Inhibitor design using MOLOC — tutorial

Table of Contents:

Getting started

Installation instructions are available at www.moloc.ch/installation.html. Note that getting a license for your own computer is time-consuming and only recommended for those who will do extensive research work with Moloc.

These instructions are intended as a quick-start guide and finish with a simple exercise in optimization of a structure-based inhibitor design. For a very detailed beginner tutorial point your web browser to the appropriate files in the installed directory on your computer (file path depends on where Moloc is installed: it is in the Applications folder on a Mac).

moloc/www/bgnr/bgnr.html

To run this tutorial, you will need the file "PMII_tutorial_1.cif" (which contains the malarial enzyme Plasmepsin II and a known inhibitor fragment) and work with it from beginning to end. If something goes wrong, you can compare to the related files numbered 2–10 that are exemplars saved at various stages of the tutorial as indicated.

To open the program MOLOC on a Macintosh, you must have the program X11 open and running, open an X11 terminal, and type the full path of the program executable (normally /Applications/moloc/bin/Moloc).

Starting the program on a Windows PC depends on the installation. Various methods for doing so are given at http://moloc.ch/installation_hints.html.

Part 1: Mouse practice and opening files. PMII_tutorial_1.cif

RIGHT CLICKS ARE FOR MENU NAVIGATION.

EXAMPLE: OPENING A .CIF FILE:

1) RIGHT CLICK on the **???** menu item on the bottom menu bar.
2) RIGHT CLICK the **[…] display** menu. This menu is used for all file opening, closing, and saving.
3) RIGHT CLICK **[d] set working directory** and navigate to the directory that houses your files.
4) RIGHT CLICK **[g] get entry from file**.
5) RIGHT CLICK file type **[i] cif-file**. A string requestor box opens. Hit ENTER and all .cif files in the working directory will be displayed.
6) Select the one(s) you want to open (PMII_tutorial_1.cif) by RIGHT CLICKING to change the 'n' to a 'y'
7) RIGHT CLICK **[x]** to end your selection and open the file(s). Hit ENTER to ignore the question about libraries.
8) There's probably nothing on the screen because the view needs to be adjusted. To auto-center things, RIGHT CLICK **[x]** one more time and then RIGHT CLICK **[z] center visible entries.**

This tutorial will no longer specify "RIGHT CLICK" when referring to menu navigation. With a little practice it will become second nature.

*As you've probably noticed, Moloc is operated using a series of hierarchical menus. You go forward one step by selecting the menu item of your choice; you go backward one step by selecting **[x] exit** at the bottom of the menu.*

Protein structures are universally available as .pdb files. Reading in a .pdb file and preparing it for use in Moloc requires some extra steps (i.e. removing extraneous waters, removing explicit protons) that are covered in the appendices of this document. All the files you'll need for this course will be provided as ready-to-use .cif files.

LEFT/MIDDLE CLICKS ARE FOR MOLECULE MANIPULATION.

LEFT BUTTON FOR ROTATION:
-Hold down the LEFT mouse button and move the mouse to ROTATE the object.

MIDDLE BUTTON FOR TRANSLATION:
-Hold down the MIDDLE mouse button to TRANSLATE laterally.

LEFT AND MIDDLE TOGETHER FOR TWO KINDS OF ZOOMING:
-Hold down LEFT and MIDDLE mouse button together and move up/down to TRANSLATE in and out of the screen.
-Hold down LEFT and MIDDLE mouse button together and move left/right to ZOOM in and out.

CTRL-CLICKING IS FOR PICKING OF ATOMS, BONDS, OR FRAGMENTS *(without the CTRL button the normal view rotation happens).*

Try it: in this view doing CTRL-LEFT CLICK on an atom multiple times will toggle through different labels for the atom/residue.

Part 2. Manipulating different molecules as "entries" and saving files

Moloc allows you to do different things to different molecules in the same workspace by classifying each molecule as a distinct "entry."

EXAMPLE: CHANGING ENTRY APPEARANCE AND "ACTIVITY"

1) From the **display** submenu choose **[a] entry activity stati**.
2) A small window with the name of each entry will appear. Clicking on the letters beside each entry will cycle through the options **A** (= Active), **V** (= Viewable, but not active), and **BLANK** (= hidden). Do this for each of the entries and you'll see that there are two: the protein and the inhibitor.
3) Leave each entry in the **A**ctive form… the calculations we'll do later will not work if they are only marked **V**iewable. Then **[x] exit**.

