Table of Contents

MDWeb Setup Tutorial

1.- Registration ................................. 3
2.- Starting Project ............................... 4
3.- Checking the Structure ....................... 5
4.- Structure Setup ............................... 7
5.- Waiting Results ............................... 9
6.- Getting Results .............................. 11

MDWeb Running Simulation Tutorial

1.- Generating Needed Files ...................... 12
2.- Downloading and Extracting Files .......... 13
3.- Preparing Downloaded Files ................. 15
4.- Launching Molecular Dynamics Simulations .... 17

MDWeb Protein Analysis Tutorial

1.- Registration ................................. 18
2.- Starting Project ............................... 19
3.- Uploading a Trajectory ........................ 20
4.- Running Analysis .............................. 22
MDWeb Setup Tutorial

MDWeb provides a friendly environment to setup new systems and run test simulations. With this short tutorial, you will be able to prepare a molecule to run a molecular dynamics simulation, following just a few steps.

1. **Registration**
2. **Starting Project**
3. **Checking the Structure**
4. **Structure Setup**
5. **Waiting Results**
6. **Getting Results**

## Tutorial Steps

1. **Registration**
   The first thing to do is choose between working as an *anonymous* user or alternatively as a *registered* user. We strongly recommend working as a registered user, as it has some important advantages.

   Anonymous user's projects are completely removed once the user is disconnected and also when session expires (after some minutes of inactivity), and therefore working as anonymous user is only suited for a first impression of the web server.

   Registration process will just take a minute --> Registration.

   Once logged in, the user workspace appears. In this workspace, all projects of the user will be shown.

   Now, we are ready to start our first MDWeb project.
2. Starting Project

**MDWeb** user can choose between four different kind of inputs, *Simulation*, *Analysis*, and *Upload a past MDWeb project*. In this tutorial, we will see an example of a Setup (Simulation) project.

![Project Setup Form](image)

You will be asked for just three easy inputs: a project title, an optional description of the project, and an input structure, either as a **PDB code** or directly as a **PDB structure file**.

**Base Structure**

- **Swiss-Prot or PDB Code Id (ex. 1ubq)**
- **User provided Structure (PDB format)**

*Input structure file should follow PDB File Format v.3.30 (July 13, 2011) more info...*

[Back] [Create Project]
3. Checking the Structure

Just after creating the project, you will be redirected to the so-called Checking page. In this page, MDWeb will show a list of possible problems found in the input structure that can affect to a future Molecular Dynamics Simulation. User can interactively inspect these problems thanks to a Jmol applet coupled to each of the problems, helping to locate them on the real tridimensional structure.

In this particular case, we are trying to work with the well-known Ubiquitin chromosomal protein (pdb code 1ubq). At the right part of the screen, you will find all the possible problems identified, whereas at the left part of the screen the Jmol applet with the structure loaded will be shown.

This small protein, with less than 80 aminoacids, has just two possible problems, marked as a warning symbol ⚠ or as an error symbol ✗ in the problem list.

![Warning messages are identified problems that can sometimes be adjusted/corrected in this same page. For example in this case there is an amide incorrect assignment in the residue number 49, that can be fixed with just a click in the desired button. The identification of this type of problems (together with all the possible problems checked) is described in the correspondent help section (Structure Checking).](image-url)
Error messages are identified problems that will probably cause Molecular Dynamics failure. The typical example is the one found in our example protein, an *atom clash*. Our ubiquitin protein has a pair of **Polar Acceptor atoms** too close (two oxygen atoms specifically, as you can perfectly see in the Jmol applet), just 3Å away. In this case, a simple energetic minimization of the system will put these pair of oxygen apart from each other, but sometimes a clash like this can be extremely dangerous for a Molecular Dynamics Simulation. **Atomic clashes** and **Sequence Gaps** are two of the most relevant problems in running MD simulations. More information about them can be found in the correspondent help section (**Structure Checking**).

Note that in this example, a **Polar Donor Clash** appears when we first load the structure, but it disappears when user fixes the **incorrect amide assignment** and the clashes are recomputed.

When clicking the next button, an **intermediate information page** will appear, listing the most important information chosen in the previous **Checking** page, i.e. **amide assignments** and **ligands discarded/taken**, and showing the part of the structure (chain/s) chosen for the project in a Jmol applet.

