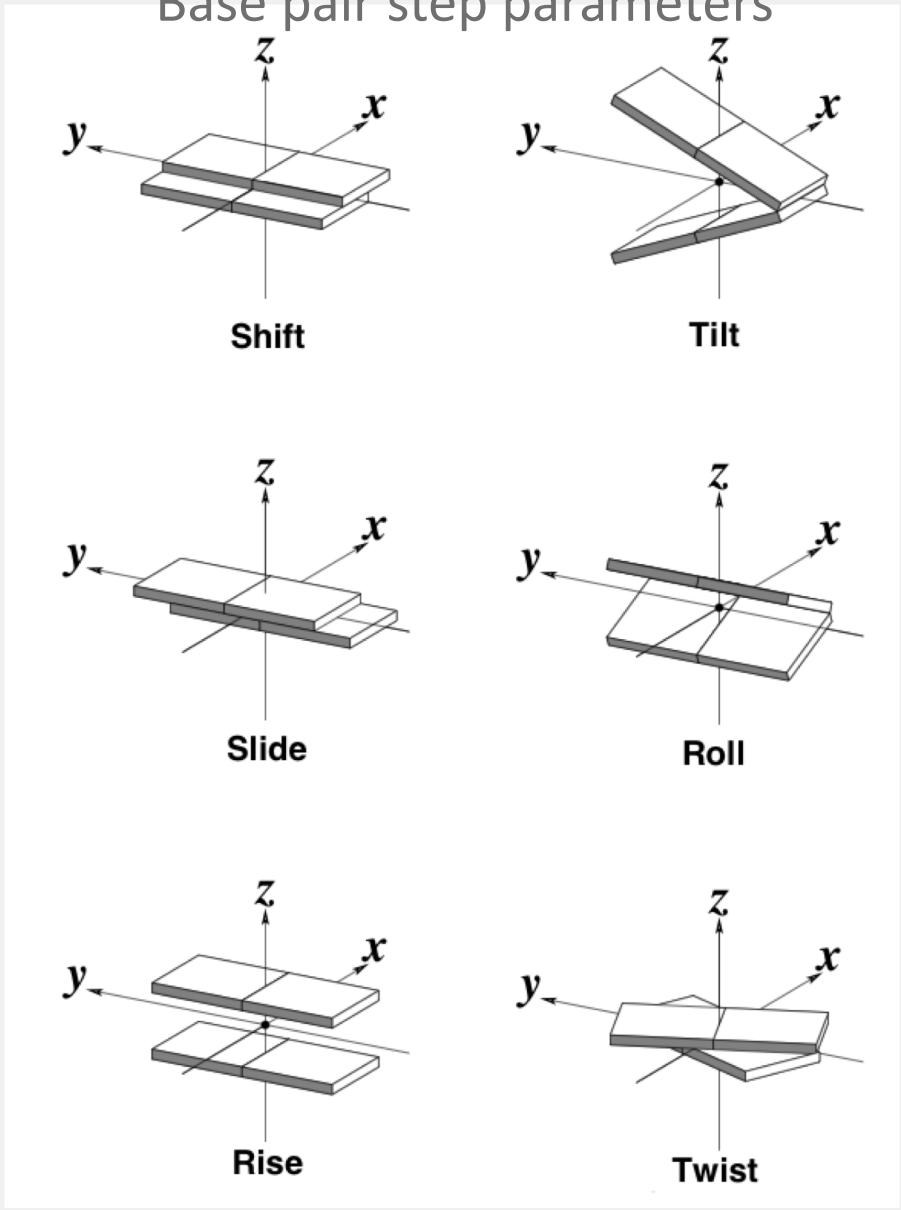


Base step parameters

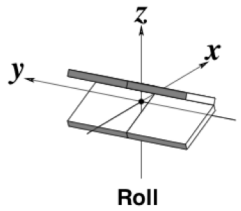


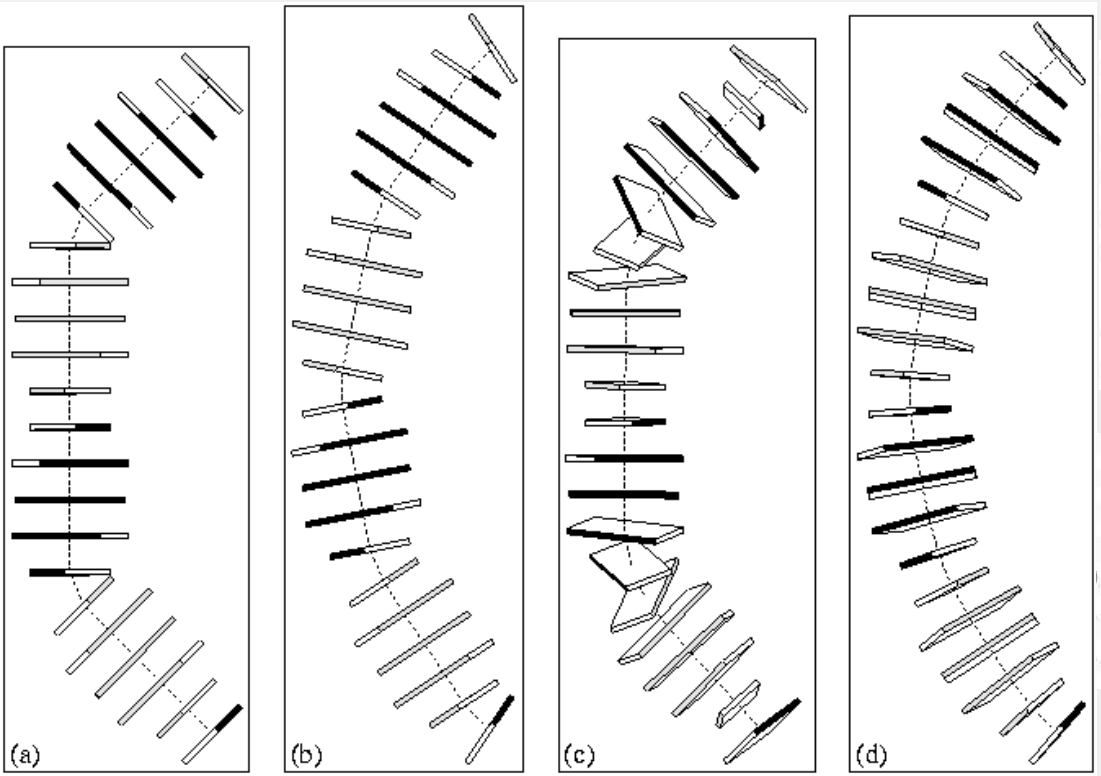
Base pair step parameters

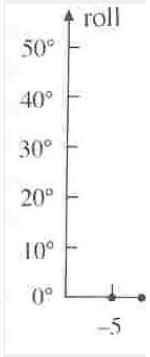


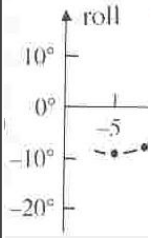


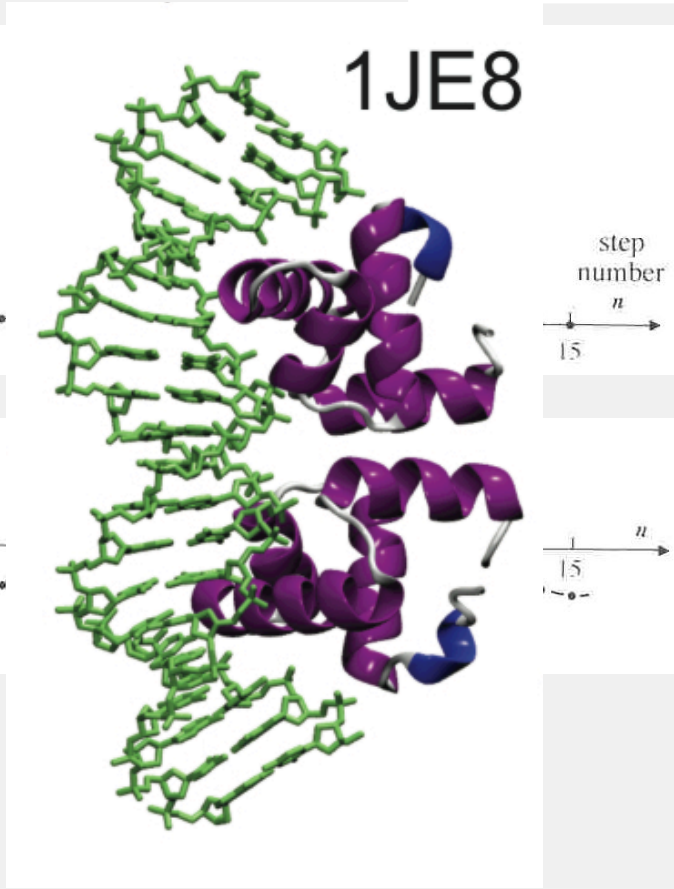
Roll angle and Bending







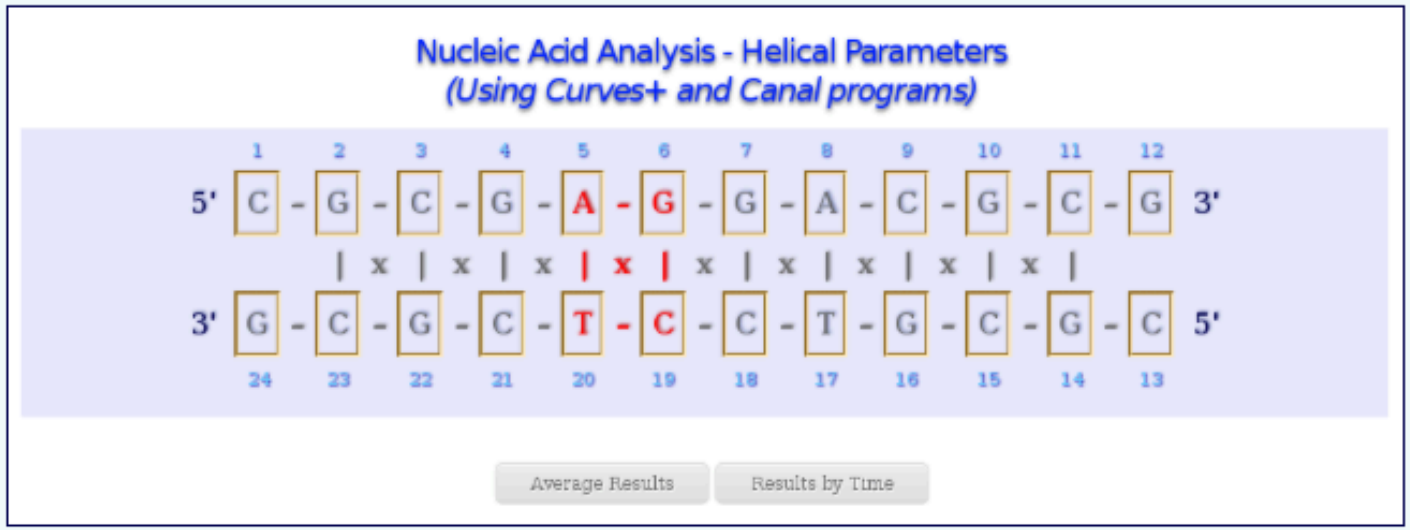




Calladine & Drew

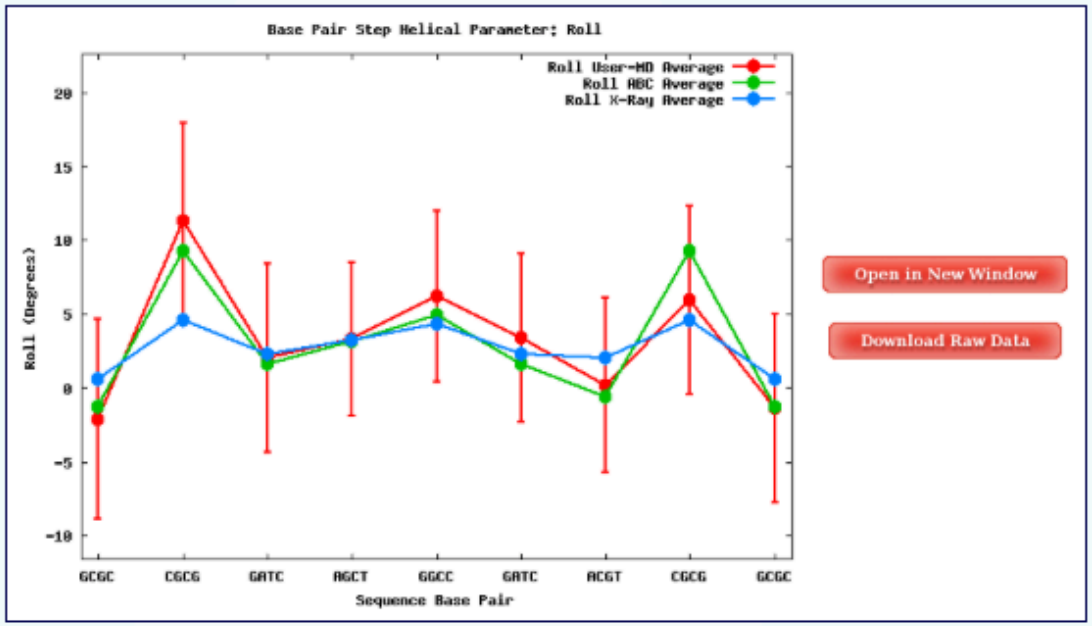
Helical Parameters

When analysing **Nucleic Acid Helical Parameters**, a graphical representation of the two sequence strands will appear on the screen. This representation allows a user-friendly selection of **nucleotides**, **base pairs** and **base pair steps** (tetramers), just clicking at the corresponding region of the sequence. For example, to select a base pair, user may click at the **|** symbol between nucleotides (from different strands) involved in the interesting base-pair, to select a base pair step, at the **x** symbol between nucleotides involved in the interesting tetramer, and just over a nucleotide one-letter code to select the corresponding nucleotide.



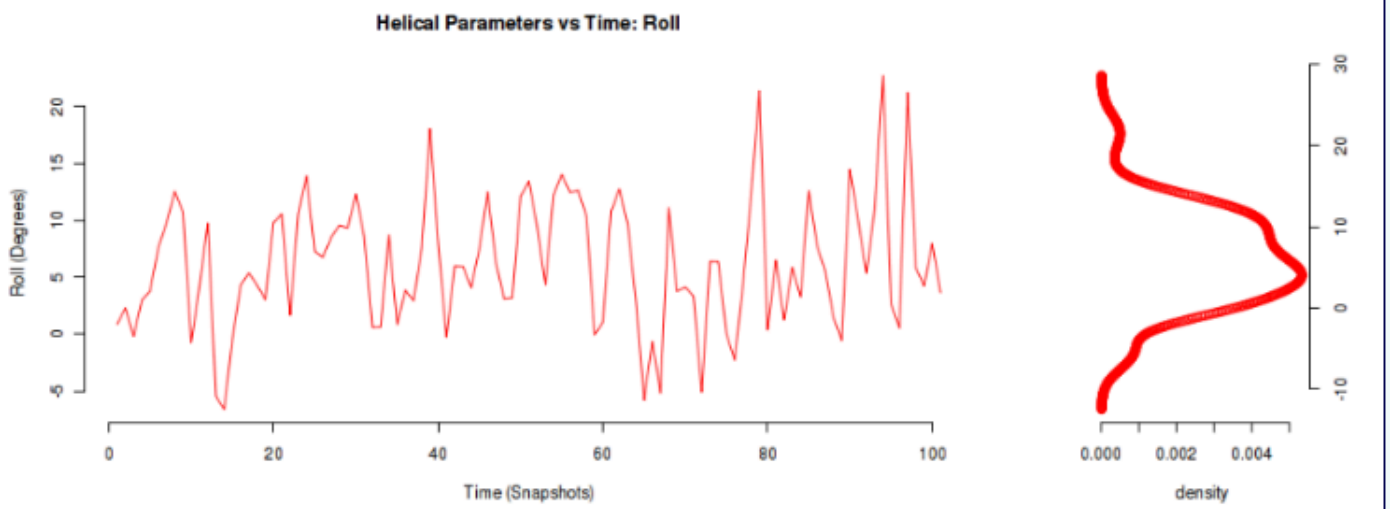
The **Average Results** section contains a set of plots generated after computing average values for all the trajectory snapshots. Plots and their associated raw data can be downloaded form the link provided.