It's a good idea to make the inhibitor look different than the protein.

4) From the **display** submenu choose **[e] change entry settings**.
5) Choose the **inhib** entry so you can change its appearance.
6) Select **[c] color entry** and choose a different color.
7) Select **[a] atom types on/off** to toggle colored atom types (blue = N, red = O, yellow = S, etc.)
8) Now **[x] exit** and repeat to change the protein's appearance too (make sure the protein is a different color than the inhibitor)

Moloc has no "undo" function. Save early, save often. The .cif file format is the most convenient for work within Moloc — .pdb files are good for exporting to other pieces of software.

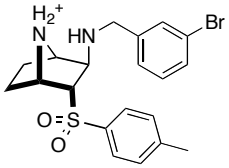
EXAMPLE: SAVING A FILE

1) From the **display** submenu choose **[s] store entries on files**.
2) Choose **[i] cif-file**.
3) Click **[x]** to ignore libraries
4) Type the file name you want and hit enter.

Part 3. Building, changing, rotating, and translating inhibitors

PMII_tutorial_2.cif

The inhibitor should now be clearly visible within the protein structure. It is a bicyclic diammonium ion that looks like this:



Look at the region of the protein around the inhibitor's toluenesulfonyl group. It is pointing at a large binding pocket not currently engaged by the inhibitor. Increasing the size of the methyl substituent will better fill this binding pocket.

*Adding and deleting atoms and functional groups is done within the **[dTp] topology changes** menu.*

EXAMPLE: ADDING SUBSTITUENTS TO AN EXISTING INHIBITOR

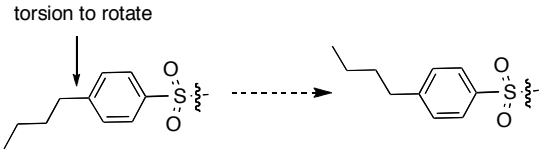
1) From the main menu, enter the **[dTp] topology changes** menu.

2) Click **[a] add and delete atoms**.
3) By default the atom type to be added is **[C]** carbon. Remember what we said about picking being done with the CTRL button. Now CTRL-LEFT CLICK on the methyl group to add an additional carbon. Repeat the action on the terminal carbon two more times to arrive at an n-butyl-substituted aryl group.
4) For practice, try removing an atom that you just added by CTRL-MIDDLE CLICKing on it.
5) When you're done adding and removing carbons, make sure that you have a butyl chain installed.
6) Now **[x] , [x]** to return to the main menu.

To manually adjust the positions of atoms before *structure optimization, the easiest method is to rotate single bonds until the substituents are pointing the way you want them to. Rotation of bonds and moving of fragments is all handled within the **[frg] forge: coordinate changes** menu.*

EXAMPLE: ROTATING BONDS TO ADJUST THE POSITION OF SUBSTITUENTS

1) From the main menu, enter the **[frg] forge** menu.
2) Manipulate the view so you have a top-down view of the tosyl aryl ring. It should look something like the structure below, left:



3) Now we want to flip the indicated torsion angle by 180° as indicated.
4) Select **[t] drive torsions**
5) CTRL-LEFT CLICK on the bond you want to rotate. If you pick it properly, it will change color, some dotted lines will appear to indicate which 4-atom torsion angle you're about to adjust, and the numerical value of the torsion angle (180°) will also appear. If you click the wrong bond, just CTRL-LEFT CLICK on a different one.
6) To change the value of the torsion angle, SHIFT-LEFT CLICK on the highlighted bond and drag the mouse left and right. For this example, set the angle somewhere around 0 (±10).

SHIFT-CLICKING (LEFT OR MIDDLE) IS USED TO MOVE PICKED ATOMS, BONDS, AND FRAGMENTS AROUND RELATIVE TO EACH OTHER.

7) Now try SHIFT-MIDDLE CLICKING on the highlighted bond and dragging left and right. Notice the important difference, but leave the value close to where it started.

