Molecular Dynamics on Web

Manual, v1.0

Table of Contents

1.- Introduction

MDWeb is a web interface to perform molecular dynamics simulations or analyze molecular dynamics trajectories.

Motivation

Since first simulation of biomacromolecules in 1977, molecular dynamics (MD) has experienced a long evolution, and, at present, it is a mature technique that can be used to obtain an accurate picture of the dynamics of complex systems with atomistic detail.

Despite all its power the practical use of MD is limited by three factors:

i) the uncertainties of classical force-fields.

ii) the required computational power.

iii) the high level of expertise needed to use the technique.

Focusing in the last point, it should be remarked that setting up a system for simulation requires from the combination of a series of operations, many of them to be performed manually, and a number of decisions that demand a significant degree of expertise. The end result is that newcomers to the field face a stiff learning curve and that in some cases (for example when supercomputer resources are used) setting-up of simulations can represent a period of time similar to that of deriving the trajectory. These problems reach a maximum in high-throughput projects, where thousands of trajectories need to be launched, what forces the development of automatic tools for mimicking human expertise in launching (and then analyzing) MD trajectories.

Within the MoDEL project we have developed a series of tools that allowed us to launch and analyze around 2,000 trajectories with little human intervention. An evolution of our original workflows is made available here as a general tool, not linked to any specific simulation package, to help naïve users to prepare completely systems for simulations, and that allows expert users the use of MD in the high-throughput regime.

The software tool developed here is presented as a fully integrated web-services-oriented software platform to perform molecular dynamics (MDMoby), and its web portal MDWeb. The technology has been adapted to be accessed as web-services following the BioMoby framework. MDMoby can be accessed with web-services clients and also programatically through suitable APIs (INB-BSC). A number of pre-prepared pipelines are also available to help non-expert users in the use of standard simulation procedures.

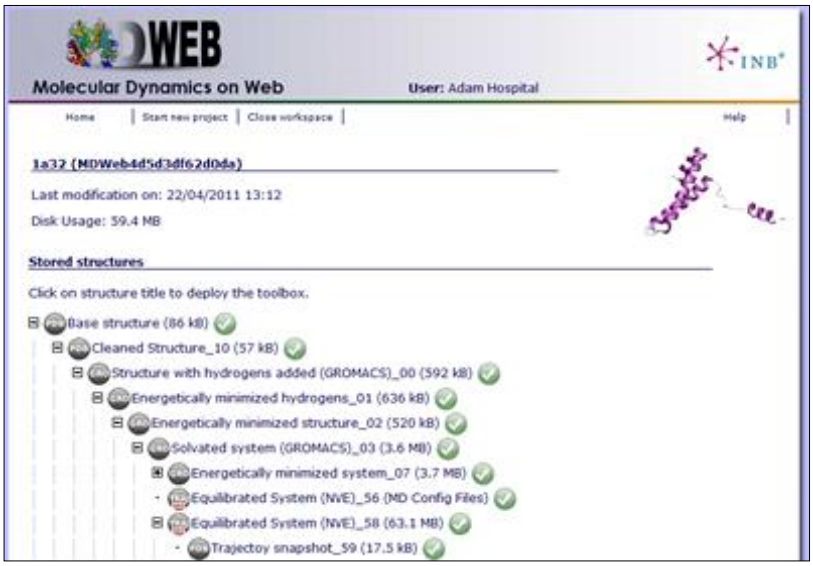
The web portal **MDWeb** provides a friendly environment to setup new systems, run test simulations and perform analysis within a guided interface. Setup files can be prepared for Amber, NAMD, and Gromacs (other formats will be incorporated in the future) and analysis can be carried out using any standard trajectory format.

Additionally, the platform is interfaced to our flexibility analysis software FlexServ, so providing coarse-grained simulation, and advanced flexibility analysis tools.

2.- Getting Started

MDWeb interface

MDWeb holds a personal **workspace** where structures and trajectories are stored.



Data is structured in **"projects"**. A **project** is opened with an initial structure or trajectory. All operations are based on that initial input.

Data in the project is organized in a **tree** that outlines the history of every object back to the initial one.

Structures and **trajectories** are added to the **tree** once available.

The available **operations** and **workflows** are defined according to every data type.

User Registration

MDWeb gives the possibility to work as an anonymous user, or as a registered user.

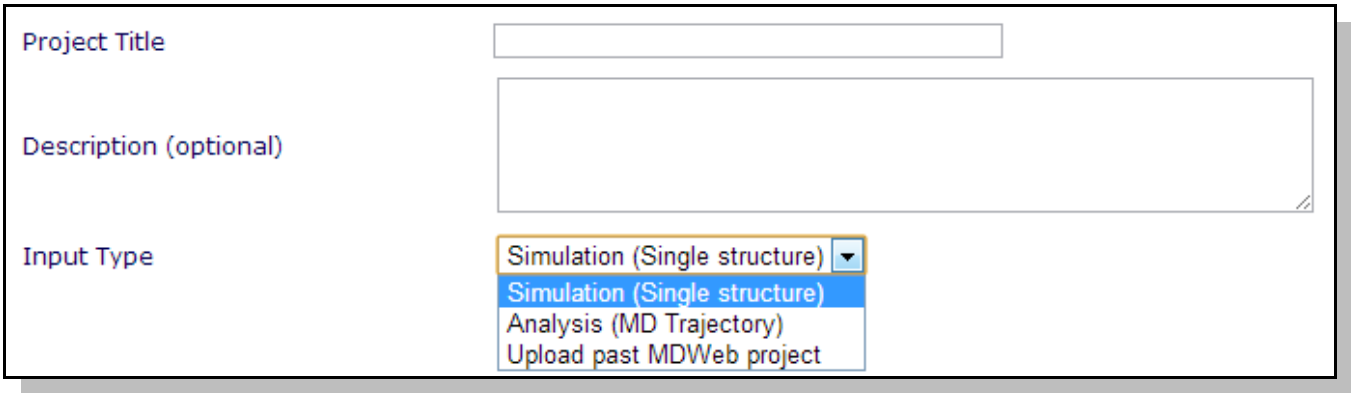
Warning: Anonymous users' projects **will be removed** once disconnected or when session expires (after some minutes of inactivity).

On the other hand, **Registered** users' projects **will be stored** in our disks for a reasonable time. A capacity of **2 GB** of disk space will be assigned for each registered user.

New Project

MDWeb offers three main entry options:

- **Simulation (Single structure)**: Work from a structure, setup and run an MD simulation, run a coarse-grained MD.
- **Analysis (MD Trajectory)**: Work from a trajectory, analyse, get information or convert between MD trajectory formats.
- **Upload past MDWeb Project**: Upload a stored MDWeb project.



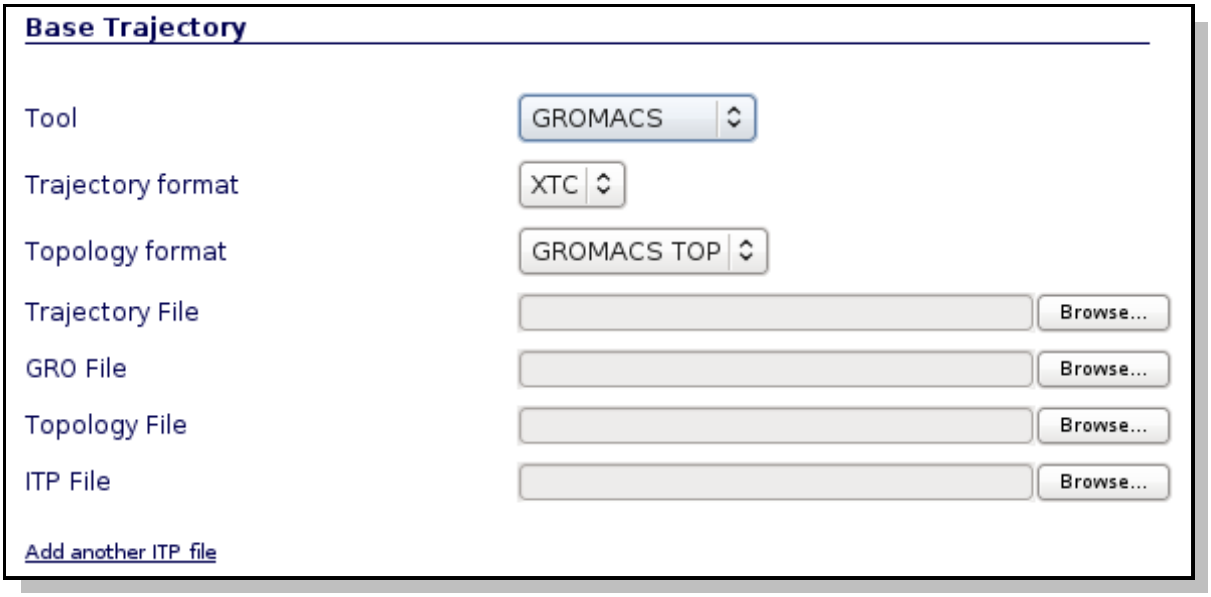
If **Base Structure** option is chosen, user will be redirected to the Structure Checking page.

(*) Please note that user provided structures must follow PDB File Format v.3.30 (July 13, 2011). The most relevant points to consider are:

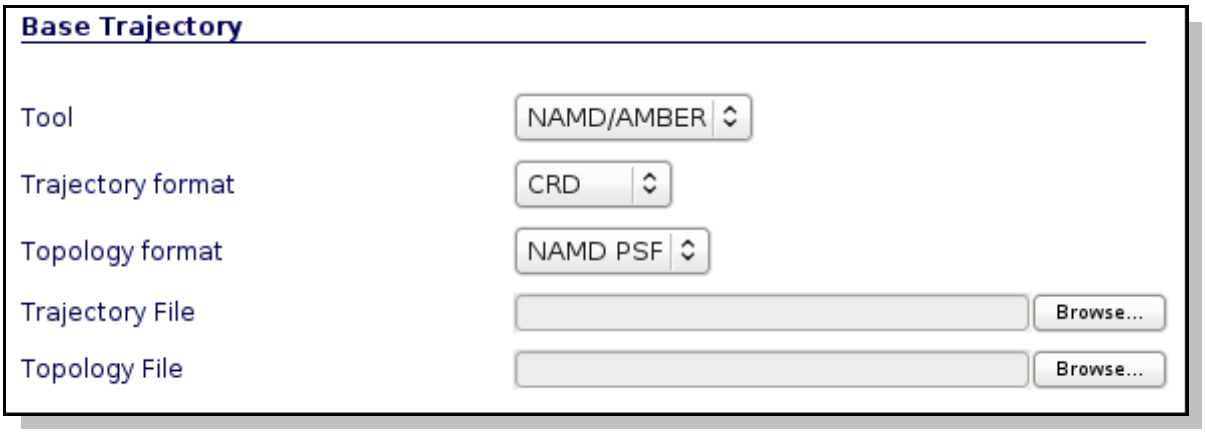
- A Non-blank alphanumerical character is used for **chain** identifier.
- Non-polymer or other "non-standard" chemical coordinates, such as water molecules or atoms presented in HET groups, use the **HETATM** record type.