When available, ***published data*** corresponding to the selected parameter set are plotted for comparison.



Plots generated contain the corresponding **histogram** attached, and an associated table with calculated **mean** and **standard deviation** values.

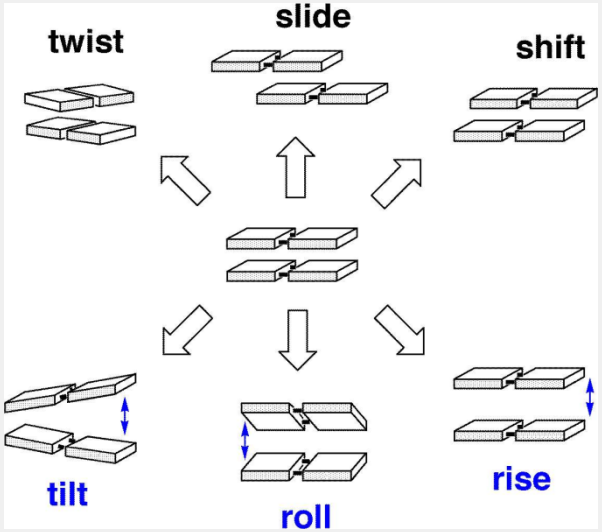
	Mean	Stdev
roll-6-GGCC	6.165	5.734



Open in New Window | Download Raw Data

Base pair step Stiffness Constants

$$(x - x_o)$$



Hooke's Law: $$\Delta E = \sum k(x - x_o)^2$$

$$\Theta = E(\Delta X)^{-2} = k_B T \subset^{-1} = \begin{pmatrix} k_w & k_{wr} & k_{wt} & k_{ws} & k_{wl} & k_{wf} \\ k_{wr} & k_r & k_{rt} & k_{rs} & k_{rl} & k_{rf} \\ k_{wt} & k_{rt} & k_t & k_{st} & k_{tl} & k_{tf} \\ k_{ws} & k_{rs} & k_{st} & k_s & k_{ls} & k_{lf} \\ k_{wl} & k_{rl} & k_{tl} & k_{ls} & k_l & k_{lf} \\ k_{wf} & k_{rf} & k_{tf} & k_{lf} & k_{lf} & k_f \end{pmatrix}$$

Stiffness Constants

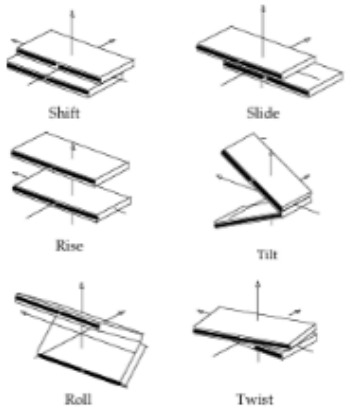
When analysing **Nucleic Acid Stiffness Constants**, the resulting web page will show again the graphical representation of the sequence strands *(see previous section)*.

The **Stiffness Constants** analysis is divided in two main sections:

- **Average** Results
- Results **by Time**

The **Average Results** section contains a set of **Stiffness Matrices** (determined by the inversion of the covariance matrix in helical space) for each sequence **Base Pair Step**.