![Image of MDWeb interface](image.png)

In this particular case, we just did a modification in one amide, on the residue 49.
4. **Structure Setup**

At this point, we already have our structure loaded into MDWeb workspace. Now we are going to choose a workflow from the list of possible operations, after clicking at the New Operation icon 🔄. Our goal is obtain a system completely prepared to begin a Molecular Dynamic Simulation with Gromacs package. So we will click at the operation named: **Gromacs FULL MD Setup**. Note that choosing operation **Namd FULL MD Setup**, a system for Namd Package with Charmm forcefield will be prepared, and choosing **Amber FULL MD Setup**, a system for Amber Package with Amber forcefields will be prepared.

**Gromacs FULL MD Setup** will run a complete Molecular Dynamics Setup to the input structure. That means going from a protein/nucleic acid structure to a system formed by the molecule submerged into a solvent box, surrounded by ions up to a given concentration, and energetically equilibrated. For a more extended information about the steps done by the workflow, see the corresponding help section (**Gromacs Workflow FULL Setup**).

![MDWeb Interface](image)

After clicking at the workflow **Gromacs FULL MD Setup**, it will give us the possibility to choose between a set of different forcefields. A force field has three components: equations defining the potential energy of a molecular system as a function of atomic coordinates, atom types, and parameter sets that fit the equations to experimental data. We will just keep the option that is active by default in MDWeb. In this case, **Amber-99SB* forcefield**.
5. **Waiting Results**

Once the workflow is launched, we just have to wait for the results. **MDWeb** will be automatically reloaded from time to time after checking if the process has already finished. Meanwhile, you can follow the progress of it clicking at the sandclock 🕒 whenever you want.

When clicking at the **sandclock 🕒**, a new page will be opened showing the progress of the workflow step by step, accompanied by a figure of the whole workflow. This way, you will be able to know at any moment which service is being executed, how many steps have been already computed, and how many steps left.
MDWeb offers you the possibility to leave your processes running (as sometimes they will be long processes), disconnect from the server, and reconnect again later on, to check the progress of your projects. The list of projects is easily accessible from the user workspace page.
6. Getting Results

Once the workflow has finished, you have the possibility to download the data generated, just clicking at the download icon 📀.

Now you can work with your system, already prepared to run an MD Simulation with Gromacs package. The downloaded compressed file has all the necessary inputs to run the simulation, but still you will need to prepare the configuration file needed by gromacs to begin the run with some key parameters like the MD ensemble, simulation time, time step, etc. But don't worry, if you are not familiar at all with Gromacs, MDWeb will help you preparing these configuration input files for you. Check Run Simulation Tutorial.

This example is done with Gromacs package, but MDWeb also offers a complete setup workflow for working with Namd package with Charmm forcefield or Amber forcefields. The complete list of workflows can be found at help pages (MDWeb List of Workflows).
MDWeb provides a friendly environment to setup new systems and run test simulations. With this short tutorial, you will be able to run a **Molecular Dynamics Simulation** in your own machines, using files provided by MDWeb after preparing the system.

1. **Generating needed files**
2. **Downloading and Extracting Files**
3. **Preparing Downloaded Files**
4. **Launching Molecular Dynamics Simulation**

**Tutorial Steps**

1. **Generating needed files**

   This short tutorial assumes that user have already setup a system using one of MDWeb Workflows to run a complete system preparation: *GromacsWorkflowFULL, AmberWorkflowFULL, or NamdWorkflowFULL* (see **Setup Tutorial**), or alternatively it has been prepared step by step using the appropriate MDWeb operations.

   The first thing to do then is run an operation called *Simple Box Solvent Molecular Dynamics* with the MD ensemble of interest *(NPT, NVE or NVT)*. It can be found in the **MDWeb Simulation Operations**, clicking at the icon 🛠️.
We will need to fill in the parameter fields with desired values. For example, if what we want is to run a Molecular Dynamics Simulation of 10 ns with a timestep of 2 fs at a 300 K of temperature (a standard molecular dynamics first step, to completely equilibrate the system), we just have to write these numbers: **Time Step:** 2 fs, **Temperature:** 300 K, **Total Time:** 10000 ps, and **Output Frequency:** 500 steps. Note that for choosing the output frequency, the total number of snapshots that will be written by the MD Simulation are computed on the fly and shown in red color. In this case, we have chosen to write every 500 steps in order to have 10000 snapshots, one per picosecond.