If **Base Trajectory** option is chosen, a new project is created with the input trajectory as a tree root (base object to work with). (See Workspace). *Note that MDWeb has a limit file size of **100 MB**.*

Different input files are required depending on program and trajectory formats (NAMD/Amber/Gromacs). **Gzipped** files are accepted.

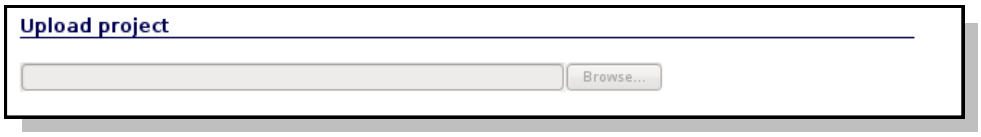


Gromacs Trajectory Input Files Example



NAMD/Amber Trajectory Input Files Example

The **Upload Project** option allows **MDWeb** users to upload projects, previously strored using the Download Project utility.



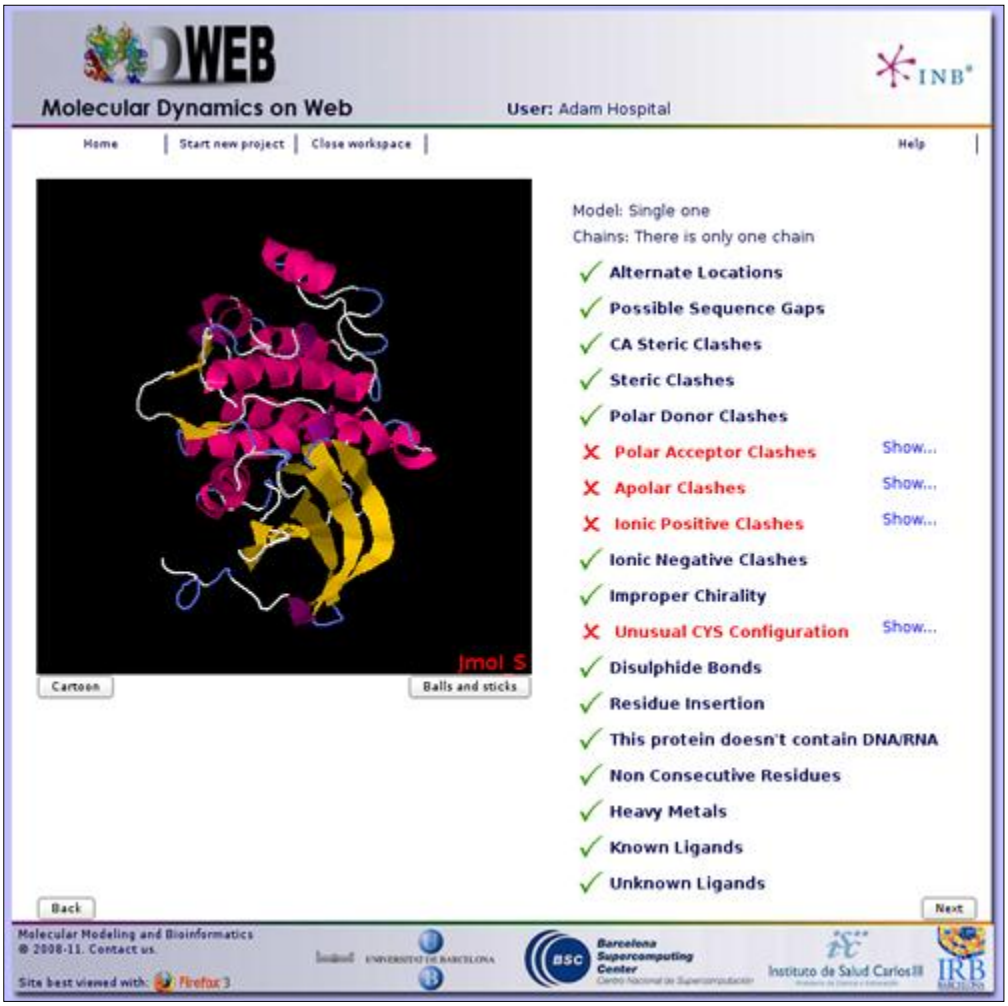
Upload Project

3.- Structure Checking

In a Molecular Dynamics simulation, the correctness of input structures is crucial. Small errors in the input structure may cause MD simulations to became unstable or give unrealitstic trajectories.

The purpose of the initial **Structure checking** page is to check for the most common problems in the input of MD simualtions, allowing the user to select possible solutions when available. Besides, structure checking allows to select fragments of the system to be simulated in the case that alternate models or multiple subunits are present in the incoming structure.

An interactive JMol applet window provides additional help in assessing the significance of errors found. Please note that MD itself can correct some of the problems found (specially steric clashes), but other can be only be corrected by editing the structure beforehand.



The **checking** consists in a list of possible options to choose the system to be simulated:

- Structure Model.
- Structure Chain/s.
- Residue/Atom Alternate Locations.

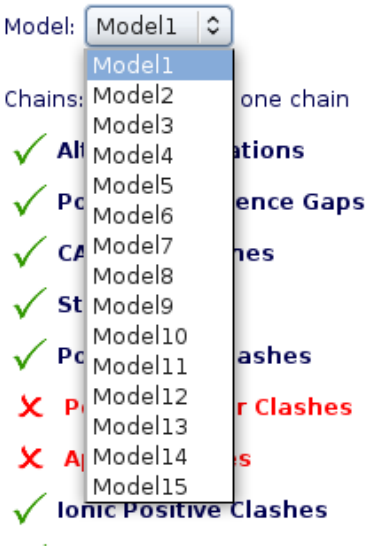
And a list of possible problems:

- Sequence Gaps/Non Consecutive Residues.
- Atom Clashes: Steric, Alpha-Carbon, Polar Donor, Polar Acceptor, Apolar, and Ionic Positive/Negative.
- Thr Improper Chirality.
- Unusual peptide-bond cis configuration.
- Disulphide Bonds.
- Residue Insertions.
- Structure containing DNA/RNA.
- Metals.
- Ligands.

Structure checking codes are:

-  **Ok**: no problems found.
-  **Warning**: important information or options to choose.
-  **Error**: will probably cause MD failure.

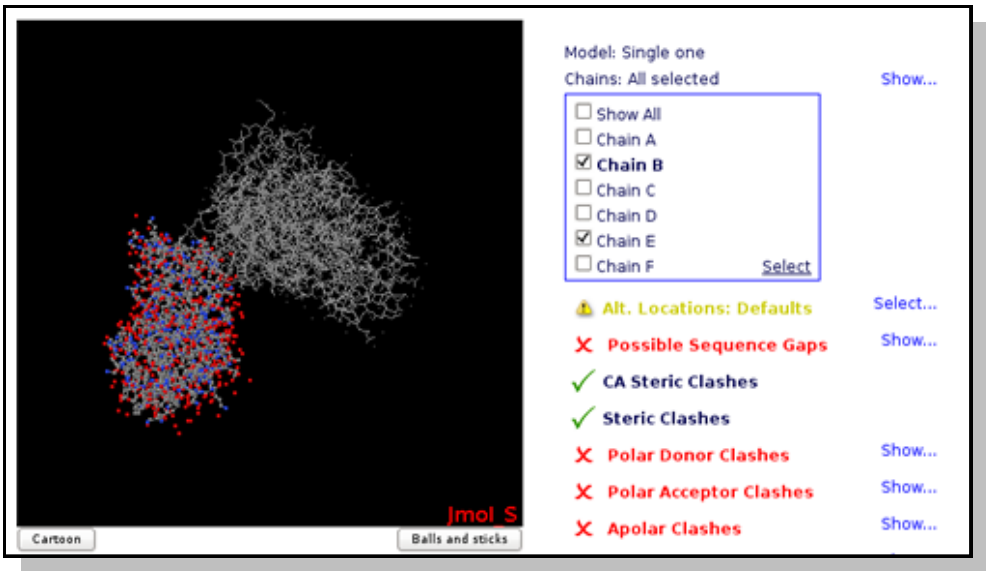
Choosing Structure Model



For PDB files containing multiple models (common for instance in NMR obtained structures), only one of the models can be simulated. The interface shows you the list of possible structure models to choose. When chosen, all checking parameters are automatically recomputed with the selected model.

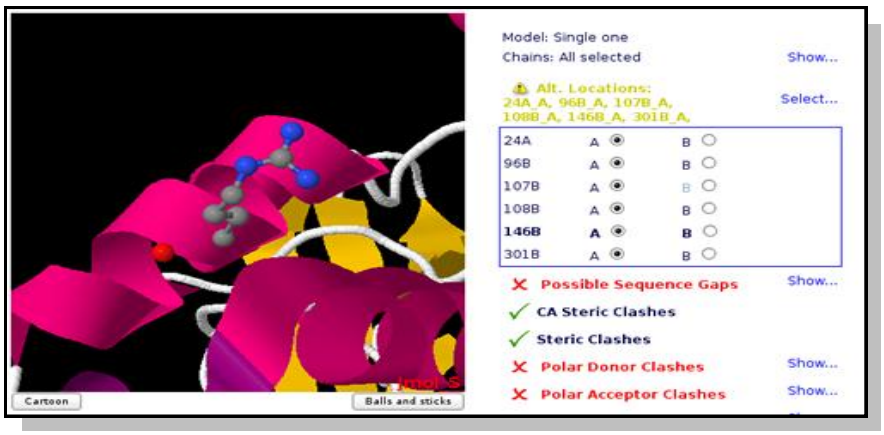
Choosing Structure Chain/s

For structures containing several Chains, user can select the ones to be included in the simulation. A Jmol applet helps in the identification of the chains, highlighting them when passing over the name with the mouse pointer. Be careful that selecting non complete structures may lead to unrealistic MD trajectories.