	Shift	Slide	Rise	Tilt	Roll	Twist
Shift	1.430	-0.396	-0.045	-0.011	0.021	0.082
Slide	-0.397	4.931	3.029	0.029	-0.024	-0.176
Rise	-0.045	3.029	8.066	0.040	0.032	-0.177
Tilt	-0.011	0.029	0.040	0.033	0.001	0.006
Roll	0.021	-0.024	0.032	0.001	0.020	0.009
Twist	0.082	-0.176	-0.177	0.006	0.009	0.041



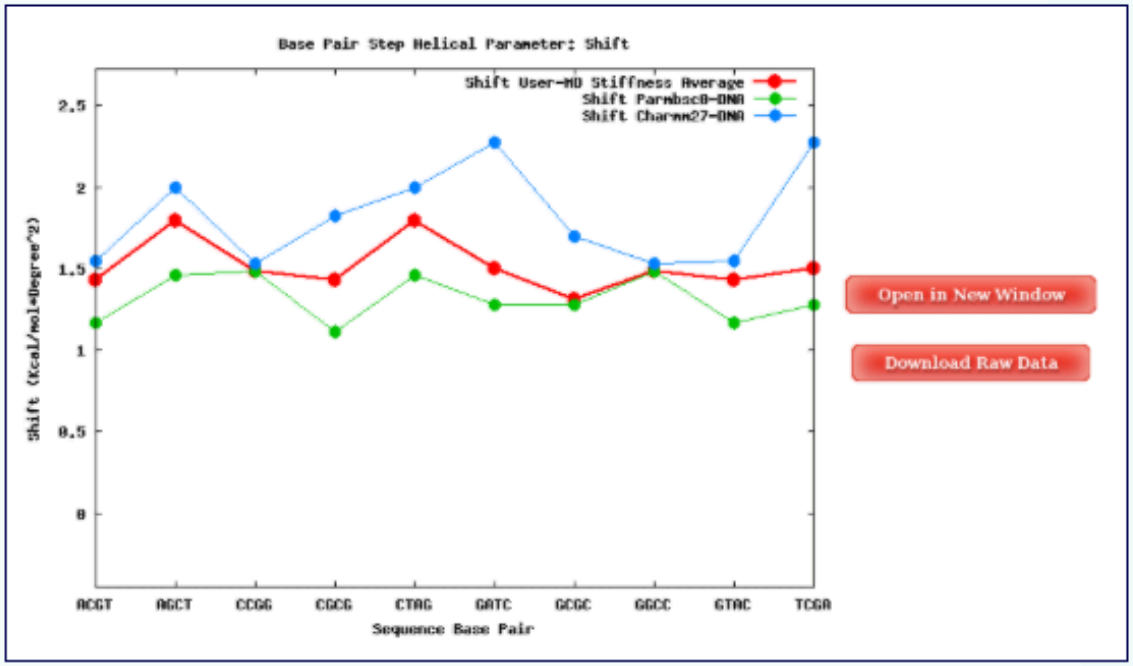
[Image courtesy of Curves+]

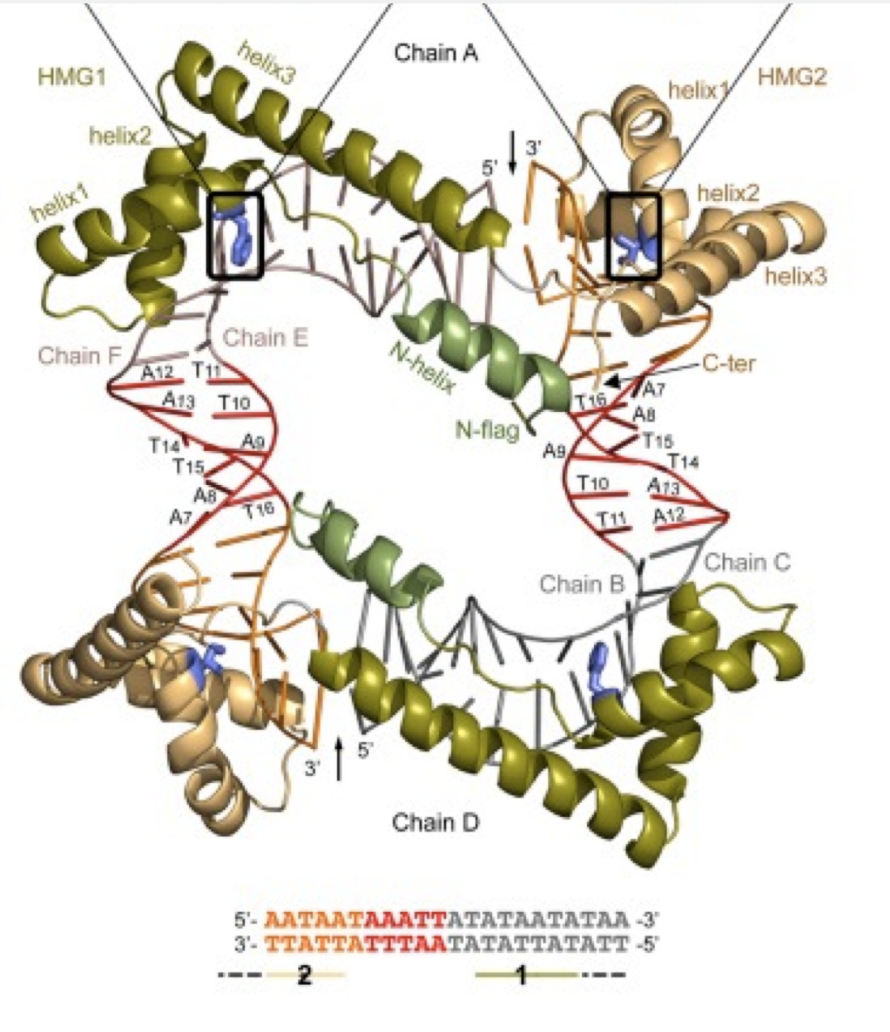
*Values in kcal/(mol*Å²)*

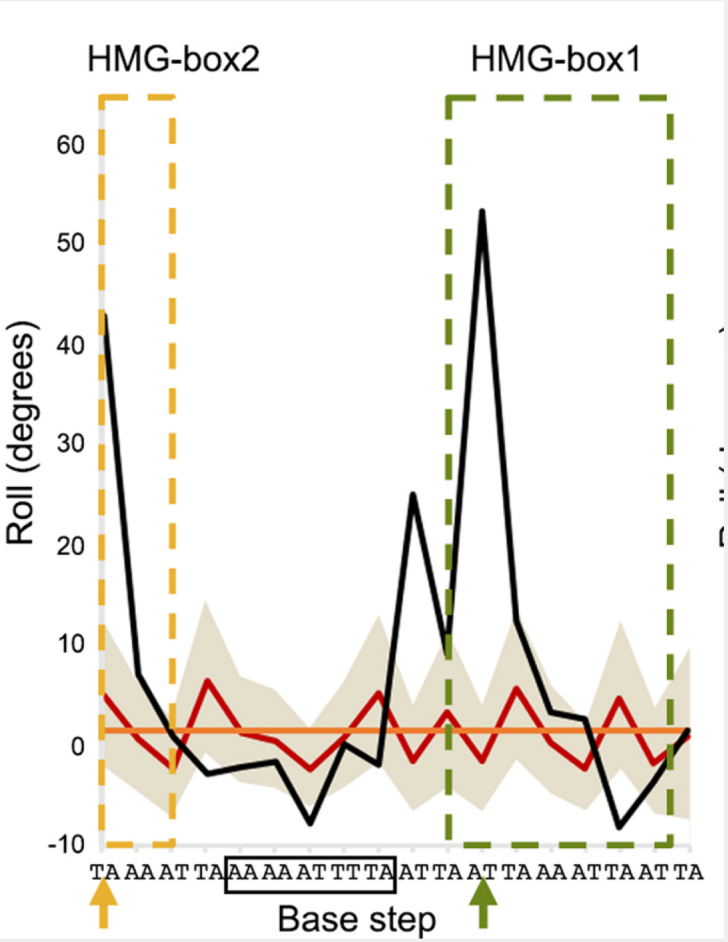
Download Raw Data | Open in New Window

A set of average plots generated computing average values for all the **Stiffness Constants** associated to a **Base Pair Step** (average of the diagonal elements of the **Stifness Matrices**) are also offered.

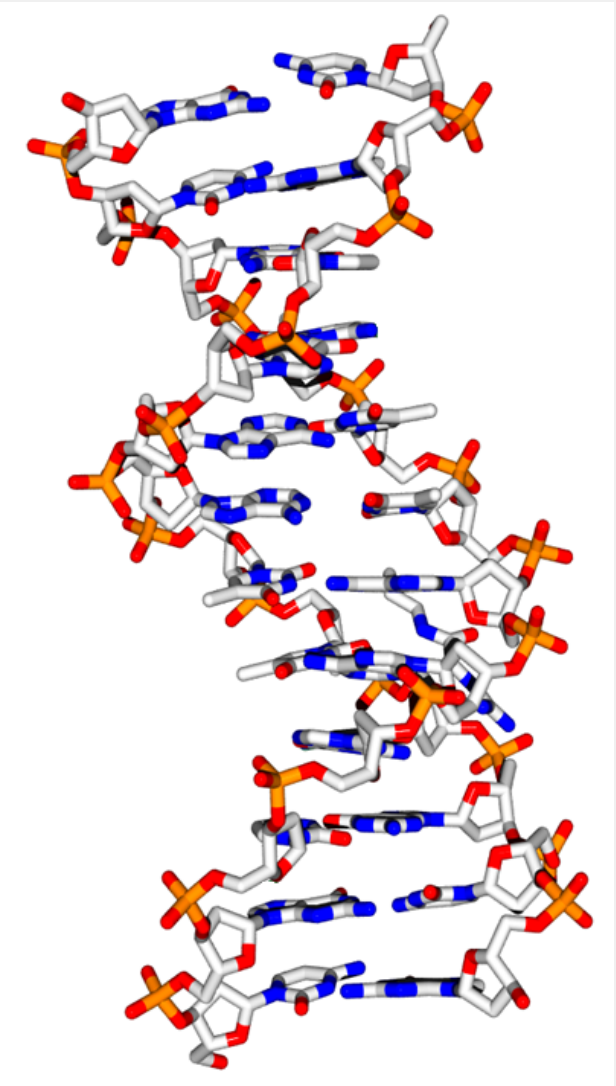
When available, published data corresponding to the selected parameter are plotted for comparison. Plots and its associated raw data can be downloaded from the link provided.