*The **[frg]** menu also contains the tools for first selecting a subset of atoms and then translating and rotating them relative to other molecules in the workspace using. These tools work in a closely analogous manner to the bond rotation that we just executed: First select the set you want to move using **([s] select set**… **[e] entries**… CTRL-LEFT-CLICK on the entry you want to move… and then **[x] exit**). Then select **[m] move set**, and do the desired rotation and translation using SHIFT-LEFT CLICK and SHIFT-MIDDLE CLICK, respectively. Explore these functions on your own time… we don't need them now.*

Part 4. Performing an optimization

PMII_tutorial_3.cif

You're ready to do a structure optimization and see how the inhibitor binds into the active site. At the heart of this approach is the idea that you can perform a reasonable structure optimization using a "frozen" protein and a mobile inhibitor. In reality, proteins are much more flexible than this method gives them credit for, but considering them as rigid objects is a necessary approximation that allows for quick and easy calculations. Such structure optimizations only find the nearest local minimum in the potential energy surface, so you have

to have the inhibitor set up in an orientation very close to what you expect for a binding mode. Hence all the work on moving and rotating bits of inhibitor BEFORE doing the optimization.

EXAMPLE: STRUCTURE OPTIMIZATION

1) From the main menu, enter the **[opt] optimize with MAB force field** menu.
2) Select **[s] define stationary atoms**.
3) Define the type of selection you want to make by clicking on **[e] entries**.
4) CTRL-LEFT CLICK on any atom in the protein. It should become highlighted with white circles.
5) Click **[x]** to go back to the **[opt]** menu.
6) Click **[o] optimize (do force field calculation)** and watch it work. When it finds a local energy minimum it will come to a stop.

Part 5. Evaluating the results

PMII_tutorial_4.cif

When using Moloc, there is no numerical metric for an inhibitor's suitability. Instead, the evaluation of potential inhibitors relies heavily on the modeler's understanding of conformations and attractive forces. The goal is to design an inhibitor that forms as many favorable contacts with the protein as possible while avoiding the following red flags:

-ANY amount of steric overlap.
-The inhibitor assuming an energetically unfavorable conformation
-Poor distribution of hydrophilic and hydrophobic groups relative to protein and surrounding solvent

More advanced designs also rule out candidate structures based on a plethora of criteria, including structural predictors of pharmacological properties (Lipinski's rules) and avoiding the incorporation of functional groups that are rapidly metabolized by liver enzymes (e.g. benzylic methylene groups). These are best implemented by med. chem. experts, and can be ignored by the beginner.

All of these evaluations are primarily carried out as "manual" processes. But Moloc does have a few tools for evaluating weak interactions and steric overlap that are worth learning.

EXAMPLE: CHECKING AN OPTIMIZED STRUCTURE FOR WEAK ATTRACTIONS AND STERIC OVERLAP

1) After an optimization, Moloc will have left you on the **[opt]** menu.
2) Click **[e] energy examination**
3) Cllck **[h] H-bonds**. All intra- and intermolecular hydrogen bonds will be shown. They are colour-coded according to the ideality of their angles and lengths, and a legend with some numerical evaluation of their strengths is given in the lower-right-hand corner of the screen. Yellow = strongest, Blue = weakest.

Examine a couple of hydrogen bonds to get a feel for what "good" and "bad" hydrogen bonds look like in this context.

4) In reality, the only H-bonds that you care about are those intermolecular ones between protein and ligand. Click **[*] all** twice to toggle the display to show **inter**molecular interactions, and then click **[h]** again. Now only the hydrogen bonds between inhibitor and protein are displayed.

The bonds in this case are all yellow – if you're structure had some blue or purple H-bonds you'd want to do a more detailed investigation of how you could improve them.

Next you want to check for the dreaded steric overlap, and to check for where better vdW surface matching may be possible.

5) Click **[r] repulsive (positive) v.d.Waals**. In the best case, you won't see any interactions at all. One or two that are so weak as to be rated "blue" are probably okay. But still, note their location and think about how you might improve upon them.
6) Now you can get a feel for where better shape (vdW) complementarity can be achieved by **SHIFT**-clicking **[r] polar-apolar or negative**. This will show you where favourable vdW contacts are made.
7) Similar checks of repulsive and attractive Coulomb (electrostatic) interactions can be done using the **[c]** button, but be warned that EVERY structure will show you some repulsive and some attractive interactions and that these are rarely useful for evaluating an inhibitor's quality.