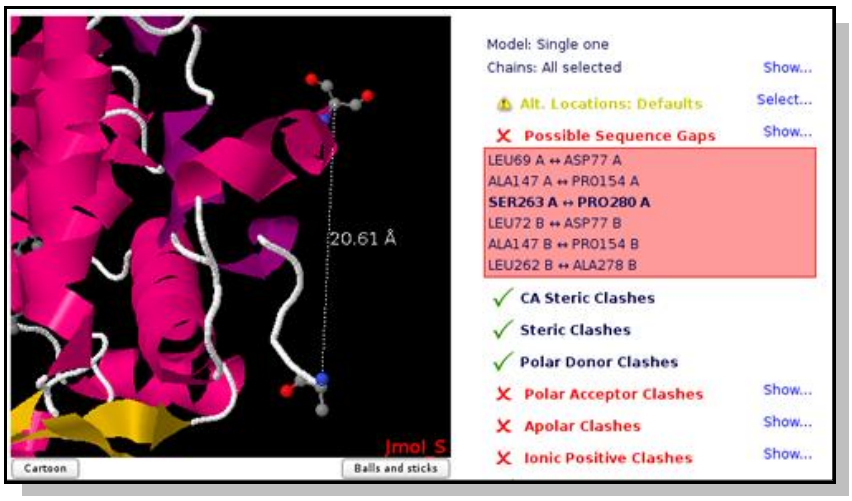
Choosing Atom/Residue Alternate Locations

Alternate Location indicators are used for atoms where more than one position is detected. Within a residue, all atoms that are associated with each other in a given conformation are assigned the same alternate position indicator. User can choose the residue alternate location of interest helped by a Jmol visualizer applet.



Sequence Gaps/Non Consecutive Residues

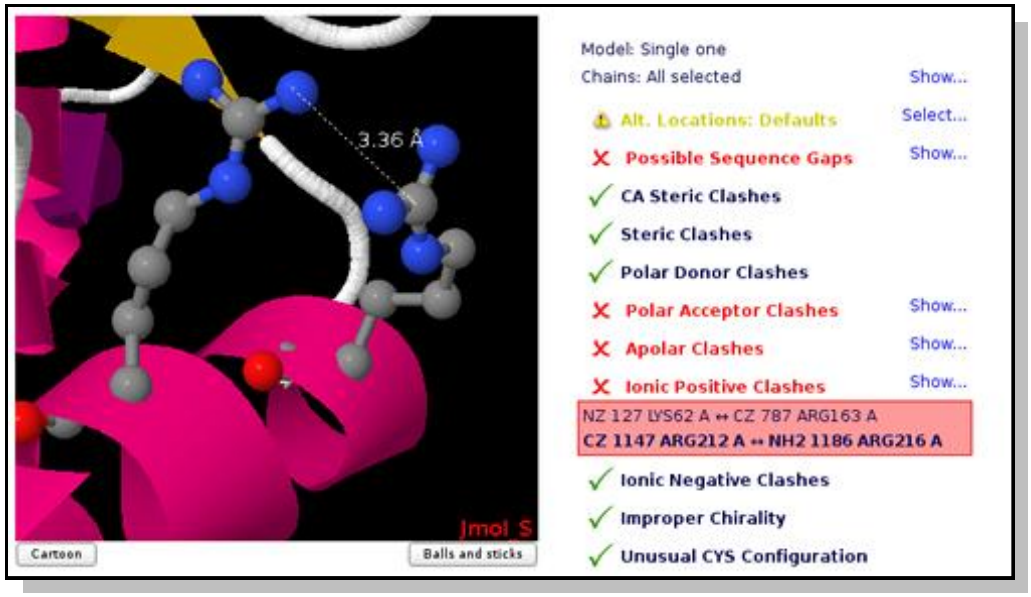
Flexible protein regions can have too low electron density to be detected in X-Ray diffraction experiments, and can be missing from PDB structures. This situation gives unrealistic structures where protein is split in several unconnected chains with new N- and C- terminal residues and buried protein regions become exposed to solvent. Simulation of such structures will give untrusty results and should be corrected by filling the missing gaps. MDWeb detects gaps both from the residue numbering and the existence or non-realistic bond distances.



Atom Clashes

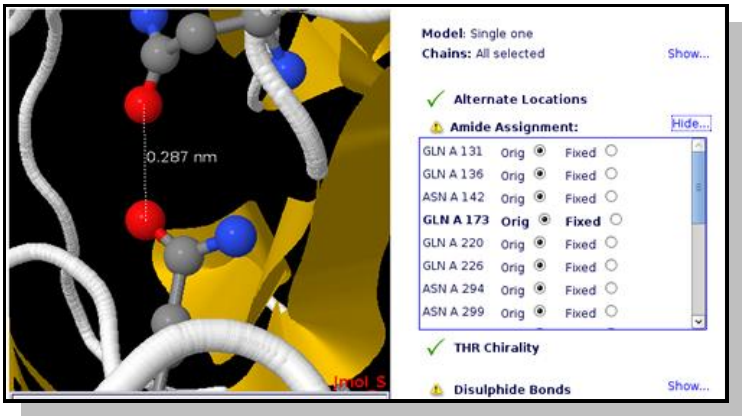
Atoms that are too close in space can have a problem of energetic repulsion. MDweb provides the list of atom pairs and the corresponding distances that can have potential problems. Most of clashes come from over-compactation of crystal structures and are naturally corrected on system setup or MD equilibration, but may lead to a significant distortion of the structure. Clashes are classified in different groups, depending on the atom types involved:

Group	Description	Distance cut-off
Steric Clashes	Any atom pair	1 Å
CA Steric Clashes	Alpha-carbon atom pairs	3.8 ± 1Å
Polar Donor/Acceptor Clashes	Polar hydrogen bond donor/acceptor atom pairs.	3.1 Å
Note that common polar clashes come from mis-assignment of side chain atoms in Asn, Gln, or Thr residues.		
Apolar Clashes	Apolar atoms	2.9 Å
Note that possible apolar clashes can come from atoms neighbouring legitimate hydrogen bonds.		
Ionic Positive/Negative Clashes	Positively/negatively charged atoms	3.5 Å



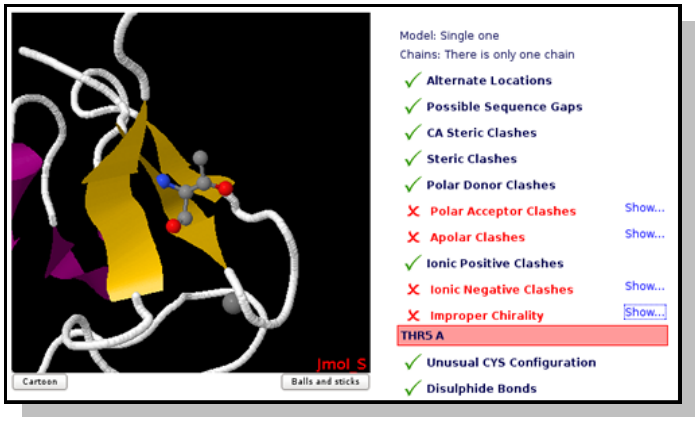
Improper Amide Assignment

Amide Groups in the side chains of Glutamine and Asparagine (Gln, Asn) residues can act simultaneously as hydrogen bond donors and acceptors. The electron density near the nitrogen and oxygen atoms of these functional groups is compatible with two rotamers related by a 2-fold symmetry axis. Therefore, electron density maps obtained from X-ray diffraction experiments can be wrongly interpreted leading to improper amide assignments. User can choose the best fitted amide group configuration helped by Jmol applet.



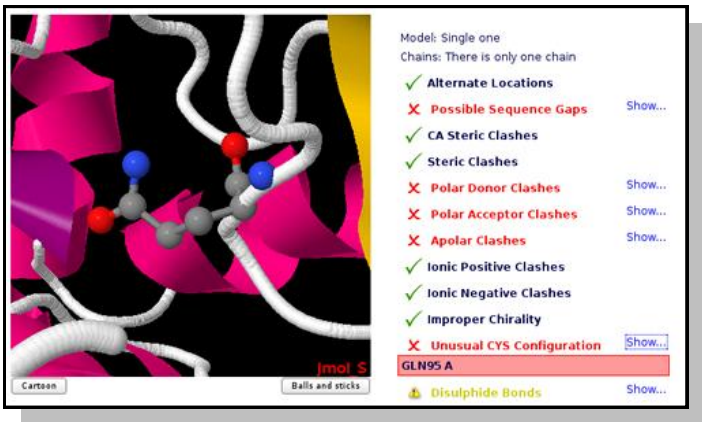
Improper Chirality

Chiral molecules lack an internal plane of symmetry and thus cannot be superimposed on their mirror images. Organic molecules which contain at least one tetrahedral carbon atom bound to four different substituent groups are chiral. That carbon atom is a "chiral centre". Alpha carbons of all amino acids except glycine are chiral, but in natural proteins only one of the configuration is found (L-configuration). D-amino acids are not naturally found in proteins. Threonine and Isoleucine side chains contain an additional chiral centre at the beta carbon. Mis-assignment of atom names in the structure may lead to an improper chirality. Usual restrictions in molecular dynamics force-fields do not allow to reverse chiral centers. MDWeb checking page will help you identifying these improper chiralities.



Unusual cis/trans Configuration

Cis/trans isomerism describes the orientation of functional groups within a molecule where a bond has a limited possibility of rotation. Peptide bonds have a considerably double bond character and present *cis/trans* isomerism. Nearly all peptide bonds appear in the *trans*, whereas *cis* configuration is sometimes found in Proline residues. Omega torsion angle (ω) is computed to identify unusual *cis/trans* configurations in peptide bonds.

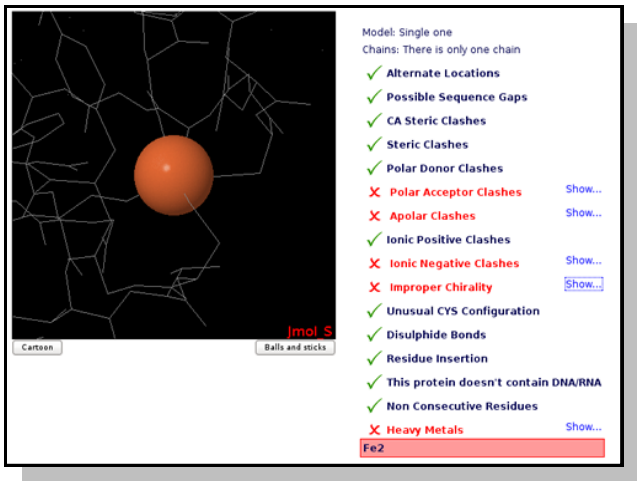


Disulphide Bonds

Disulphide bonds are covalent bonds usually formed by a pair of thiol groups. They are also called SS-bonds or disulphide bridges. As they are covalent bonds, it is of great importance in a simulation to take them into account. In fact, SS-bonds are known to have an important structural role in protein structure and stability. Setup procedures can detect and link thiol groups forming disulphide bridges, but it is important to know whether the structure contains such covalent bonds. Jmol applet will show the structure's disulphide bonds, as well as the distance between the sulphur atoms involved.