DNA structure directs positioning of the mitochondrial genome packaging protein Abf2p, Chakraborty, NAR (2017)



- Helical Parameters
- Stiffness Constants

- **Principal Component (PCA)**

- Nuclear Magnetic Resonance Observables (J-Couplings, NOE's)

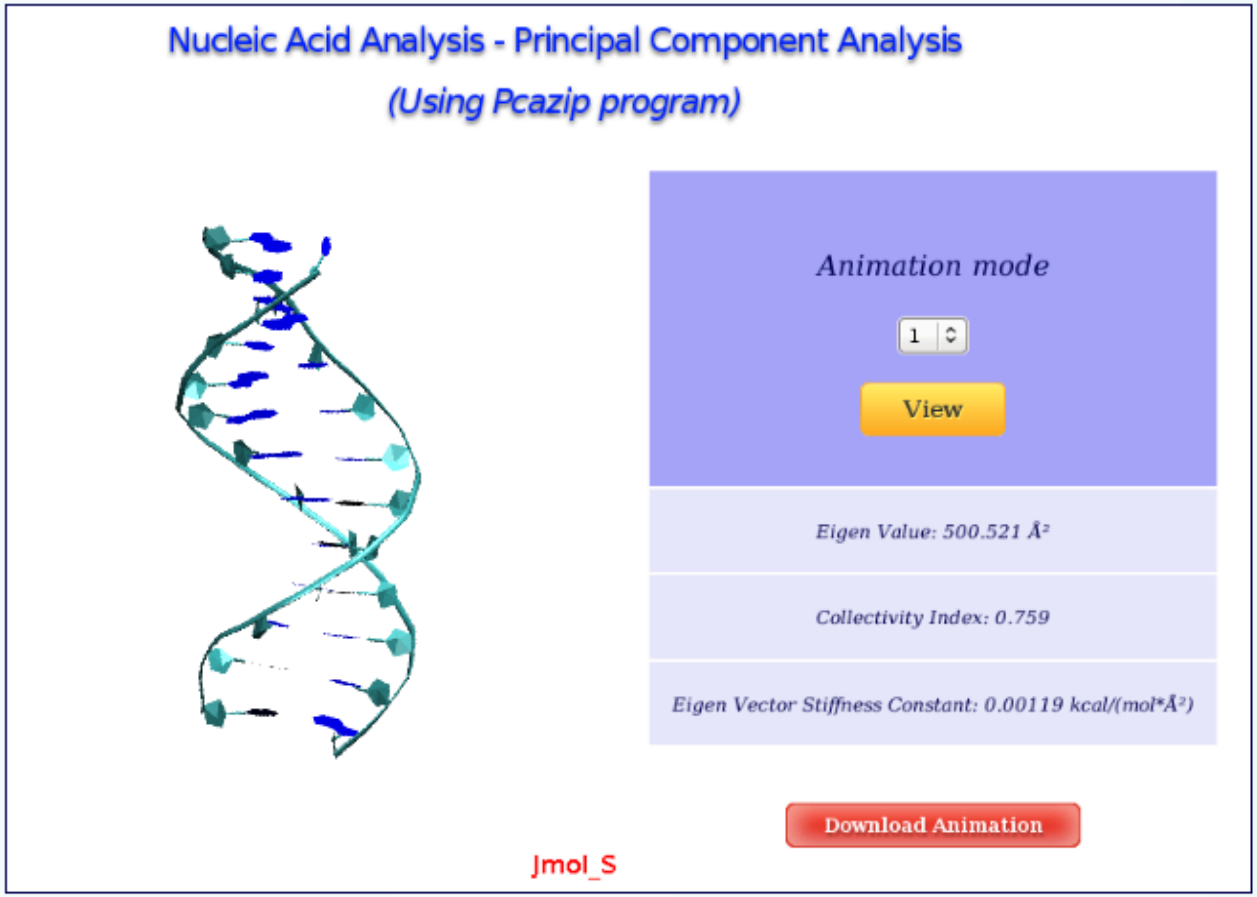
- Canonical Hydrogen Bond (HB) Analysis
- Atom Pairs Distances
- Backbone Analysis

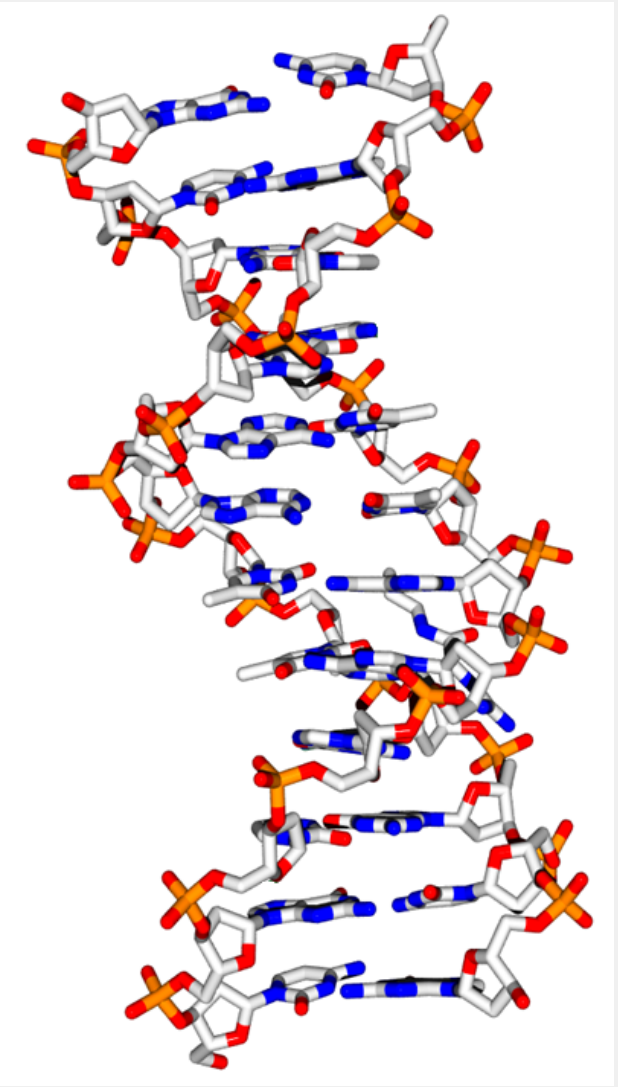
- HB/Stacking Energies

- Distance Contact maps

Principal Component Analysis

The **Principal Component Analysis** graphical interface offers the possibility of studying the real movements of the structure through the projections of the trajectory onto the different essential modes. An interactive **JMol applet** shows these movements, allowing user to translate, rotate and in general manipulate the visualization. The first 10 animation modes are offered for visualization and download. Associated values as **eigenvalues**, **collectivity indexes** and **eigenvector stiffness constants** are also shown.





- Helical Parameters
- Stiffness Constants

- **Principal Component (PCA)**

- **Nuclear Magnetic Resonance Observables (J-Couplings, NOE's)**

- Canonical Hydrogen Bond (HB) Analysis
- Atom Pairs Distances
- Backbone Analysis

- HB/Stacking Energies

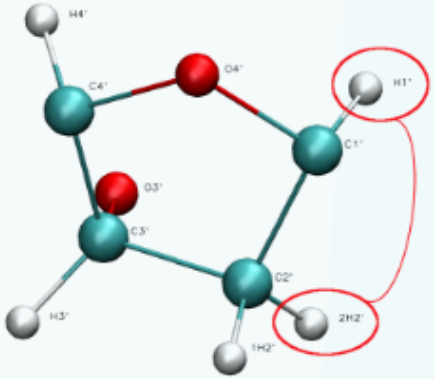
- Distance Contact maps

NAFlex Nucleic Acids Flexibility Analysis: Nuclear Magnetic Resonance Observables

J-Couplings

Vicinal ^{1}H-^{1}H Coupling Constants (^{3}J-Couplings) are scalar couplings between protons located three bonds away (H-C-C-H).

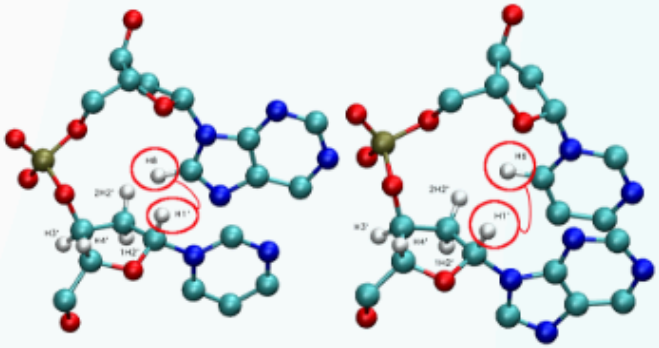
Vicinal J-Couplings are empirically correlated with the dihedral angle (H-C-C-H) through the so-called **Karplus equation**.

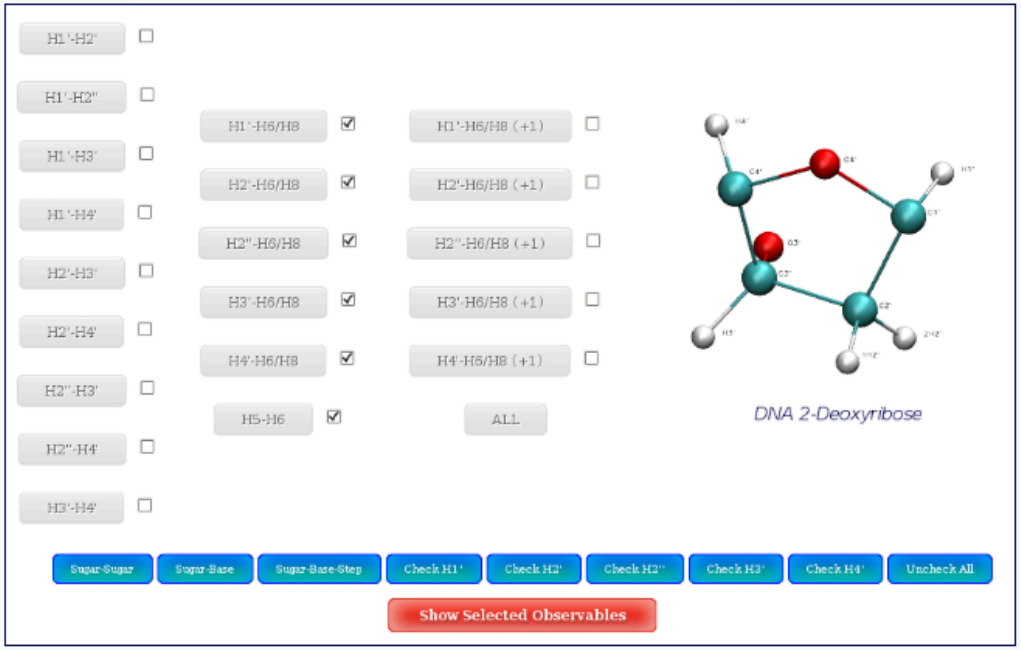


NOEs

Nuclear Overhauser Effect is the transfer of magnetization from one nuclear spin to another via cross-relaxation. The intensity of NOE cross-peaks between two particular protons depends on their relative distance (Inoe = $1/d^{6}$).

Proton-proton distances derived from NOEs are the most useful NMR parameters for structure elucidation.