*Note while in the **[r]** view that the butyl tail on your inhibitor doesn't experience many favorable vdW contacts even though it is pointed into a pocket. Rotate the structure around while examining the end of the butyl chain and you'll see that there's lots of room for it to get bigger. Now read on for one last section on structure optimization.*

Part 6. An exercise in design optimization

PMII_tutorial_4.cif

Many structural designs are possible, but the designer must temper her/his enthusiasm for creative designs with a realistic understanding of how easy to make a designed structure will be. Because all designed molecules must be made before they can be tested, the most productive designers are synthetic chemists themselves, or computational chemists who work closely with synthetic chemists throughout the design process. Given that we know the methyl and butyl inhibitors can be made, we can assume for this exercise that installing a variety of straight-chain alkyl substituents at the same position will be easy.

Now extend the chain one methylene at a time from butyl to decyl (following the protocol in Part 3), minimize each structure (Part 4), evaluate the results (Part 5), and save each inhibitor as a separate .cif file (Part 2) along the way. Consider repulsive and attractive vdW forces AND alkyl chain conformations in order to predict what the optimum alkyl chain for this inhibitor would be.

C5 chain = PMII_tutorial_5.cif
C6 chain = PMII_tutorial_6.cif
C7 chain = PMII_tutorial_7.cif
C8 chain = PMII_tutorial_8.cif
C9 chain = PMII_tutorial_9.cif
C10 chain = PMII_tutorial_10.cif

Appendices — Miscellaneous other Moloc functions

A. Importing a protein structure from protein databank (.pdb) format (many are available at www.rcsb.org).
There is some prep work to do if you want to get a pdb file into shape for use with Moloc. You must remove water molecules, explicit hydrogen atoms (if present), and other extraneous molecules, and have the protein, cofactors, and any other molecules that you need present as separate entries in order to make their subsequent manipulation much easier.

1) RIGHT CLICK on the **???** menu item on the bottom menu bar.
2) RIGHT CLICK the **[…] display** menu. This menu is used for all file opening, closing, and saving.
3) RIGHT CLICK **[d] set working directory** and navigate to the directory that houses your file.
4) RIGHT CLICK **[g] get entry from file**.
5) RIGHT CLICK file type **[p] pdb-file**. A string requestor box opens. Hit ENTER and all .pdb files in the working directory will be displayed.
6) Select the one you want to open by RIGHT CLICKING to change the 'n' to a 'y'
7) RIGHT CLICK **[x]** to end your selection and open the file. Hit ENTER to ignore the question about libraries.
8) Select <u>one</u> molecule from the list of chains, water molecules, etc. that is presented to you. *Odds are you only want the protein and its cofactors, and no waters at all. It is easiest if you go through this process and bring in only the things you want, one thing at a time, because they will already then be available to you as separate entries that be easily manipulated.*
9) There's might still be nothing on the screen because the view needs to be adjusted. To auto-center things, RIGHT CLICK **[x]** one more time and then RIGHT CLICK **[z] center visible entries.**

Once the molecules are imported and set up as separate entries, it is highly recommended that you save as a .cif file for all subsequent Moloc modeling.

1) From the **display** submenu choose **[s] store entries on files**.
2) Choose **[i] cif-file**.
3) Click **[x]** to ignore libraries
4) Type the file name you want and hit enter.

See also Appendix H for instructions on how to generate a new inhibitor fragment as a separate entry.

B. Save separate entries as separate .cif files. If each inhibitor and each protein are saved as individual .cif files, multiple inhibitors can be easily read in and superimposed within the active site of a single copy of the protein. Convenient and efficient.

Simply do the normal save routine: **display** menu… **[s] store entries**… **[i] cif-file**… **[x] ignore** libraries… THEN ONLY TICK NEXT TO THE ENTRY THAT YOU WANT TO SAVE before typing in the name for the file.

C. Export protein and inhibitor to separate pdb files. This allows you to view them in another program like PyMol. The benefits of PyMol are numerous… you can make nice pictures, you can control coloring of different structural elements much more easily, and you can easily generate surfaces to check shape complementarity.

This is achieved just like saving a .cif file, except you choose **[p] pdb-file** instead of **[i] cif-file**.

D. Creating a protein surface file to view within Moloc. Hard, but sometimes rewarding.

1) From the main menu select **[Srf] surfaces**.
2) Select **[q] generate surfaces** then CTRL-CLICK on the protein.
3) Select **[e] entries**, then CTRL-LEFT CLICK on some part of the protein entry.
4) Click **[x] exit**, and then choose a color, and the surface appears as a bunch of dots. Click **no** to discard that set of atoms.
5) Click **[d] display settings of surfaces…** and select your surface from the list that pops up.
6) Click **[t] tessellate surface** and the click on **accept** and then be patient… it will take up to 15 minutes.
7) Click **[m] display mode** and leave only "mesh" toggled to yes.