Note the line saying: **Only return Configuration Files (without running simulation).** Leaving this checked MDWeb will give you all the necessary input files to run the simulation, together with a README file with instructions about how to run it. Unchecking this line, MDWeb will perform the simulation, however, in the present case you will get an error message, because only simulations up to 5 ps are allowed in our server due to the limitations in computer resources.

2. **Downloading and Extracting Files**

Now, in the last operation results, we have to click at the download icon to get all the necessary files to run the prepared Simulation. The file downloaded is a gzip-compressed tar file, which can be easily uncompressed and unpacked with the appropriate program.

The list of files extracted, depending on the Workflow Program, looks like this:
## GROMACS Files:

<table>
<thead>
<tr>
<th>File Extension</th>
<th>File information</th>
</tr>
</thead>
<tbody>
<tr>
<td>pdb</td>
<td>MD Input coordinate file in PDB format.</td>
</tr>
<tr>
<td>gro</td>
<td>MD Input coordinate file in GRO format.</td>
</tr>
<tr>
<td>top</td>
<td>MD Input structure topology.</td>
</tr>
<tr>
<td>itp</td>
<td>MD Input include topology files.</td>
</tr>
<tr>
<td>ndx</td>
<td>MD Input index file, containing user definable sets of atoms.</td>
</tr>
<tr>
<td>trr</td>
<td>MD Input restart file, storing periodic box information, velocities and positions of all atoms.</td>
</tr>
<tr>
<td>edr</td>
<td>MD Input restart energy file, storing energies, temperature, pressure, box size, density and virials.</td>
</tr>
<tr>
<td>mdp</td>
<td>GROMACS Input configuration file.</td>
</tr>
<tr>
<td>gromacs</td>
<td>README.gromacs, explaining how to run the simulation.</td>
</tr>
<tr>
<td>tpr</td>
<td>MD Run Input File, in binary format, containing the starting structure of your simulation (coordinates and velocities), the molecular topology and all the simulation parameters.</td>
</tr>
</tbody>
</table>

*queueSP* --&gt; C-shell script to run MD simulation on a single processor.

*queueMPI* --&gt; C-shell script to run MD simulation on multiple processors with MPI.

*queueSGE* --&gt; C-shell script to run MD simulation on a queue manager (SGE).

## NAMD Files:

<table>
<thead>
<tr>
<th>File Extension</th>
<th>File information</th>
</tr>
</thead>
<tbody>
<tr>
<td>pdb</td>
<td>MD Input coordinate file in PDB format.</td>
</tr>
<tr>
<td>psf</td>
<td>MD Input structure topology.</td>
</tr>
<tr>
<td>xsc</td>
<td>MD Input eXtended System Configuration.</td>
</tr>
<tr>
<td>pos</td>
<td>MD Input structure positions.</td>
</tr>
<tr>
<td>vel</td>
<td>MD Input structure velocities.</td>
</tr>
<tr>
<td>in</td>
<td>NAMD Input configuration file.</td>
</tr>
<tr>
<td>namd</td>
<td>README.namd file, explaining how to run the simulation.</td>
</tr>
</tbody>
</table>

*queueSP* --&gt; C-shell script to run MD simulation on a single processor.

*queueMPI* --&gt; C-shell script to run MD simulation on multiple processors with MPI.

*queueSGE* --&gt; C-shell script to run MD simulation on a queue manager (SGE).
A **README** file for each of the programs explains briefly how to use the given **scripts** to run the simulation. Let's take a look at one example. We will use the preparation done in the [Setup Tutorial](#) help part, with a **GROMACS Workflow**.

### 3. Preparing Downloaded Files

First of all, let's take a look at the **README** file. In one section, it is said that a few lines of the shell scripts queue* need to be modified before launching the simulation. They refer to MD program paths and libraries. These are enclosed between these marks:

```
#############################################################
# Modify the following lines according to your installation
# --> Lines to modify <--
#############################################################
```

So we are going to edit the file named **queueSP***, that corresponds to a **single processor** run. The same should be done with the corresponding **queueMPI*** file to launch the simulation in **Multiple Processors** or with **queueSGE*** to use SGE queue manager.
In our case, we know that our Gromacs mdrun executable is located at `/opt/gromacs/bin/`, so we just need to modify the `Path to GROMACS` variable:

```
set GROMACS=/opt/gromacs
```

to this new one:

```
set GROMACS=/opt/gromacs
```

The `trjconv` binary is located at the same directory, so we don't have to modify the line:

```
set TCONV=$GROMACS/bin/trjconv
```

And that's all! We already have our shell script prepared to launch the simulation.
4. **Launching Molecular Dynamics Simulation**

Once again, a section of the README file can help us in this task. It is said that if executing in Single Processor mode, we just need to type:

```
csh queueSP* &
```

where `queueSP*` can be replaced by the complete real name of the csh file.