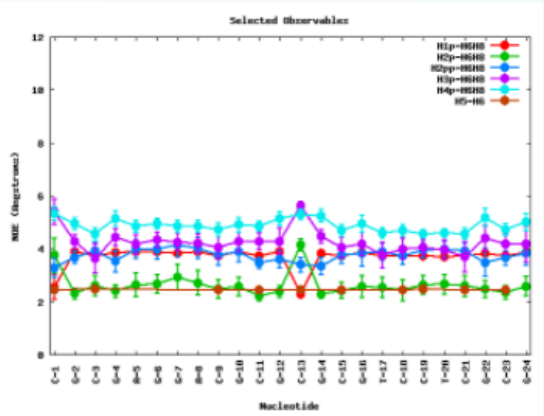
A set of buttons to automatically check interesting proton pairs are offered:

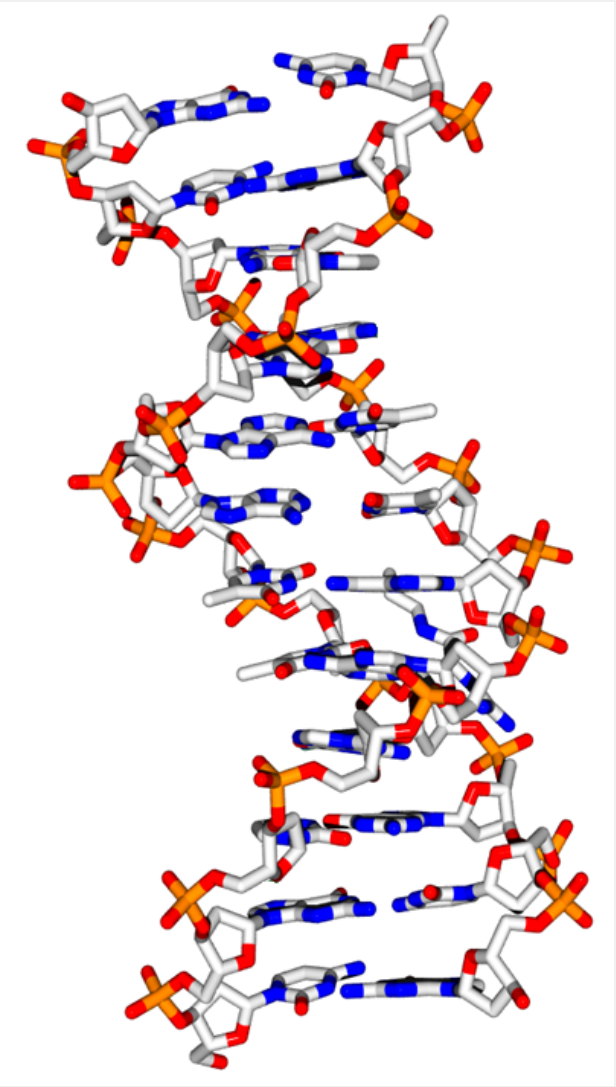
By proton pair position:

- **Sugar-Sugar:** Intra-Sugar proton pairs.
- **Sugar-Base:** Sugar-Base proton pairs.
- **Sugar-Base Step:** Sugar-Base Step proton pairs.

By specific proton:

- **H1':** All proton pairs involving H1'
- **H2'** All proton pairs involving H2'
- **H3'** All proton pairs involving H3'
- **H4'** All proton pairs involving H4'
- **H2''** All proton pairs involving H2'' *(Only in DNA case)*





- Helical Parameters
- Stiffness Constants

- Principal Component (PCA)

- Nuclear Magnetic Resonance Observables (J-Couplings, NOE's)

- **Canonical Hydrogen Bond (HB) Analysis**
- **Atom Pairs Distances**
- **Backbone Analysis**

- HB/Stacking Energies

- Distance Contact maps

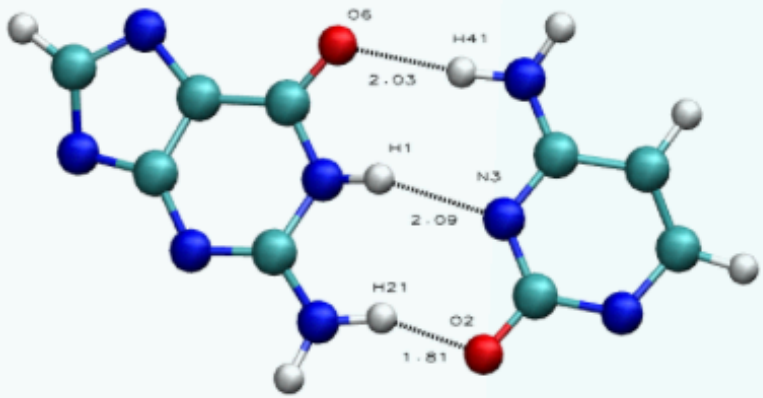
Canonical Hydrogen Bond Analysis

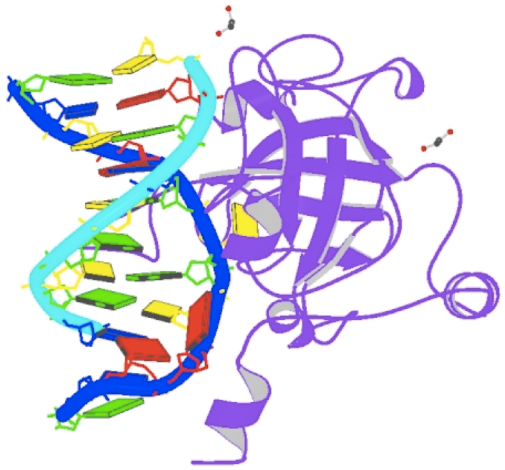
Purine-Pyrimidine HBs:

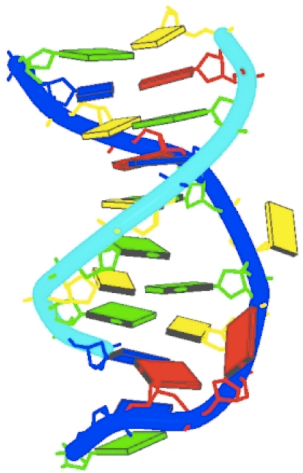
- Adenine N1 - Thymine/Uracil H3
- Adenine H61 - Thymine/Uracil O4

Pyrimidine-Purine HBs:

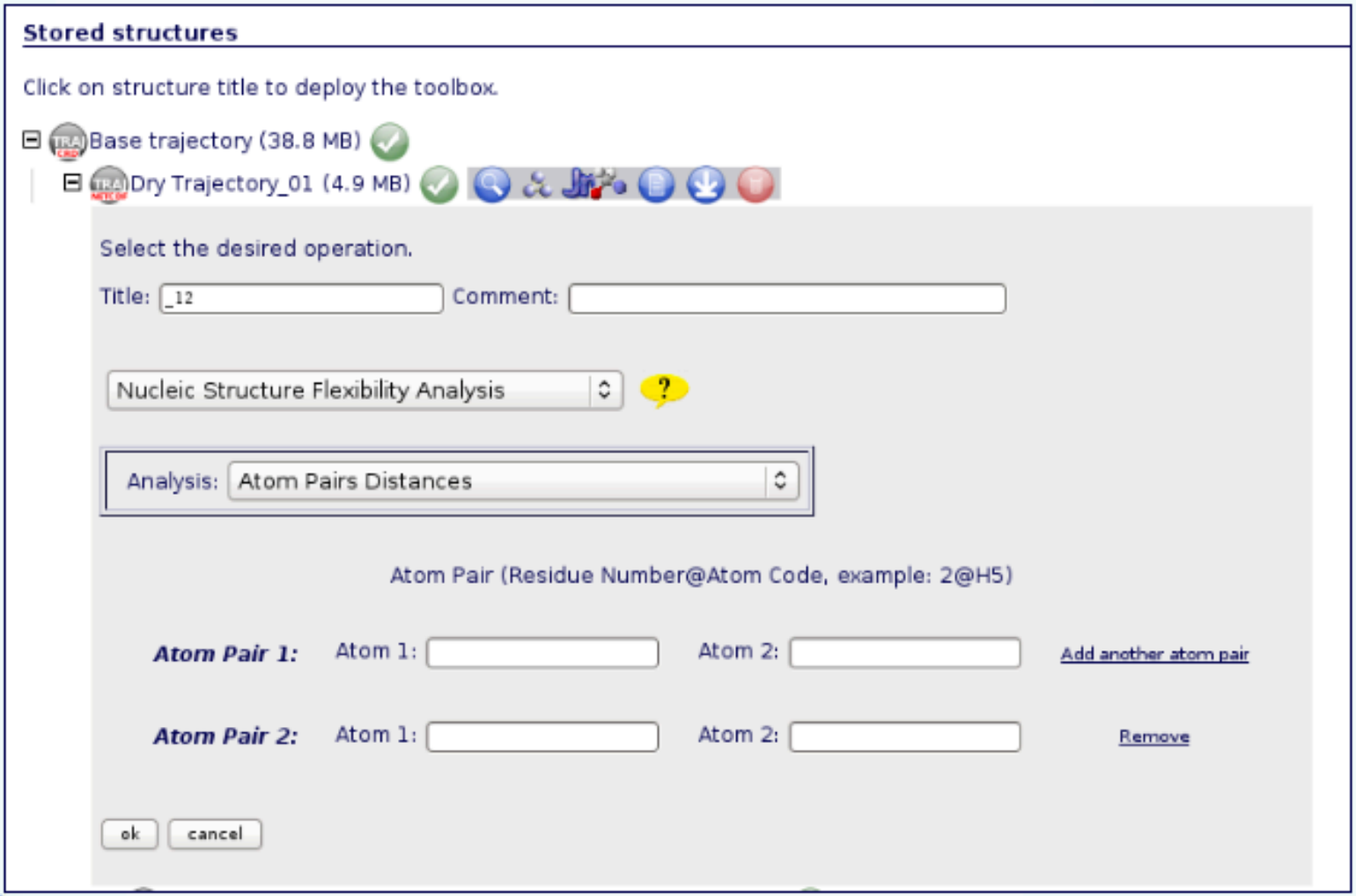
- Guanine O6 - Cytosine H41
- Guanine H1 - Cytosine N3
- Guanine H21 - Cytosine O2







PDB ID 2ZO1


Stored structures
Click on structure title to deploy the toolbox.
Base trajectory (38.8 MB)
Dry Trajectory_01 (4.9 MB)
Select the desired operation.
Title: _12
Comment:
Nucleic Structure Flexibility Analysis
Analysis: Atom Pairs Distances
Atom Pair (Residue Number@Atom Code, example: 2@H5)
Atom Pair 1:
Atom 1:
Atom 2:
Add another atom pair
Atom Pair 2:
Atom 1:
Atom 2:
Remove
ok
cancel