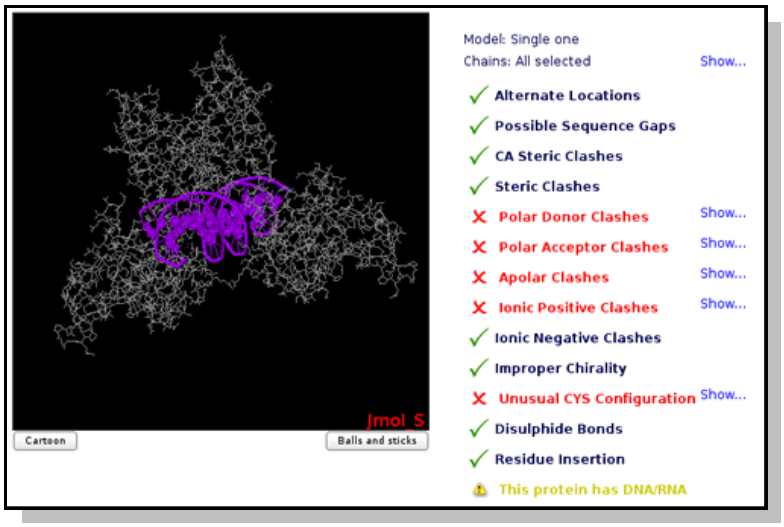
Metals

Metals (Mg, Zn, Mn, Mo, Ni, Fe, Co, Cu, Hg, Cd, Ag, Au) are usually found making coordination complexes with protein residues. A typical example is the so called *Zinc finger*, where a Zn ion is coordinated by cysteines and histidine residues, forming a structural motifs that help stabilize the structure. Metall complexes involve a complex chemistry and are not normallly covered by standard force-fields as they require a complete re-parametrization of both metal and ligands. Simulations of such complexes would normally require the setting of distance restrains to maintain the coordination structure.



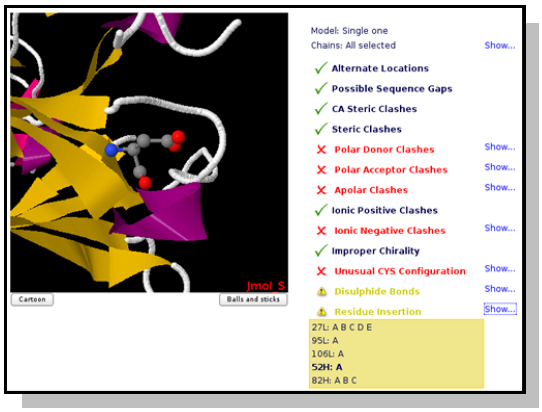
Structure containing DNA/RNA

MDWeb can work not only with proteins but also with nucleic acids and even with protein-DNA and protein-RNA complexes, however this feature has not been extensively tested. The structure checking process informs about the existence of nucleic acids in the input structure. Removal of nucleic acids if desired can be done in the chain selector.



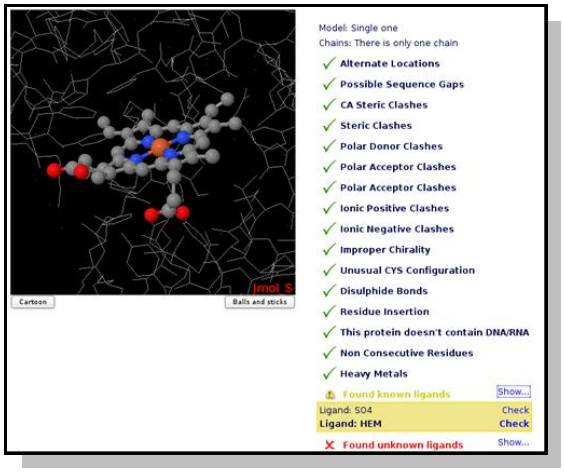
Residue Insertion

PDB specification contains an optional field, named *Residue insertion code* to allow structure providers to match residue numbering between proteins from different sources when insertions and deletions occur. Simulation setup will remove such field and renumber residues accordingly. The MDWeb checking page shows the set of residues with insertion codes in the Jmol applet.



Ligands

The majority of the structures in the Protein Data Bank have a non-standard residues included, identified as "heteroatoms" in the structure. Molecular dynamics force-fields contain parameters for standard amino acids and nucleotides but not normally for such compounds. To include a ligand in the simulation, the complete description of the ligand structure and the corresponding force-field parameters should be provided. MDWeb contains a extense library of already parameterized ligands that can be included in the simulation ("known ligands"). In the checking step, the presence of ligands in the structure, as well as the availability of the corresponding parameters library will be tested. In the case of unknown ligands, not available in the library, used will be prompted for the corresponding parameter data. Please note that the default behaviour of MDWeb checking phase is to remove ligands from the structure.

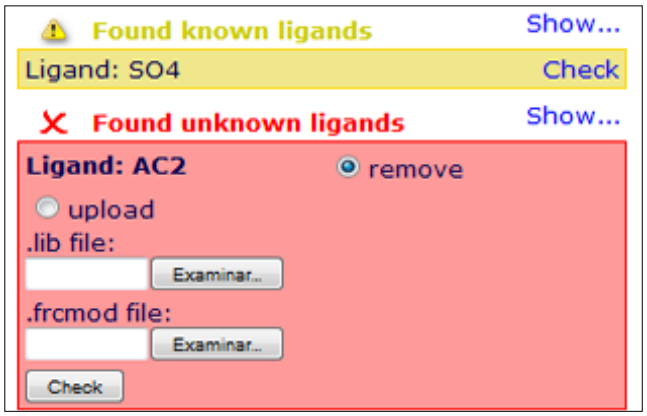


4.- Ligand Checking

The majority of the structures in the Protein Data Bank have a non-standard residues (known as heteroatoms or ligands). Molecular dynamics force-fields are designed to work with amino acids or nucleotides. To include a ligand in the simulation, a complete description of the molecule as a parameterization file must be provided.

Ligands are divided in two groups: Known and Unknown, depending on the ligand parameters availability.

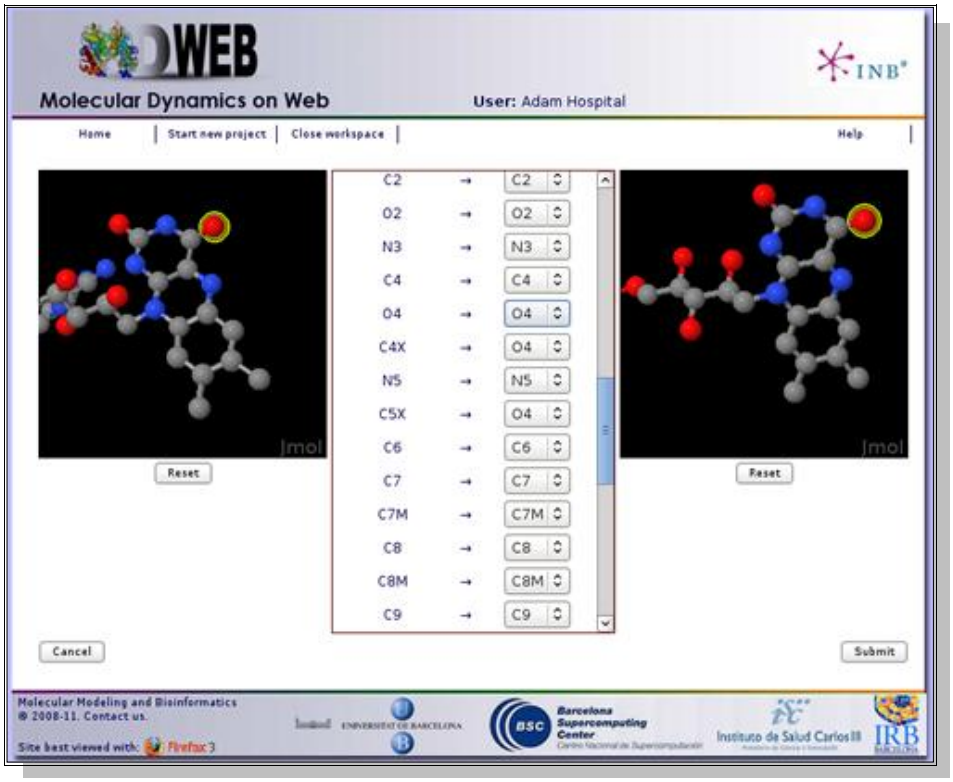
Known ligands correspond to a database of heteroatom used on our MoDEL library, and can be incorporated to the simulated system. For the unknown ones, **MDWeb** gives the user the possibility to upload parameter files (in Amber format, lib/frcmod).



If parameter files are not uploaded, the ligand will be removed from the structure. Please note that to avoid inconsistent systems, the default option is to remove ligands, even the known ones.

Atom name match window

Finally, in both groups (known or unknown), atom naming of the ligand needs to be checked. Atom names in non-standard residues found in the PDB structure can differ to those in the available parameter files. To generate a consistent structure, atom names should match. To make this process easier, **MDWeb** opens a new window with two synchronized JMol applets, showing the incoming (PDB) ligand structure (left window) and the library ligand (right window).



If the two ligands are siimilar to each other, both structures will be showed with the same orientation, and rotation of the structures will be synchronized.

In the middle of JMol windows, there is a matching table. The first column shows the ligand atom names as appear in the incoming PDB file, while the second one shows the atom names contained in the ligand parameter file. For atoms having different names, user can use this central form to do the necessary corrections. Atoms located under cursor are highlighted in the corresponding JMol applet, thus allowing the correct identification.

This process can only be skipped when the ligand is mono-atomic, as in the case of ions (Auto link).

5.- Workspace

Users in **MDWeb** have their own **workspace**.

User **workspace** appears just after login. In the **workspace**, all projects user's projects are shown. Name of the project, last modification date and disk usage is displayed.

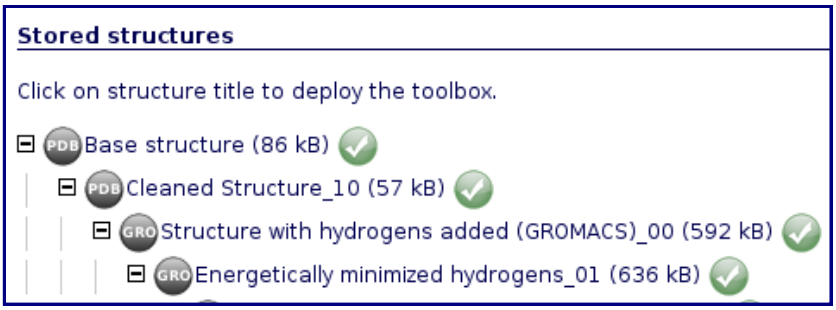
This page acts as a project repository. Users can identify a specific project to open, remove it from the workspace (note that this implies erasing all data included in the project) or download it.

Projects downloaded will be given in a single compressed file, that can be stored by **MDWeb** users and can be uploaded again from **MDWeb** start project page at a later time. This will help to save disk space and in the case of non-registered users to resume their projects in further sessions.

Structures

Once into a project, **MDWeb** interface has a hierarchical organization, starting from the base structure/trajectory (root of the tree) and going down applying operations and building new branches.