And if all was fine, you will be able to follow the progress of your simulation from the output file named `*_mdrun.log`, where information about the completed and remaining steps is written, together with an estimation of the finishing time:

```
starting mdrun 'Protein in water'
5000000 steps, 10000.0 ps.
step 0

NOTE: Turning on dynamic load balancing

vol 0.80 imb F 2% step 100, will finish Tue Oct 18 11:20:51 2011
vol 0.81 imb F 1% step 200, will finish Tue Oct 18 11:49:35 2011
vol 0.80 imb F 1% step 300, will finish Tue Oct 18 07:22:22 2011
vol 0.80 imb F 0% step 400, will finish Tue Oct 18 05:08:26 2011
...
```

For launching simulations with other packages (**AMBER** and **NAMD**), the steps of this short tutorial should be completely equivalent.
MDWeb Tutorial [Protein Analysis Tutorial]

MDWeb Protein Analysis Tutorial

MDWeb provides a friendly environment to analyse your own generated molecular dynamics trajectories. With this short tutorial, you will be able to upload a trajectory and run a set of analysis, checking for example the stability of your system or information about flexibility.

1. **Registration**
2. **Starting Project**
3. **Uploading a trajectory**
4. **Running Analysis**

Tutorial Steps

1. **Registration**
   The first thing to do is to choose between working as an anonymous user or alternatively as a registered user. We **strongly** recommend working as a registered user, as it has some important advantages.

   Anonymous user's projects are completely removed once the user is disconnected and also when session expires (after some minutes of inactivity), and therefore working as anonymous user is only suited for a first impression of the web server.

   Registration process will just take a minute --> Registration.

   Once logged in, the user workspace appears. In this workspace, all projects of the user will be shown.

   Now, we are ready to start our first MDWeb analysis project.
2. **Starting Project**

**MDWeb** user can choose between two different kinds of inputs, *Simulation* and *Analysis*. In this tutorial, we will see an example of an Analysis project.

You will be asked for a small set of data and input files: First of all, you may fill in a project title and a description (optional) at the top part of the form. Then, at the *Input Type* drop-down selector, just under the description, you should choose the option *Analysis (MD Trajectory)*.

When choosing *Analysis* as input type, a new screen (the one corresponding to Analysis, *MD Trajectory*) will appear.

User may select the appropriate **tool, trajectory and topology** format to work with the input files. In the next section we will see an example of how to build the trajectory and topology files with a size small enough to be transferred through the net to **MDWeb** server.
3. **Uploading a trajectory**

In this example, we are going to compute a set of basic analyses to a 10 ns molecular dynamics simulation of a **Serine Hydrolase** (pdb code: 1eve). First of all we need to reduce the size of the 10 ns trajectory to be uploaded to MDWeb. In this case, we will just extract 20 snapshots from the original obtained trajectory (one snapshot every 0.5ns). Alternatively, if you wish to have more resolution, you can extract more snapshots by removing solvent molecules, thus drastically reducing the trajectory file size, but to be able to properly analyse this solvent-stripped trajectory, a coherent topology must be generated, and that's usually not an easy job.

Example of reducing trajectory size:

**XTC Trajectory:**

Use `trjconv` from Gromacs package to reduce the size of a trajectory in XTC format.

Usage: `trjconv -sp -s structure.gro -f trajectory.xtc -o trajectory.20.xtc -b 1 -e 10000 -dt 500 -ur compact -pbc atom`

where `structure.gro` is the structure topology in GRO format, `trajectory.xtc` is the trajectory in XTC format, `1-10000` are the starting and ending snapshot times (in ps), and `500` is the offset (write every 500 ps). The rest of the commands are useful when dealing with non-cubic boxes (-ur) and periodic boundary conditions (-pbc).