HB/Stacking Energies

Contact Maps

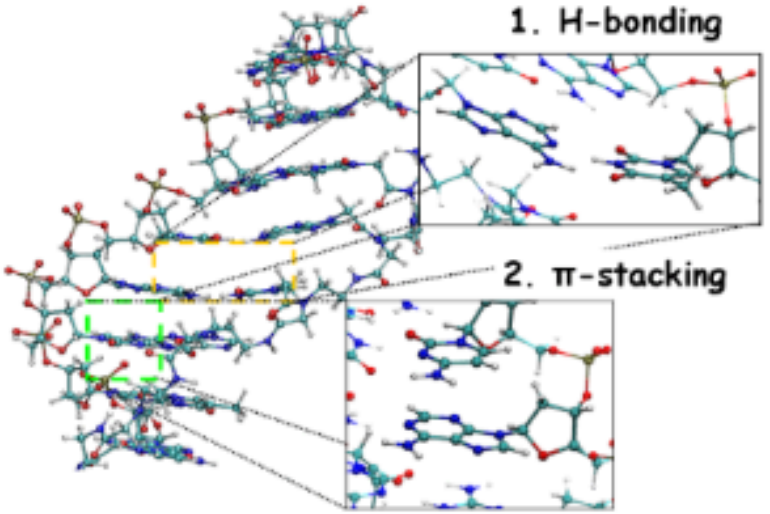
Results by Time

HB/Stacking Energies - Mean

HB/Stacking Energies - Min

HB/Stacking Energies - Max

HB/Stacking Energies - Stdev



HB/Stacking Energies

[Image courtesy of iOpenShell]

HB/Stacking Energies

Contact Maps

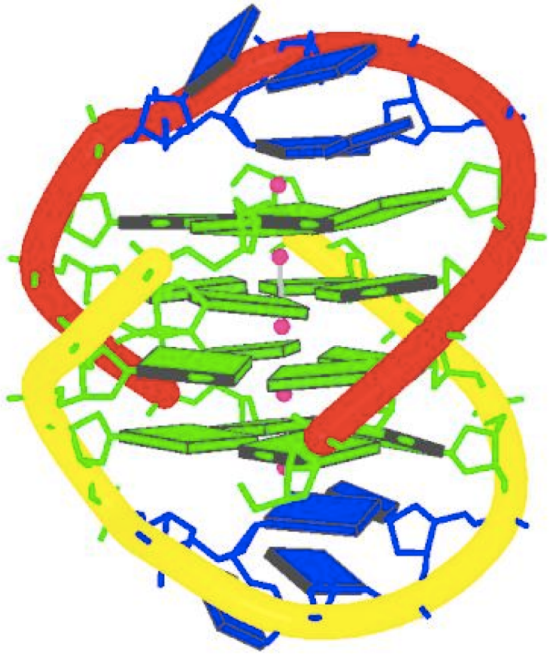
Results by Time

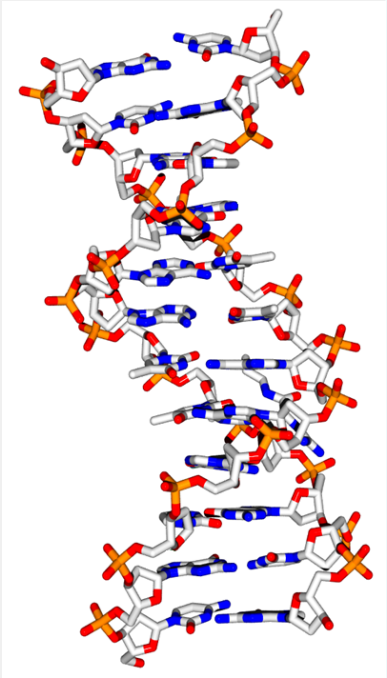
HB/Stacking Energies - Mean

HB/Stacking Energies - Min

HB/Stacking Energies - Max

HB/Stacking Energies - Stdev





Backbone Torsions

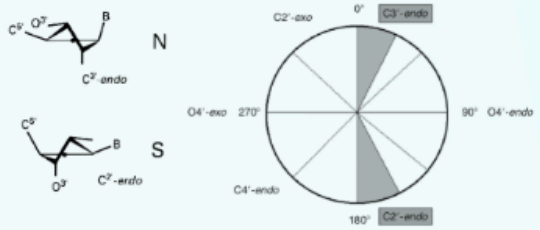
The three major elements of flexibility in the backbone are:

- **Sugar Puckering**

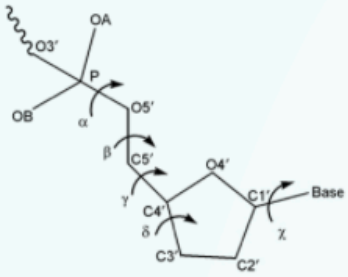
Sugar Puckering annotation is done by dividing the pseudo-rotational circle in four equivalent sections:

- North: 315:45º
- East: 45:135º
- South: 135:225º
- West: 225:315º

These four conformations are those dominating sugar conformational space, in agreement with all available experimental data.



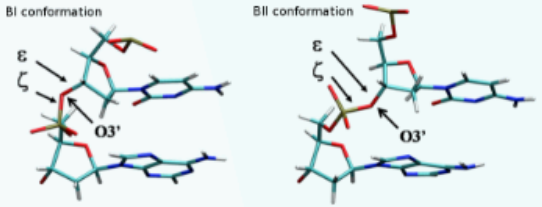
- **Canonical Alpha-Gamma**



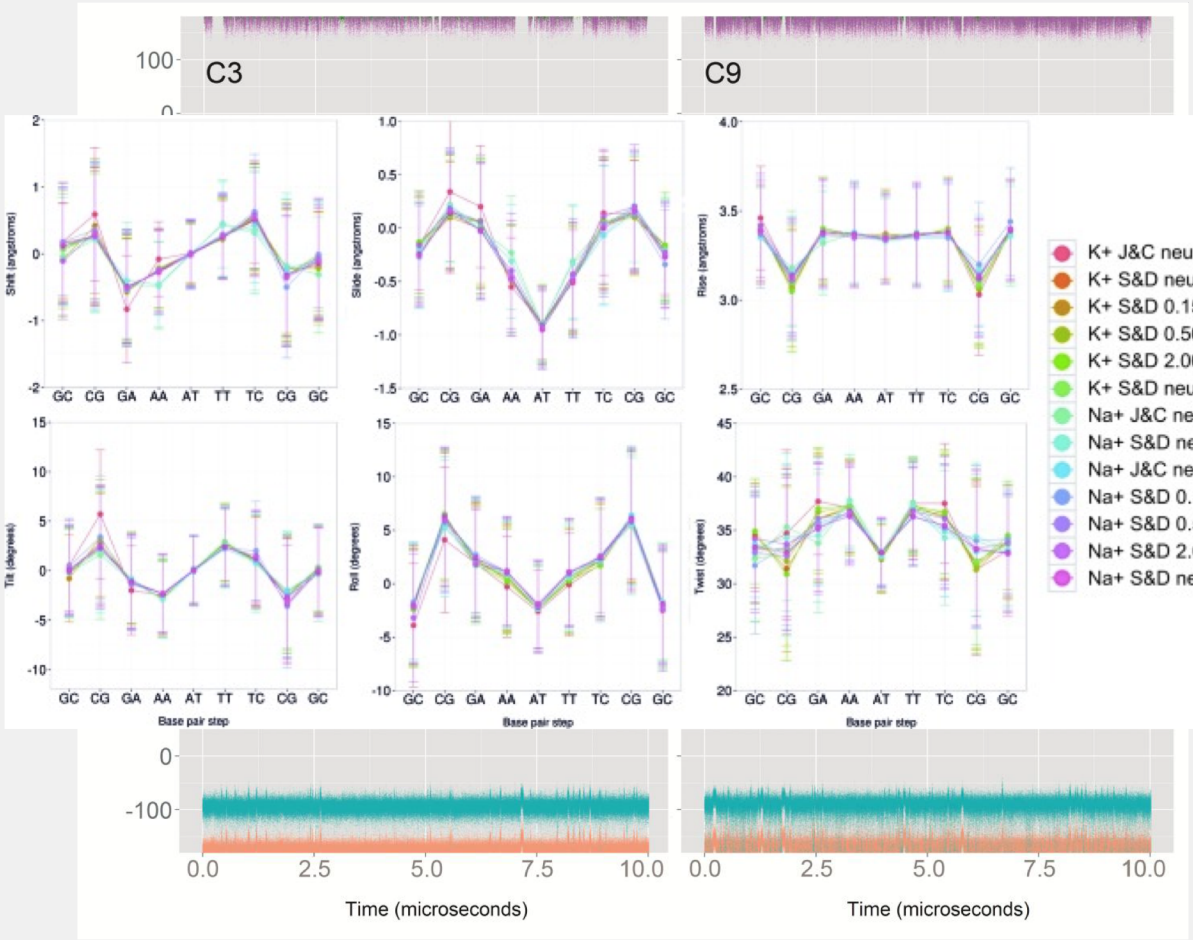
Rotations around **α/γ** torsions generate non-canonical local conformations leading to a reduced twist and they have been reported as being important in the formation of several protein-DNA complexes.

- **BI-BII Population**

The concerted rotation around **ζ/ε** torsions generates two major conformers: **BI and BII**, which are experimentally known to co-exist in a ratio around **80%:20% (BI:BII)** in B-DNA.