From the initial base object, a set of operations can be applied. These available operations are selected automatically depending on the nature of the structure. This helps the user to follow the right sequence of operations. The type of structure is shown as an icon at the left part of the object name (See Icons).



Operations

Clicking on one of the project branches, a toolbox with the available operations appears. Operations can be divided in Molecular Dynamics tools (setup, simulation/optimization, analysis), and visualization tools (rasmol, Jmol).

Other operations include downloading results in a compressed tgz file, view logs of the operation, or delete the entire branch.

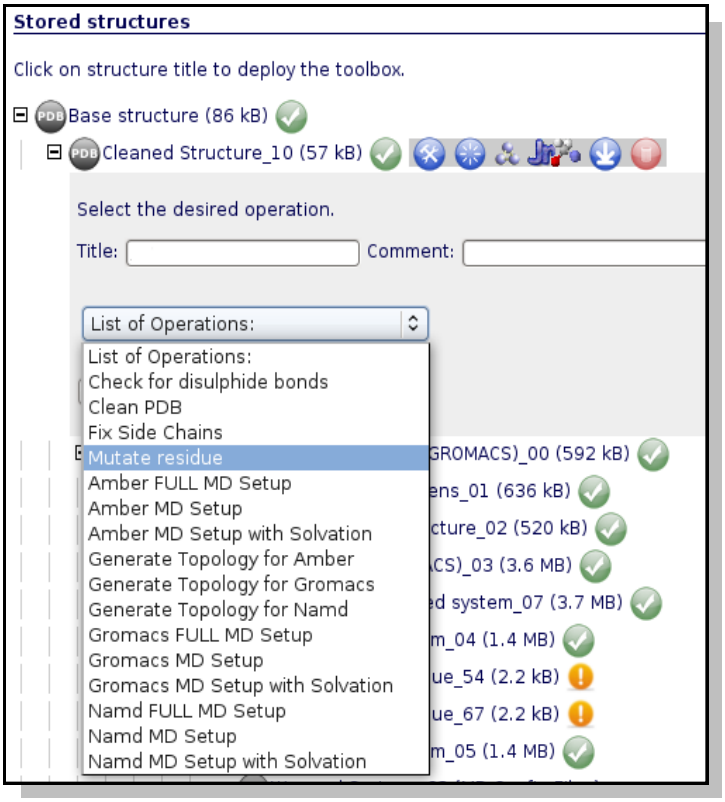
Toolbox icons::

Operations

- Perform a new setup operation on the selected structure.
- Perform a new simulation/optimization.
- Perform a new analysis on the selected trajectory.
- Visualize structure using Rasmol compatible viewers (plug-in required).
- Visualize structure using JMol.
- Show log file
- Download results in a compressed tgz file.
- Delete item from the workspace.

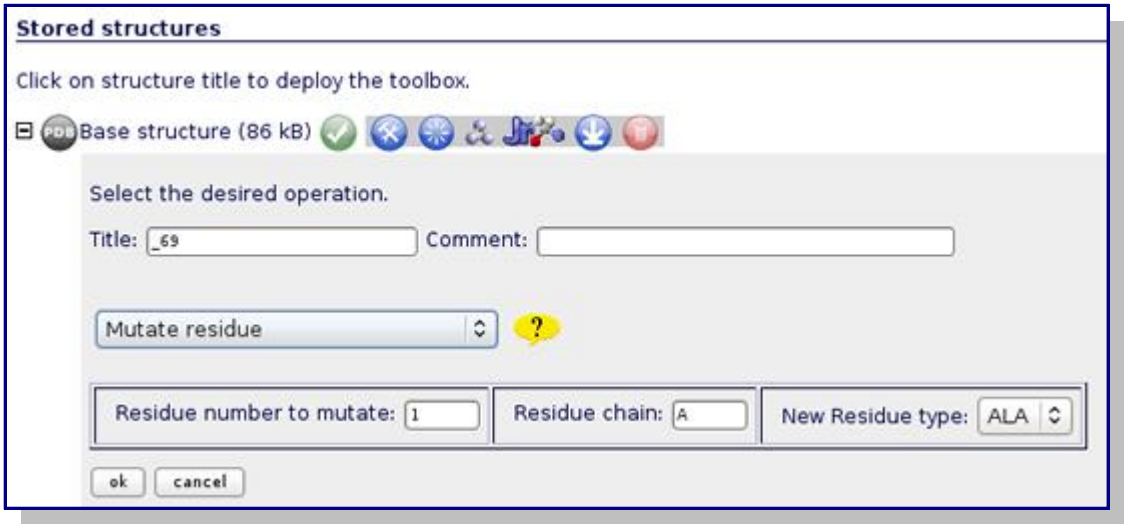
Services List

When choosing either setup, simulation/optimization or analysis operations, a selector with all the services that can use the specific structure type as an input will appear. Some of these services are simple processes (MDMoby webservices), whereas others are workflows involving preorganized sets of operations.



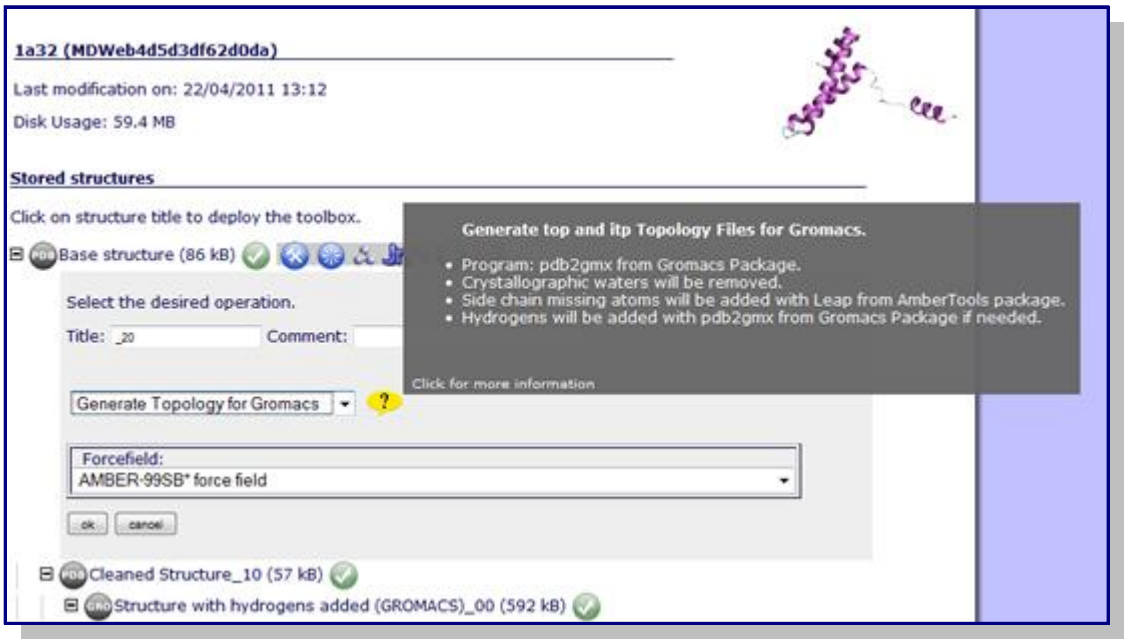
Services Parameters

For services having input parameters, a specific form will appear, including the recommended values.



Service Parameters Example

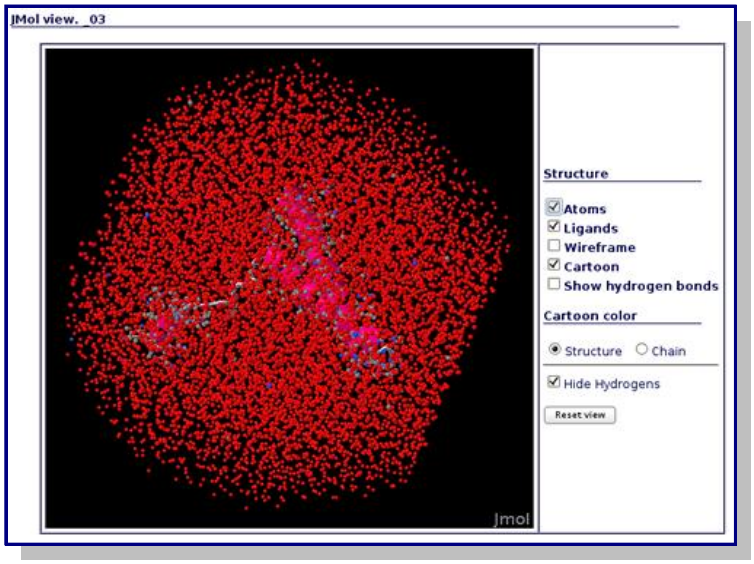
When selecting a service, a question mark icon appears next to the selection list. Passing over this icon with the mouse pointer triggers a small window with the most important information about the service and parameters. For workflows, the set of steps is described. Clicking at the question mark icon a complete service description is shown.



Quick Help Information Example

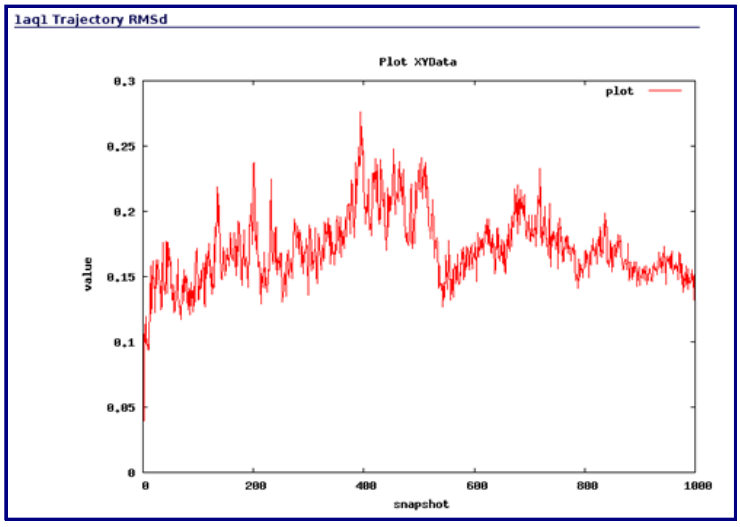
Visualization

Data can be visualized at any point during the process of setup or analysis. For three-dimensional structures or trajectories, Rasmol scripts and JMol applet is offered with just a click. In case of JMol applet, a new window will be opened with the visualization, together with some pre-configured JMol options such as structure representation (Atoms, Ligands, Wireframe, Cartoon, HBonds) and Color (Structure, Chain).



JMol Visualization Example

In the analysis part, 2D plots (RMSd, Bfactors x Residue, etc.) can also be automatically visualized with **MDWeb**.



Analysis Plot Example

6.- MDWeb Operations

Structure modification:

Mutate Residue

Mutate a Residue from a structure given in PDB format

Program: **VMD**.

Set-up for Simulation

Check for disulphide bonds

Checking for Disulphide Bonds with a distance criteria.