Cool, eh? The surface must be saved as a separate file of type .sfc using **display** menu… **[s] store entries**… **[q] sfc-file**. These files are not, as far as I can tell, readable by other programs.

Sometimes the surface confuses your viewpoint. It's appearance can be toggled on and off on the fly using **display** menu… **[m] map contour and surface visibilities**.

E. Adding functional groups instead of single atoms. Instead of adding carbon 6 times, why not add a benzene?

1) From the main menu, click **[dTp] topology changes**
2) Click **[s] add selected substituents**
3) Choose from the list… phenyl or a variety of cycloalkyls, then CTRL-LEFT CLICK on the atom to which you'd like to attach the new substituent.

F. Controlling the orientation of atoms and functional groups as you add them. What if you don't want the phenyl group that you've just added pointed in the way the computer chose?

For any new atom or substituent, you must execute a tricky mouse manoeuver:

1) CTRL-LEFT CLICK AND HOLD THE LEFT BUTTON (the new atom or substituent appears).
2) WHILE STILL HOLDING THE LEFT BUTTON, CLICK THE MIDDLE BUTTON to toggle through three possible orientations for the new substituent. When you let go of the left button, the choice is set in stone. (There is no "undo," but you can change the angle by driving the torsion angle manually.)

G. Checking and altering bond multiplicities and protonation states.

You may have noticed that MOLOC normally operates without any hydrogen atoms showing at all. The MAB force field that does the calculations has parameters that treat each heavy atom as a "united atom" that intrinsically includes the attached hydrogen atoms. Because of this, issues of protonation state and bond multiplicity are cleverly handled by simply defining how many "silent" hydrogen atoms are on each heavy atom. Check out the silent H-atom counts on any of your structures by:

1) **[dTp] topology changes**
2) **[h] check modify silent h's**

Each heavy atom now has a tiny number beside it that represents how many protons are attached. An aromatic carbon atom has 1, a methylene has 2, a methyl has 3, an amide nitrogen has 1, a primary amine can have 2 or 3 depending on whether or not it is protonated (charged). You should routinely check the silent h's of EVERY new inhibitor structure that you draw in order to avoid problems.

To modify the silent h's, simple CTRL-LEFT CLICK on an atom to increase the number of protons by one and CTRL-MIDDLE CLICK to decrease the number by one.

Bond multiplicities are automatically calculated based on how many hydrogen atoms are attached. So if you want to make a C-C bond into a C=C bond, you have to set the number of silent hydrogen atoms to the number appropriate for a double-bonded carbon.

H. Starting a new inhibitor by installing a completely new fragment.

Adding atoms and groups is normally done as an extension of an existing molecule. To add a new fragment "floating in space," you can first build your new atom or group off of an existing protein residue and then sever the link you just made. Presto, a new inhibitor fragment unattached to the protein and ready to be moved and rotated to get it into the desired position.

Example: installing a new benzene ring as the beginning of a new inhibitor structure:

1) From the main menu, select **[dTp] topology changes**
2) **[s] add selected substituents**… select phenyl
3) CTRL-LEFT CLICK on an atom near to the desired position of your benzene fragment
4) **[x] exit**
5) **[b] add and delete bonds**…
6) CTRL-MIDDLE CLICK on the bond between the benzene ring and the protein to delete it.
7) **[x] exit**

Now you have the fragment you want, but it's still part of the protein entry. You need to do a few more steps to make it into an entry of its own.

1) From the main menu select **[set]** define sets of atoms and then **[f] fragments**.
2) CTRL-LEFT CLICK on the benzene fragment
3) **[j] generate a new entry**… click beside "new entry" and then give it a name.

You now have two duplicate benzene fragments… one in a new set and one in the protein's set. You'll have to delete the latter.

1) **[.] display**
2) **[a] entry activity stati**
3) Toggle the new entry to blank (= invisible) and exit all the way back to main menu
4) **[dTp] topology changes**
5) **[d] delete set of atoms**… **[f] fragment**… and CTRL-LEFT CLICK on the fragment.
6) **[x] exit** and the highlighted atoms are deleted.
7) **[.] display**… **[a] entry activity stati** toggle your new entry back to **A**ctive.

Now you have a benzene fragment as a separate entry. You can color it, move it, and build off of it in order to create an entirely new inhibitor.