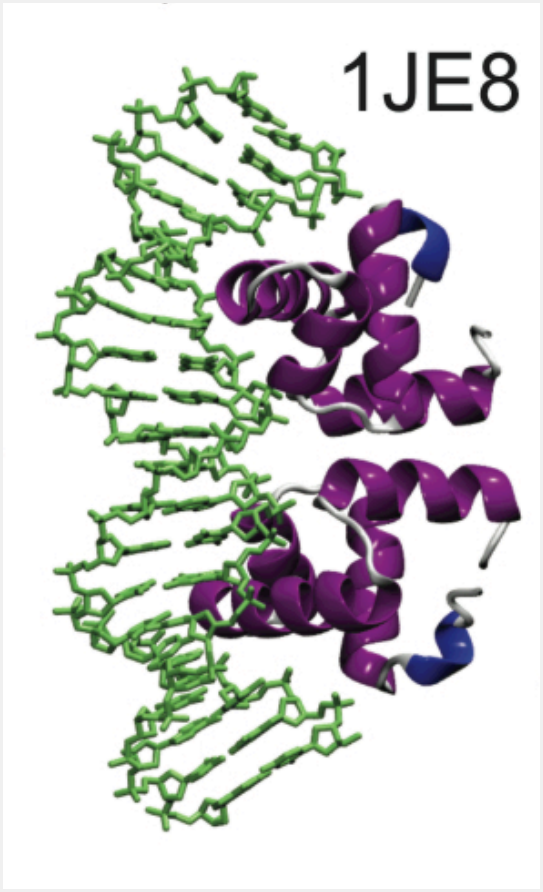


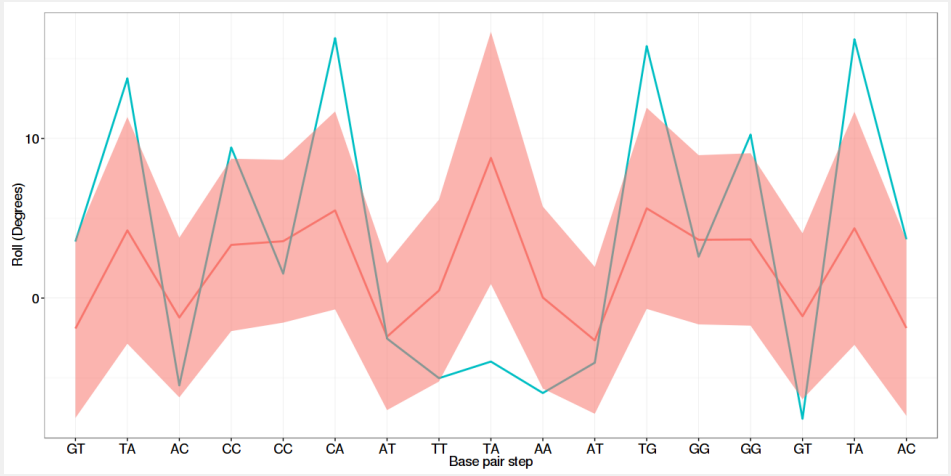
ANALYSIS DNA DREW DICKERSON DODECAMER

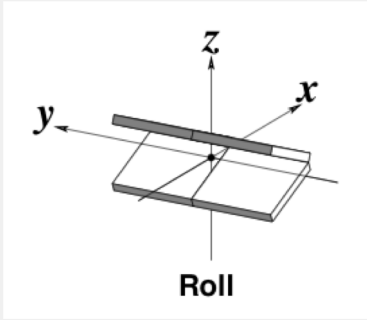


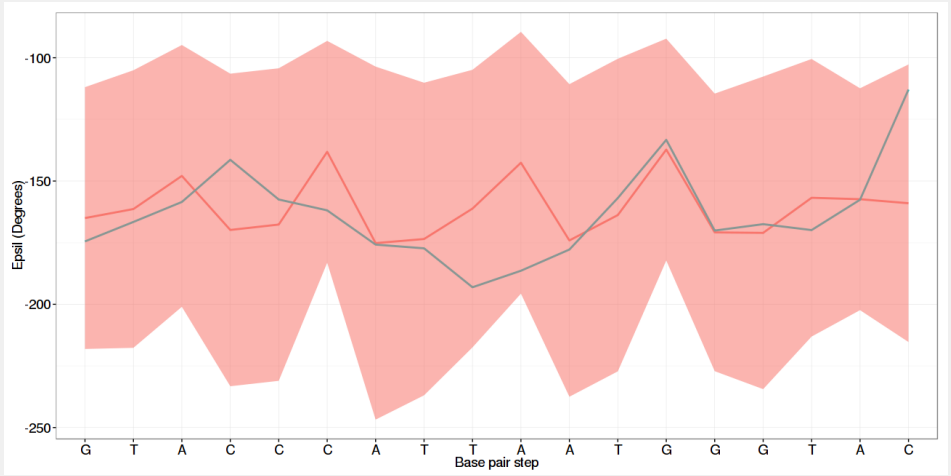
Long-timescale dynamics of the Drew Dickerson dodecamer, DANS PD, NAR (2016).

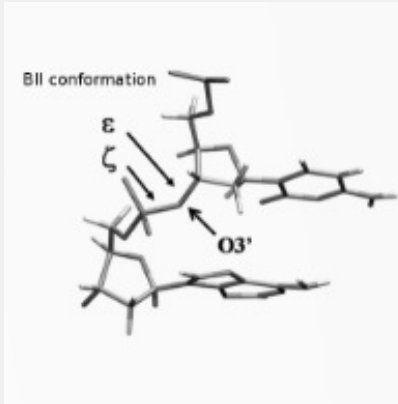
ANALYSIS DNA: PROTEIN-BOUND VS UNBOUND









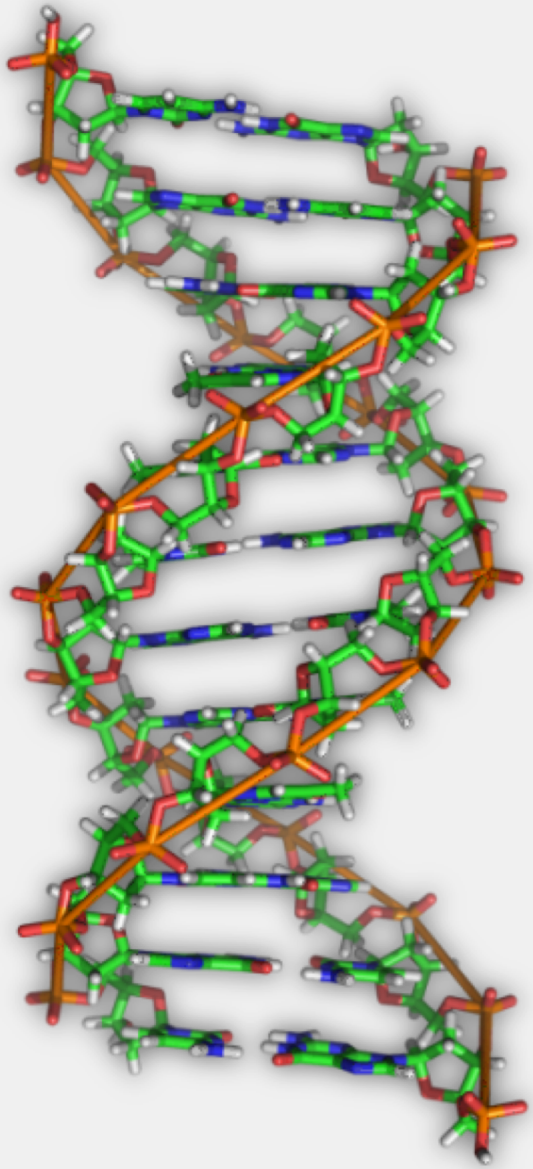


Software

NAFlex is powered by an Apache 2. web server with PHP 5. and MySQL 5.0.51. Calculations are redirected to a 8 core Intel(R) Xeon(R) CPU @ 2.67GHz - 16GB RAM cluster managed by a Sun Grid Engine batch manager.

List of external software used in **NAFlex** operations:

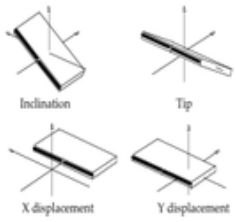
Program	Description	Package & Version
Curves+, Canal	Analysis of Nucleic Acids Structure and Flexibility	Curves+ 2.0
Open Babel	Chemical toolbox designed to work with many languages of chemical data.	Open Babel 2.1.1
R Statistics	Statistical Computing and Graphics software.	R 2.15.0
BLAST	Basic Local Alignment Search Tool	BLAST 2.2.17
CMIP	Classical Molecular Interaction Potential	CMIP 2.5.4
Gnuplot	Plotting tool	Gnuplot 4.2 patchlevel 2
Grace	Plotting tool	Grace 5.1.21
GROMACS	Molecular Dynamics Simulator	GROMACS 4.5.5
JMol	Molecular Graphics Viewer	JMol 10.00.46
MobyLite PerlAPI	BioMoby Perl API	MobyLite PerlAPI 1.0
NAMD	Molecular Dynamics Simulator	NAMD 2.8
PCAsuite	Trajectory compression tool	PCASuite 1.1
PropKa	Prediction of protein pKa values	Propka 2.0
Ptraj, cpptraj	Structure and dynamic analysis of trajectories	Ambertools 14
tgatoppm, pnmcrop, pnmtopng	Image management	Netpbm 10.0
Tleap	MD preparation program	Ambertools 14
VMD	Molecular Graphics Viewer	VMD 1.8.5



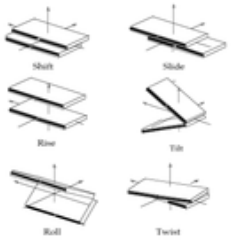
BIGNASim

A BigData approach to efficiently manage large nucleic acids simulation data

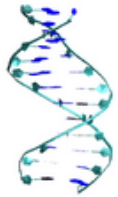
Trajectory Analyses >> *(Click to expand/shrink)*



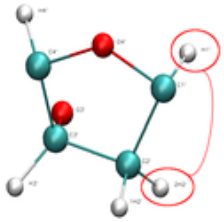
Curves Analysis



Stiffness Analysis

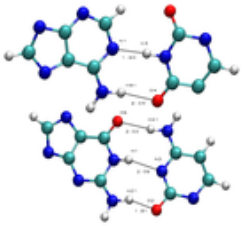


PCAzip Analysis

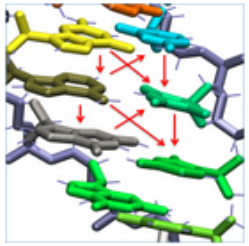


NMR_JC Analysis

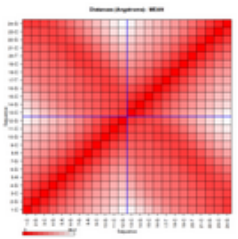
NMR NOEs Analysis



HBs Analysis



Stacking Analysis



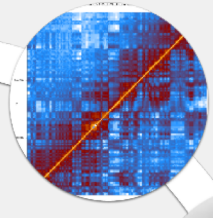
Contacts Analysis

Multiscale complex Genomics

Multiscale Complex Genomics

WP7
Pilot Projects
VRE lead users. Demonstration of MuG utility in delivering real biological added value for the community.

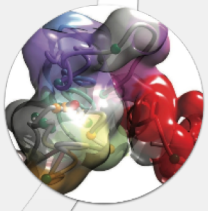
WP6
Analysis tools
User-driven, ready-to-use workflows combining existing software tools for simulation and structural analysis of experimental data.