- Distance Cutoff: 2.5 Å
- SS bonds added to the PDB data as a REMARK.

Clean PDB

Remove Crystal Waters.

- Remove Crystal Hydrogens.
- Remove Non-Parametrized Ligands.

Fix Side Chains

Fix Side Chain Problems, adding missing heavy atoms (if possible).

Program: **Leap** (Ambertools).

Add structural water molecules and ions

Add structural water molecules and ions in the energetically most favourable positions at the surface of the structure.

Program: **CMIP**, Classical Molecular Interaction Potentials.

- By default, necessary ions (Positive ions: Na+. Negative Ions: Cl-) are added in order to neutralize the system.

Set ionization state

Set the appropriate ionization states according to pH

Programs: **ProtPka** (Fast). **CMIP** (Slow)

Energy Minimization (Hydrogen Atoms)

Energy minimization of hydrogen atoms keeping all heavy atoms fixed to their original positions

NAMD

Usable for both AMBER and CHARMM topologies

Program: **namd2** from NAMD Package.

- Available types of minimization: Conjugate Gradients.

GROMACS

Programs: **grompp** and **mdrun** from Gromacs Package.

- Available types of minimization: Conjugate Gradients and Steepest Descent.

Energy Minimization (All atoms)

Possibility of restraining atom movements with a given force constant.

NAMD

Usable for both AMBER and CHARMM topologies

Programs: **namd2** from NAMD Package.

- Type of minimization: Conjugate Gradients.

GROMACS

Programs: **grompp** and **mdrun** from Gromacs Package.

- Types of minimization: Conjugate Gradient and Steepest Descent.

Solvate Structure

Adds a solvent shell to the system according to the specific settings. By default, the necessary ions (Positive ions: Na+, Negative Ions: Cl-) are added in order to neutralize the system

AMBER

Structure Solvation in Amber.

Program: **Leap** (Ambertools)

- Available box types:: Shell, Spheric, Cubic and Truncated Octahedron.
- Box size specified by Å of spacing distance around the molecule.
- Water molecule type: TIP3P

NAMD

Program: **VMD** (solvate and autoionize plugins) from NAMD Package.

- Available box types: Cubic, and Spheric.
- Box size specified by Å of spacing distance around the molecule.
- Water molecule type: TIP3P.

GROMACS

Programs: **genbox** and **genion** from Gromacs Package.

- Available box types: Cubic, Triclinic, Dodecahedron and Truncated Octahedron.
- Box size specified by nanometers of spacing distance around the molecule.
- Water molecule type: SPC.

Molecular Dynamics simulations

Simple Box Solvent System Equilibration (NPT, NVT, or NVE)

Simple Box Solvent Molecular Dynamics (NPT, NVT, or NVE)

- Particle Mesh Ewald (PME) for full-system periodic electrostatics.
- Available Ensembles: NPT, NVT, NVE

NAMD - Amber

Program: **namd2** from NAMD package.
ForceField: Amber Parm99SB* for Proteins, Amber Parm99bsc0 for Nucleic Acids (DNA/RNA)

- Constant temperature dynamics via Langevin Dynamics.
- Constant pressure dynamics via Nose-Hoover Langevin piston.
- SHAKE is used to maintain all bonds involving hydrogen atoms at their equilibrium value

NAMD - Charm

Program: **namd2** from NAMD package.
Forcefield: Charmm

- Constant temperature dynamics via Langevin Dynamics.
- Constant pressure dynamics via Nose-Hoover Langevin piston.
- SHAKE is used to maintain all bonds involving hydrogen atoms at their equilibrium values

GROMACS

Program: **mdrun** from Gromacs package.

- Constant temperature dynamics via Velocity-rescale algorithm.
- Constant pressure dynamics via Parrinello-Rahman algorithm.
- LINCS Linear Constraint Solver was used to maintain all bonds at their equilibrium values.

Coarse Grained Molecular Dynamics Simulations

Coarse-Grained MD: Brownian Dynamics (C-Alpha)

Protein is put in a stochastic bath that keeps the temperature constant and modulates the otherwise extreme oscillations of the residues. This bath is simulated with two terms accounting for a velocity-dependent friction and stochastic forces due to the solvent environment. The potential energy used to compute forces assumes a coarse-grained representation of the protein (Calpha-only) and a quasi-harmonic representation of the interactions.

- Time: total simulation time.
- Delta_t: Time ellapsed between two consecutive time steps.
- Output Frequency: Number of frames after which a frame is written and considered for analysis.
- Force Constant: Measure of the strength of the spring connecting atoms measured in Kcal/mol*Å².

Coarse-grained MD: DISCRETE dynamics (C-Alpha)

Protein is modelled as a system of beads (**Calpha atoms**) interacting through a discontinuous potential (square wells). Outside the discontinuities, potentials are considered constant, thereby implying a ballistic regime for the particles (constant potential, constant velocity) in all conditions, except at such time as when the particles reach a potential discontinuity (event or collision). At this time, the velocities of the colliding particles are modified by imposing conservation of the linear momentum, angular momentum, and total energy. DMD has a major advantage over techniques like MD because, as it does not require the integration of the equations of motion at fixed time steps, the calculation progresses from event to event.

- Time: total simulation time.
- Delta_t: Time ellapsed between two consecutive time steps.
- Cut-off Distance: Maximum distance for the pairs of atoms to be included in the calculation.
- Sigma is the well amplitude for consecutive Calpha atoms.
- Sigma Go is the well amplitude for non-consecutive Calpha.

Coarse-grained MD: DISCRETE dynamics (Heavy atoms)

Protein is modelled as a system of beads (**ALL Heavy Atoms**) interacting through a discontinuous potential (square wells). Outside the discontinuities, potentials are considered constant, thereby implying a ballistic regime for the particles (constant potential, constant velocity) in all conditions, except at such time as when the particles reach a potential discontinuity (event or collision). At this time, the velocities of the colliding particles are modified by imposing conservation of the linear momentum, angular momentum, and total energy. DMD has a major advantage over techniques like MD because, as it does not require the integration of the equations of motion at fixed time steps, the calculation progresses from event to event.

- Time: total simulation time.

Coarse-grained MD: NMA (C-Alpha)

Normal Mode Analysis (NMA) can be defined as the multidimensional treatment of coupled oscillators from the analysis of force-derivatives in equilibrium conformations. This methodology assumes that the displacement of an atom from its equilibrium position is small and that the potential energy in the vicinity of the equilibrium position can be approximated as a sum of terms that are quadratic in the atomic displacements. Vectorial Anisotropic Network Model using coarse-grained (Calpha) protein representation is the formalism implemented here. Through the diagonalization of the hessian matrix, the ANM provides eigenvalues and eigenvectors that not only describe the frequencies and shapes of the normal modes, but also their directions.

Possibility to choose between three ANM formalism definitions of the force constants:

- 1. Linear algorithm uses a unique potential is used for all interactions considered.
- 2. Kovacs algorithm uses a distance-dependent potential.
- 3. Mixed algorithm uses a potentials adjusted to atomistic MD simulations.

- Force Constant: Measure of the strength of the spring connecting atoms measured in Kcal/mol*Å².
- Cutoff: Maximum distance for the pairs of atoms to be included in the calculation.

Molecular Dynamics Trajectory Simple Analysis

Plot RMSd along the trajectory

Root Mean Square deviation along the trajectory.

Programs: **Ptraj** from AmberTools package (Binpos,Crd,DCD,NetCDF trajectory formats). **g_rms** from Gromacs package (XTC trajectory format).

- Possibility to select a subset of atoms/residues from the whole system.
- User defined Mask only available for ptraj-compatible trajectory formats (Binpos,Crd,DCD,NetCDF).

Plot RMSd x Residue

Average Root Mean Square deviation per residue along the trajectory.

Programs: **Ptraj** from AmberTools package (Binpos,Crd,DCD,NetCDF trajectory formats). **g_rmsf** from Gromacs package (XTC trajectory format).

- Possibility of use a reference structure (PDB format).
- Possibility to select a subset of atoms/residues from the whole system.
- User defined Mask only available for ptraj-compatible trajectory formats (Binpos,Crd,DCD,NetCDF).

Plot BFactor per residue

Gets Bfactor values per residue from a Trajectory.

Programs: **Ptraj** from AmberTools package (Binpos,Crd,DCD,NetCDF trajectory formats). **g_rmsf** from Gromacs package (XTC trajectory format).

- Possibility to select a subset of atoms/residues from the whole system.
- User defined Mask only available for ptraj-compatible trajectory formats (Binpos,Crd,DCD,NetCDF).

Plot Radius of Gyration along the trajectory

Programs: **Ptraj** from AmberTools package (Binpos,Crd,DCD,NetCDF trajectory formats). **g_gyrate** from Gromacs package (XTC trajectory format).

- Possibility to select a subset of atoms/residues from the whole system.
- User defined Mask only available for ptraj-compatible trajectory formats (Binpos,Crd,DCD,NetCDF).

Trajectory format conversions

Compress trajectory to PCZ

Program: **pcazip.** Pcazip (http://mmb.irbbarcelona.org//software/pcasuite) compresses Molecular Dynamics (MD) trajectories using Principal Component Analysis (PCA) algorithms. Pcazip offers a good compression ratio at the expense of losing some precision in the trajectory.

- Quality param: Specifies the quality of the compression as a percentage of explained variance (default 90%)
- Gaussian RMSd param: weighted RMSd, giving less weight to atoms in flexible regions (e.g. loops).
- Number of Eigen Vectors Param: Specifies the number of eigenvectors that must be stored in the file. With 0 value pcazip will get the necessary number of vectors to obtain the percentage of explained variance (quality param).

Decompress PCZ trajectory

Uncompress trajectory from PCZ format to CRD format.

Program: **pcazip** (http://mmb.irbbarcelona.org//software/pcasuite).

Convert trajectory to a set of PDB Files

Program: **ptraj** from Ambertools package.

Convert trajectory to BINPOS Format

Program: **ptraj** from Ambertools package. BINPOS is the Scripps Binary format.

Converts trajectory to CRD Format

Program: **ptraj** from Ambertools package. CRD: Amber ASCII format.

Converts trajectory to DCD Format

Program: **ptraj** from Ambertools package. DCD: CHARMM/X-PLOR/NAMD Binary format.

Converts trajectory to NetCDF Format

Program: **ptraj** from Ambertools package. NetCDF (Network Common Data Form) is a machine-independent binary data format for array-oriented scientific data.

Trajectory manipulation

Programs: **Ptraj** from AmberTools package (Binpos,Crd,DCD,NetCDF trajectory formats). **g_rmsf** from Gromacs package (XTC trajectory format).

- Possibility to select a subset of atoms/residues from the whole system.
- User defined Mask only available for ptraj-compatible trajectory formats (Binpos,Crd,DCD,NetCDF)

Get Average Structure

Get a trajectory fragment

Gets a set of snapshots from a Trajectory.