You will be asked for a group to write to the output file, e.g.:

*Will write xtc: Compressed trajectory (portable xdr format)*

*Select group for output*

*Opening library file /usr/local/gromacs/share/gromacs/top/aminoacids.dat*

*Group 0 ( System) has 99637 elements*

*Group 1 ( Protein) has 8380 elements*

*Group 2 ( Protein-H) has 4263 elements*

*Group 3 ( C-alpha) has 534 elements*

*Group 4 ( Backbone) has 1602 elements*

*etc.*

We are going to choose **System** (group 0), to write all the system as output and hence being coherent with the system topology.

Executing these commands will generate a new trajectory with 20 snapshots (10000/500), named trajectory.20.xtc, in **Gromacs XTC** format, ready to be uploaded to **MDWeb** server.
CRD, Binpos, NetCDF, DCD trajectories:

Use `ptraj` from Ambertools package to reduce the size of a trajectory in DCD, NetCDF, Binpos or CRD formats.

Usage: `ptraj structure.top < ptrajReducing.in`
where `structure.top` is the structure topology in PRMtop or PSF format, and `ptrajReducing.in` is a file containing the appropriate commands:

```
trajin trajectory.crd 1 10000 500 # Input trajectory, first and last snapshot to read and offset.
trajout trajectory.20.crd # Output trajectory, in Amber ASCII CRD format (ptraj default).
go
```

Executing these commands will generate a new trajectory with 20 snapshots (10000/500), named trajectory.20.crd, in Amber ASCII CRD format, ready to be uploaded to MDWeb server.

To start our analysis, we are going to put this new generated 20 snapshots trajectory in the **Trajectory File** input box, and the associated topology (the same file used for the molecular dynamics simulation) in the **Topology File** input box.
4. **Running Analysis**

Once we have our trajectory uploaded, we can begin with the analysis. **MDWeb** offers a set of basic analysis that can be divided in 3 different sections:

- Trajectory manipulation: format conversions, slice/snapshot extraction, etc.
- Analysis per Residue/Structure.
- Flexibility Analysis (link to [FlexServ](#)).

In this specific example we are not interested in converting the trajectory format, but we will begin applying some manipulations to it. The most usual and important one is stripping the solvent, usually together with the counterions, when user is only interested in studying the protein molecule.

To do that, click on the operation called **Remove water molecules and ions from trajectory**. We can choose whether we would like to remove only water molecules, only counterions or both. In this case, we will remove both, obtaining the trajectory (and the correct topology associated) for just the structure of interest.
The first noticeable thing once we have the "dry" trajectory, is that it is really smaller in size than the solvated one. That allows the user to animate the trajectory with Jmol applet. If we try to visualize the original solvated structure, MDWeb will show a warning message because the trajectory is too large, recommending to remove solvent. So, clicking at the corresponding Jmol icon, a new window with the Jmol interactive visualizer will be opened. It is a good first approximation to take a look at the dynamics of the structure, before obtaining analysis data.

Now we can begin with the basic data analysis. From the list, we can choose two basic analysis as Root Mean Square deviation and Radius of Gyration. With these two analysis we will be able to see the deviation of the structure during the dynamics, whether it maintains its conformation during the 10 ns length or it suffers some distortion.

Each of the analysis plot can be visualized clicking at the correspondent icon.
We can apply a couple of additional analysis to get a first impression about the flexibility of the structure. These are the **Root Mean Square deviation per Residue** and the **Bfactor per Residue** analyses. Taking a look at the plots, we can easily identify the regions of the structure having more mobility (usually loops) and the regions that are almost kept fixed during the whole simulation.
If we are interested in more flexibility analysis, **MDWeb** offers the possibility to send the project to our flexibility server **FlexServ**. Just click on the operation called **Flexibility Analysis**.

**FlexServ** is a server aimed to show in an easy and visual way flexibility-related properties of proteins. It computes a big set of flexibility analysis like:

- Principal Components.
- B-factor landscape.
- Lindemann Coefficients.
- Hinge Point Predictions.
- Residue Correlations.
- etc.
So with this set of analysis, we have been able to see that the simulated structure has not suffered critical structural deformations. It has not gone more than 2Å away from the initial structure, and the **Radius of Gyration** has been also maintained around 20Å. From the flexibility part, we identified some flexible regions from the **Bfactor** and **RMSd per residue** analysis, that are located mostly in loop regions. We launched a **FlexServ** analysis to obtain more information about these flexible part.