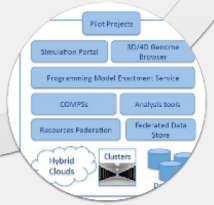
WP4
Data management
Scalable, high-performance data storage solution.
Advanced tools for data mining and exploration.

Virtual Research Environment
3D/4D Genomics

WP3
Multi-scale genome browser
Allowing for the first time to connect 1D to 3D/4D genomic data and to navigate across all the resolutions of DNA in the cell through a graphical user interface.



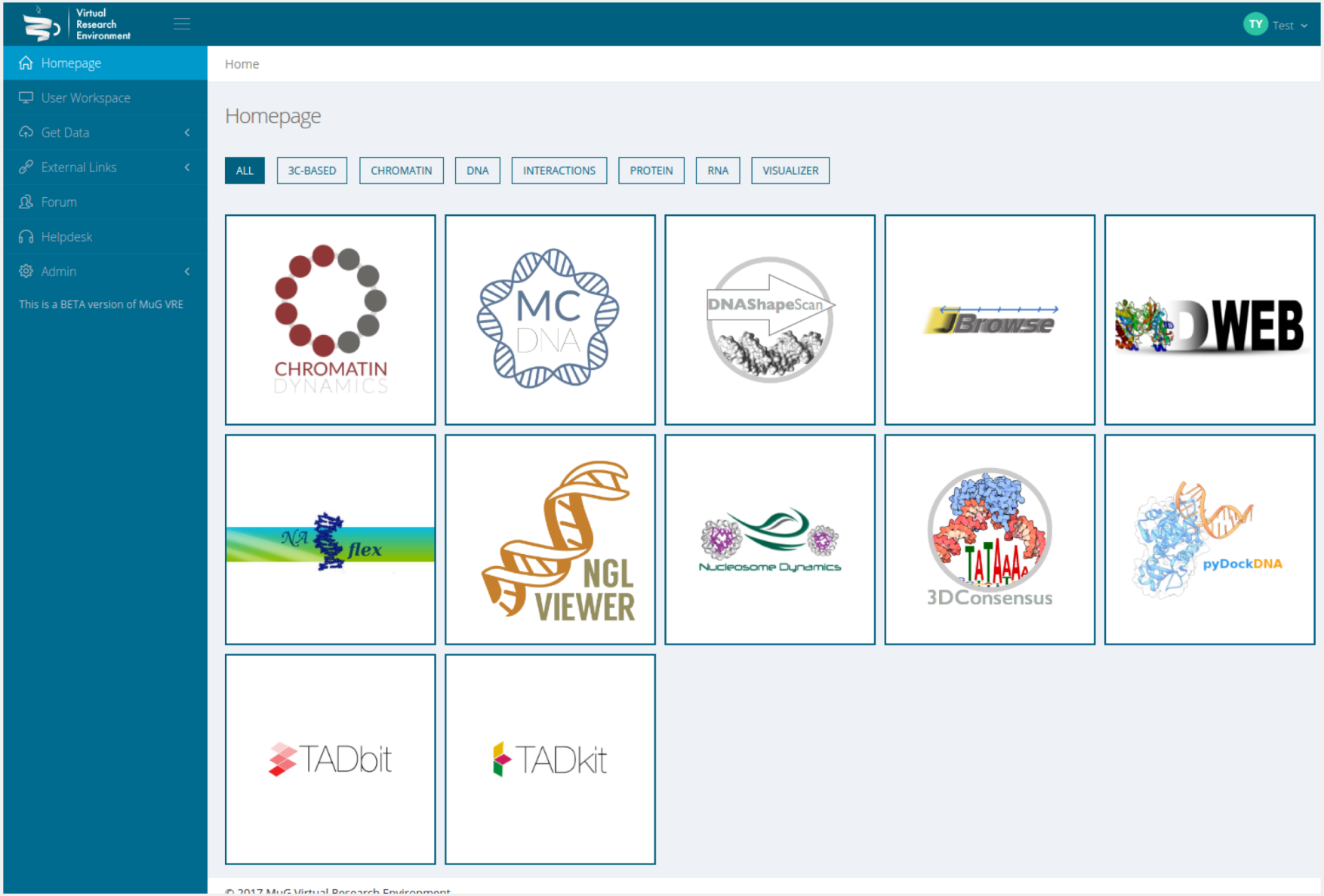
WP5
Computational infrastructure supporting the VRE services
Compatible with Cloud and HPC.
Interoperable with Data Infrastructure.
Integrates security and User Support.



WP2
Outreach, training, exploitation
Community engagement, dissemination of results, training of VRE users, long-term sustainability and exploitation.

IRB BARCELONA INSTITUTE FOR RESEARCH IN BIOMEDICINE
BSC
cnag
CRG Centre for Genomic Regulation
cnrs
The University of Nottingham
UNITED KINGDOM · CHINA · MALAYSIA
EMBL

http://www.multiscalegenomics.eu/
MuG community – interest group


Virtual Research Environment
TY Test
Homepage
User Workspace
Get Data
External Links
Forum
Helpdesk
Admin
This is a BETA version of MuG VRE
Home
Homepage
ALL
3C-BASED
CHROMATIN
DNA
INTERACTIONS
PROTEIN
RNA
VISUALIZER
CHROMATIN DYNAMICS
MC DNA
DNAShapeScan
JBrowse
DWEB
NA flex
NGL VIEWER
Nucleosome Dynamics
3DConsensus
pyDockDNA
TADbit
TADkit

Acknowledgements

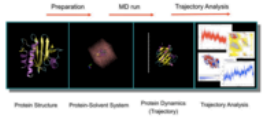
Prof. Modesto Orozco
Prof. Josep Lluís Gelpí
Adam Hospital

Bioexcel forum (ask.bioexcel.eu)

ask.bioexcel.eu

Sign Up | Log In

Molecular Dynamics Novice Users IG



This category is for discussions amongst members of the Molecular Dynamics Novice Users Interest Group that do not relate directly to GROMACS, HADDOCK or CPMD. If your question relates to one of these codes, please use the code-specific forums on this site.

Follow up on Webinar: "NAFlex, a web server for the study of nucleic acid flexibility" by Federica Battistini (2017-07-20) 1d

CMIP titration service failure Mar 15

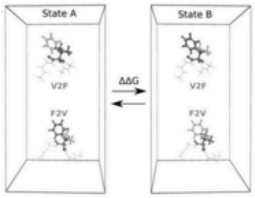
PATC Simulation Environment for Life Sciences 2017 Mar 9

BioExcel Webinar #3: Atomistic Molecular Dynamics Setup with MDWeb Oct '16

1 / week

1 / month

Free Energy IG



This category is for all discussions related to the Free Energy Interest Group. Its primary audience is members of the IG but this forum is open for anyone to contribute.

Follow-Up from Webinar: Mutation free energy calculations with pmx Jun '16

Workflows IG

This category is for all discussions related to the Workflows Interest Group. Its primary audience is members of the IG but this forum is open for anyone to contribute.

Follow up, Webinar "Building pharmacological workflow blocks for virtual screening" 2d

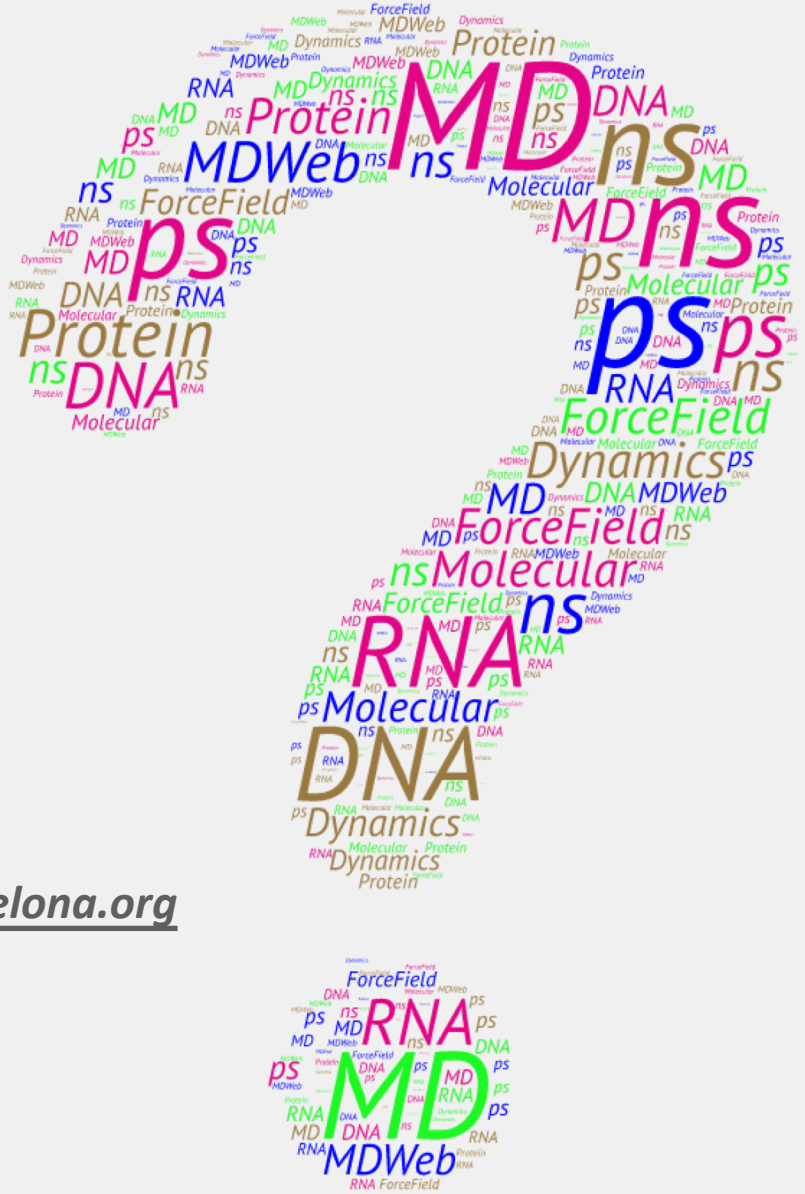
Follow up, Webinar "Introduction to the Common Workflow Language" 16d

Follow up on "Large-scale analytical workflows on the cloud using Galaxy and Globus" Nov '16

Musings on GROMACS and workflow description integration Nov '16

1 / week

2 / month



federica.battistini@irbbarcelona.org

Ask.bioexcel.eu