Get a trajectory snapshot

Remove Water molecules and ions from trajectory

Return trajectory for a set of atoms

Protein Flexibility Analysis

Flexibility Analysis (FlexServ)

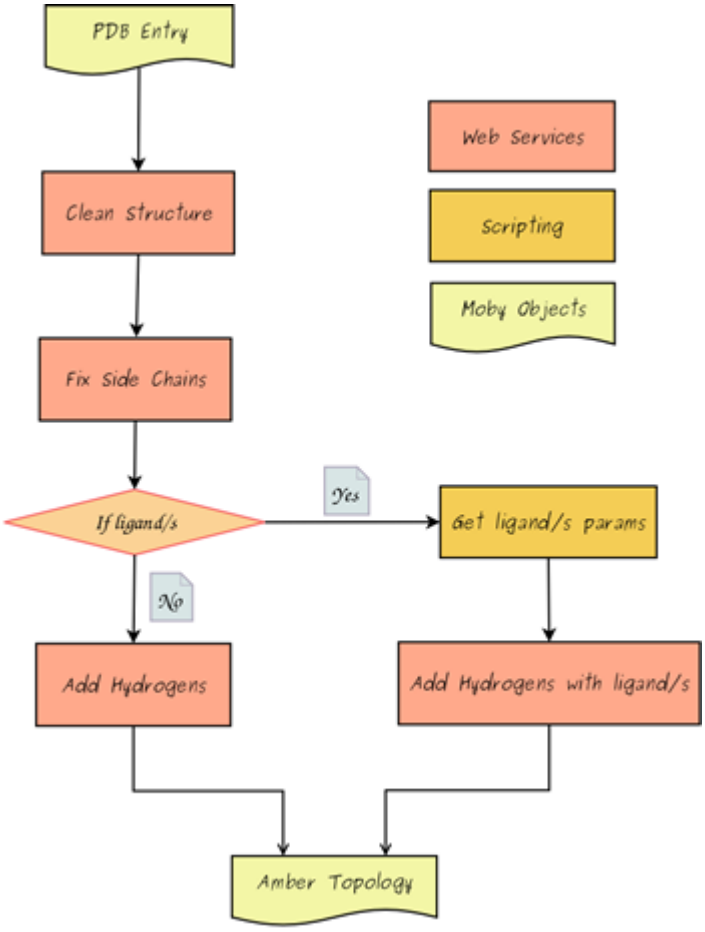
- Principal Components.
- Variance Profile.
- B-factors Landscape.
- Lindemann Coefficients.
- Apparent Stiffness.
- Hinge Point Prediction.
- Residue Correlations.
- Collectivity Indexes.

7.- Workflows

MDWeb is powered by a set of BioMoby Molecular Dynamics Web-Services (MDMoby).

One of the main advantages of web services is the possibility to interconnect them building complex pipelines called **workflows**.

To facilitate most usual MD operations, **MDWeb** offers a collection of pre-packed workflows. Settings have been adapted to run successfully on most systems, and their use for non-experts is recommended. However, MDWeb provides also the same functionality as separate operations, so the individual parameters could be adjusted. The pre-packed operations include:



- Workflows for generate topologies.
- Workflows for solvate and neutralize structures.
- Workflows for running a complete MD Setup.
- Workflows for equilibrate systems.

This help section shows a short description of the **workflows** available. Clicking on the workflow name, a new window will automatically open, showing the workflow as a graphical diagram, where input object/s, output object/s, web services and scripting pieces can be easily identified, together with the necessary interconnections.

- Setup for Amber Simulation
- Setup for NAMD Simulation
- Setup for Gromacs Simulation

Setup for Amber Simulation:

Generate Topology for Amber.

- *ForceField: Amber Parm99SB* for Proteins, Amber Parm99bsc0 for Nucleic Acids (DNA/RNA).*
- *Program: Leap from AmberTools Package.*

1. Remove crystallographic water molecules.
2. Add hydrogen atoms and missing side chain atoms as appropiate.

Amber MD Setup. Structure Setup for AMBER Forcefield.

- *ForceField: Amber Parm99SB* for Proteins, Amber Parm99bsc0 for Nucleic Acids (DNA/RNA).*
- *Programs: namd2 from NAMD Package, leap from AmberTools package, protpKa and CMIP.*

1. Generate Topology for AMBER.
2. Protonate Histidine residues according to protpKa program algorithm.
3. Add 20 water molecules at the energetically best favourable positions of the structure surface using CMIP program.
4. Energy minimize hydrogen atoms for 500 steps of conjugate gradients, while the rest of the structure is kept fixed.
5. Energy minimize the structure for 500 steps of conjugate gradients, restraining heavy atoms with a force constant of 50 Kcal/mol to their initial positions.

AMBER MD Setup with Solvation. Structure Setup + Solvation for AMBER Forcefield.

- *ForceField: Amber Parm99SB* for Proteins, Amber Parm99bsc0 for Nucleic Acids (DNA/RNA).*
- *Programs: namd2 from NAMD Package, leap from AmberTools package, protpKa and CMIP.*

1. AMBER MD Setup.
2. Set a truncated Octahedron box of TIP3P water molecules with a spacing distance of 15 Å around the system.
3. Add Cl- and/or Na+ ions necessary to neutralize the system. Then, add the appropiate amount of ions up to a concentration of 50 mM.
4. Further energy minimize the structure for 500 steps of conjugate gradients, restraining heavy atoms with a force constant of 50Kcal/mol to their initial positions.

AMBER Advanced Equilibration. System Equilibration.

- *Equilibration steps done in NPT ensemble with Periodic Boundary Conditions.*
- *Particle Mesh Ewald (PME) for full-system periodic electrostatics.*
- *Constant temperature dynamics via Langevin Dynamics.*
- *Constant pressure dynamics via Nose-Hoover Langevin piston.*
- *SHAKE is used to maintain all bonds involving hydrogen atoms at their equilibrium values.*

1. Heat solvent to 300K. Solute atoms restrained (force constant of 10 Kcal/mol). Length 5ps.
2. Reduce force constant to 5 Kcal/mol. Length 1ps.
3. Reduce force constant to 2.5 Kcal/mol and limit restraints to backbone atoms. Length 1ps.
4. Reduce force constant to 1 Kcal/mol. Length 1ps.
5. Simulation without restraints. Length 1ps.

Complete Setup for AMBER forcefield (Structure Setup + Solvation + Equilibration).

- *ForceField: Amber Parm99SB* for Proteins, Amber Parm99bsc0 for Nucleic Acids (DNA/RNA).*
- *Programs: namd2 from NAMD Package, leap from AmberTools package, protpKa and CMIP.*
- *Equilibration steps done in NPT ensemble with Periodic Boundary Conditions.*
- *Particle Mesh Ewald (PME) for full-system periodic electrostatics.*
- *Constant temperature dynamics via Langevin Dynamics.*
- *Constant pressure dynamics via Nose-Hoover Langevin piston.*
- *SHAKE is used to maintain all bonds involving hydrogen atoms at their equilibrium values.*

1. AMBER MD Setup with Solvation.
2. AMBER Advanced Equilibration.

Setup for Gromacs Simulation:

Generate Topology for GROMACS. Generate top and itp Topology Files for Gromacs.

- *Programs: pdb2gmx from Gromacs Package.*

1. Remove crystallographic water molecules.
2. Add side chain missing atoms using Leap from AmberTools package.
3. Add Hydrogen atoms using pdb2gmx.

GROMACS MD Setup. Structure Setup for Gromacs.

- *Programs: pdb2gmx, grompp, editconf, trjconv, make_ndx and mdrun from Gromacs Package, leap from AmberTools package, protpKa and CMIP.*

1. Generate Topology for GROMACS.
2. Protonate Histidine residues according to protpKa program algorithm.
3. Add 20 water molecules at the energetically best favourable positions of the structure surface using CMIP program.
4. Energy minimize hydrogen atoms for 500 steps of hydrogen conjugate gradients, while the rest of the structure is kept fixed.
5. Energy minimize the structure for 500 steps of structure conjugate gradients, restraining heavy atoms with a force constant of 500KJ/mol*nm2 to their initial positions.

GROMACS MD Setup with Solvation. Structure Setup + Solvation for Gromacs.

- *Programs: pdb2gmx, grompp, editconf, trjconv, make_ndx, genbox, genion and mdrun from Gromacs Package, leap from AmberTools package, protpKa and CMIP.*

1. GROMACS MD Setup.
2. Set a truncated Octahedron box of TIP3P water molecules (Amber FF) or SPC water molecules (other FF's) with a spacing distance of 15Å around the molecule.
3. Add Cl- and/or Na+ ions necessary to neutralize the system. Then, add the appropiate amount of ions up to a concentration of 50 mM.
4. Further energy minimize the structure for 500 steps of conjugate gradients, restraining heavy atoms with a force constant of 500KJ/mol*nm2 to their initial positions.

GROMACS Advanced Equilibration. System Equilibration.

- *Equilibration steps done in NPT ensemble with Periodic Boundary Conditions.*
- *Particle Mesh Ewald (PME) for full-system periodic electrostatics.*
- *Constant temperature dynamics via Velocity-rescale algorithm.*
- *Constant pressure dynamics via Parrinello-Rahman algorithm.*
- *LINCS Linear Constraint Solver is used to maintain all bonds at their equilibrium values.*

1. Heat solvent to 300K. Solute atoms restrained (Force constant of 400 KJ/mol*nm2. Length 5ps.
2. Reduce force constant to 300 KJ/mol*nm2. Length 1ps.
3. Reduce force constant to 200 KJ/mol*nm2 and limit restraints to backbone atoms. Length 1ps.
4. Reduce force constant to 100 KJ/mol*nm2. Length 1ps.
5. Simulation without restraints. Length 1ps.

GROMACS FULL MD Setup. Complete Setup for Gromacs Package (Structure Setup + Solvation + Equilibration).

- *Programs: pdb2gmx, grompp, editconf, trjconv, make_ndx, genbox, genion and mdrun from Gromacs Package, leap from AmberTools package, protpKa and CMIP.*
- *Equilibration steps done in NPT ensemble with Periodic Boundary Conditions.*
- *Particle Mesh Ewald (PME) for full-system periodic electrostatics.*
- *Constant temperature dynamics via Velocity-rescale algorithm.*
- *Constant pressure dynamics via Parrinello-Rahman algorithm.*
- *LINCS Linear Constraint Solver was used to maintain all bonds at their equilibrium values.*

1. GROMACS MD Setup with Solvation.
2. GROMACS Advanced Equilibration.

Setup for NAMD Simulation:

Generate Topology for NAMD. Generate PSF Topology for Charmm Forcefield.

- *ForceField: Charmm-27.*
- *Program: psfgen from NAMD Package.*
- *Warning: Ligands not allowed.*

1. Remove crystallographic water molecules.
2. Add hydrogen atoms and missing side chain atoms as appropiate.

NAMD MD Setup. Structure Setup for Charmm Forcefield.

- *ForceField: Charmm-27.*
- *Programs: psfgen, vmd (solvate and autoionize plugins) and namd2 from NAMD Package, protpKa and CMIP.*

1. Generate Topology for NAMD.
2. Protonate Histidine residues according to protpKa program algorithm.
3. Add 20 water molecules at the energetically best favourable positions of the structure surface using CMIP program.
4. Energy minimize hydrogen atoms for 500 steps of conjugate gradients, while the rest of the structure is kept fixed.
5. Energy minimize the structure for 500 steps of conjugate gradients, restraining heavy atoms with a force constant of 50Kcal/mol to their initial positions.

NAMD MD Setup with Solvation. Structure Setup + Solvation for Charmm Forcefield.

- *ForceField: Charmm-27.*
- *Programs: psfgen, vmd (solvate and autoionize plugins) and namd2 from NAMD Package, protpKa and CMIP.*

1. NAMD MD Setup.
2. Set a cubic box of TIP3P water molecules with a spacing distance of 15 Å.
3. Add Cl- and/or Na+ ions necessary to neutralize the system. Then, add the appropiate amount of ions up to a concentration of 50 mM.
4. Further energy minimize the structure for 500 steps of conjugate gradients, restraining heavy atoms with a force constant of 50Kcal/mol to their initial positions.

NAMD Advanced Equilibration. System Equilibration.

- *Equilibration steps done in NPT ensemble with Periodic Boundary Conditions.*
- *Particle Mesh Ewald (PME) for full-system periodic electrostatics.*
- *Constant temperature dynamics via Langevin Dynamics.*
- *Constant pressure dynamics via Nose-Hoover Langevin piston.*
- *SHAKE is used to maintain all bonds involving hydrogen atoms at their equilibrium values.*

1. Heat solvent to 300K. Solute atoms restrained (force constant of 10 Kcal/mol). Length 5ps.
2. Reduce force constant to 5 Kcal/mol. Length 1ps.
3. Reduce force constant to 2.5 Kcal/mol and limit restraints to backbone atoms. Length 1ps.
4. Reduce force constant to 1 Kcal/mol. Length 1ps.
5. Simulation without restraints. Length 1ps.

NAMD FULL MD Setup. Complete Setup for Charmm Forcefield (Structure Setup + Solvation + Equilibration).

- *ForceField: Charmm-27.*
- *Programs: psfgen, vmd (solvate and autoionize plugins) and namd2 from NAMD Package, protpKa and CMIP.*
- *Equilibration steps done in NPT ensemble with Periodic Boundary Conditions.*
- *Particle Mesh Ewald (PME) for full-system periodic electrostatics.*
- *Constant temperature dynamics via Langevin Dynamics.*
- *Constant pressure dynamics via Nose-Hoover Langevin piston.*
- *SHAKE is used to maintain all bonds involving hydrogen atoms at their equilibrium values.*

1. NAMD MD Setup with Solvation.
2. NAMD Advanced Equilibration.

8.- Icons

Projects menu

 Open project

 Download all project items in a compressed file, that can be restored back to **MDWeb**

 Delete project from the workspace (all items will be deleted).

Data types

 PDB Structure

 PDB Structure + PSF Topology (NAMD format).

 PDB Structure + PARMTop Topology (AMBER format).

 PDB Structure + Top (& ITPs) Topology (GROMACS format).

 MD Trajectory.

 MD Trajectory CRD Amber ASCII format.

 MD Trajectory DCD CHARMM/X-PLOR/NAMD Binary format.

 MD Trajectory XTC Gromacs Binary format.

 MD Trajectory NetCDF (Network Common Data Form), machine-independent binary data format for array-oriented scientific data.

 MD Trajectory compressed with Pcazip, using Principal Component Analysis (PCA) algorithms.

 Amino Acid sequence with an associated feature value as an array of real vaues (e.g. Bfactor x residue).

 Data as an XY array (e.g. RMSd vs time)

Data status

 Data computed and stored correctly.

 Data not yet available, process is still running.

 Data not available, process has failed.

Operations

 Quick info about the operation.

 Perform a new setup operation on the selected structure.

 Perform a new simulation/optimization procedure.

 Perform a new analysis.

 Visualize structure using Rasmol compatibles viewers (plug-in required).

 Visualize structure using JMol.

 Download object in a compressed tgz file.

 View log file.

 Delete item from the workspace.

9.- Software

MDWeb 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 MDWeb operations:

Program	Description	Package & Version
Biomoby	Biomoby web-services framework	BioMoby 1.0
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 4.0.2
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	Structure and dynamic analysis of trajectories	Ambertools 1.2
tgatoppm, pnmcrop, pnmtopng	Image management	Netpbm 10.0
Tleap	MD preparation program	Ambertools 1.2
VMD	Molecular Graphics Viewer	VMD 1.8.5

10.- References

Citing MDWeb:

Adam Hospital, Pau Andrio, Carles Fenollosa, Damjan Cicin-Sain, Modesto Orozco, Josep Lluís Gelpí
MDWeb and MDMoby: an integrated web-based platform for molecular dynamics simulations.
Bioinformatics, 2012, 28(9):1278-1279.

Alder B.J. and Wainwright T.E.
Studies in Molecular Dynamics.
I. General Method. Journal of Chemical Physics. (1959) 31, 459-466.

Benjamin T. Miller, Rishi P. Singh, Jeffery B. Klauda, Milan Hodošček, Bernard R. Brooks, H. Lee Woodcock III.
CHARMMing: A new, flexible web portal for CHARMM.
J. Chem. Inf. Model. (2008) 48, 1920-1929.

Aleksey Porollo, Jaroslaw Meller.
POLYVIEW-MM: web-based platform for animation and analysis of molecular simulation.
Nucleic Acids Research. (2010) 38, W662-W666.

Pradeep Kota.
GUIMACS - A java based front end for GROMACS.
In Silico Biology. (2006) 10, 0008.

Sanjit Roopra, Bernhard Knapp, Ulrich Omasits, Wolfgang Schreiner.
jSimMacs for GROMACS: A java application for advanced molecular dynamics simulations with remote access capability.
J. Chem. Inf. Model. (2009) 49, 2412-2417.

Christopher Bruns, Russ Altman, Vjay Pande.
OpenMM Zephyr Project.
Ongoing Project.

Sunhwan Jo, Taehoon Kim, Vidyashankara G. Iyer, Wonpil Im.
CHARMM-GUI: A web-based graphical user interface for Charmm.
J. Comput. Chem. (2008) 29, 1859-1865.

Diamantis Sellis, Dimitrios Vlachakis, Metaxia Vlassi.
Gromita: A fully integrated graphical user interface to Gromacs 4.
Bioinformatics and Biology Insights (2009) 3, 99-102.

BioMoby Consortium.
Interoperability with Moby 1.0-It's better than sharing your toothbrush.
Brieffings in Bioinformatics. (2008) 1, 1-12.

Hui Li, Andrew D. Robertson, Jan H. Jensen.
Very Fast Empirical Prediction and Interpretation of Protein pKa Values.
Proteins. (2005) 61, 704-721.

Delphine C. Bas, David M. Rogers, Jan H. Jensen.
Very Fast Prediction and Rationalization of pKa Values for Protein-Ligand Complexes.
Proteins. (2008) 73, 765-783.

Dmitrij Frishman, Patrick Argos.
Knowledge-based protein secondary structure assignment.
Proteins. (1995) 23(4), 566-579.

Sorin EJ, Pande VS.
Exploring the helix-coil transition via all-atom equilibrium ensemble simulations.
Biophys. J. (2005) 88(4), 2472-93.

DePaul AJ, Thompson EJ, Patel SS, Haldeman K, Sorin EJ.
Equilibrium conformational dynamics in an RNA tetraloop from massively parallel molecular dynamics.
Nucleic Acids Res. (2010) 38(14), 4856-67.

D.A. Case, T.E. Cheatham, III, T. Darden, H. Gohlke, R. Luo, K.M. Merz, Jr., A. Onufriev, C. Simmerling, B. Wang, R. Woods.
The Amber biomolecular simulation programs.
Computat. Chem. (2005) 26, 1668-1688.

Hess, B., Kutzner, C., van der Spoel, D., Lindahl, E.
GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation.
J. Chem. Theory Comput. (2008) 4, 435-447.

James C. Phillips, Rosemary Braun, Wei Wang, James Gumbart, Emad Tajkhorshid, Elizabeth Villa, Christophe Chipot, Robert D. Skeel, Laxmikant Kale, and Klaus Schulten.
Scalable molecular dynamics with NAMD.
Journal of Computational Chemistry. (2005) 26, 1781-1802.

Humphrey, W., Dalke, A. and Schulten, K.
VMD - Visual Molecular Dynamics.
J. of Molec. Graphics. (1996) 14, 33-38.

Jmol: an open-source Java viewer for chemical structures in 3D.
http://www.jmol.org/

Tim Meyer, Carles Ferrer-Costa, Alberto Pérez, Manuel Rueda, Axel Bidon-Chanal, Francisco Javier Luque, Charles A. Laughton, Modesto Orozco.
Essential Dynamics: A Tool for Efficient Trajectory Compression and Management.
J.Chem.Theor.Comput (2006) 2, 251-258.

Manuel Rueda, Pablo Chacón, Modesto Orozco.
Thorough Validation of Protein Normal Mode Analysis- A Comparative Study with Essential Dynamics.
Structure. (2007a) 15, 565-575.

Manuel Rueda, Carles Ferrer-Costa, Tim Meyer, Alberto Pérez, Jordi Camps, Adam Hospital, Josep Lluís Gelpí, Modesto Orozco.
A consensus view of protein dynamics.
Proc. Natl. Acad. Sci. (2007b) 104, 796-801.

Tim Meyer, Marco D'Abramo, Adam Hospital, Manuel Rueda, Carles Ferrer-Costa, Alberto Pérez, Oliver Carrillo, Jordi Camps, Carles Fenollosa, Dmitry Repchevsky, Josep Lluís Gelpí, Modesto Orozco.
MoDEL (Molecular Dynamics Extended Library): a database of atomistic molecular dynamics trajectories.
Structure. (2010) 18(11), 1399-1409.

Agusti Emperador, Oliver Carrillo, Manuel Rueda, Modesto Orozco.
Exploring the suitability of coarse-grained techniques for the representation of protein dynamics.
Biophysical Journal. (2008) 95, 2127-2138.

Jordi Camps, Oliver Carrillo, Agustí Emperador, Laura Orellana, Adam Hospital, Manuel Rueda, Damjan Cicin-Sain, Marco D'Abramo, Josep Lluís Gelpí, Modesto Orozco.
FlexServ: An integrated tool for the analysis of protein flexibility.
Bioinformatics. (2009) 25(13), 1709-1710.