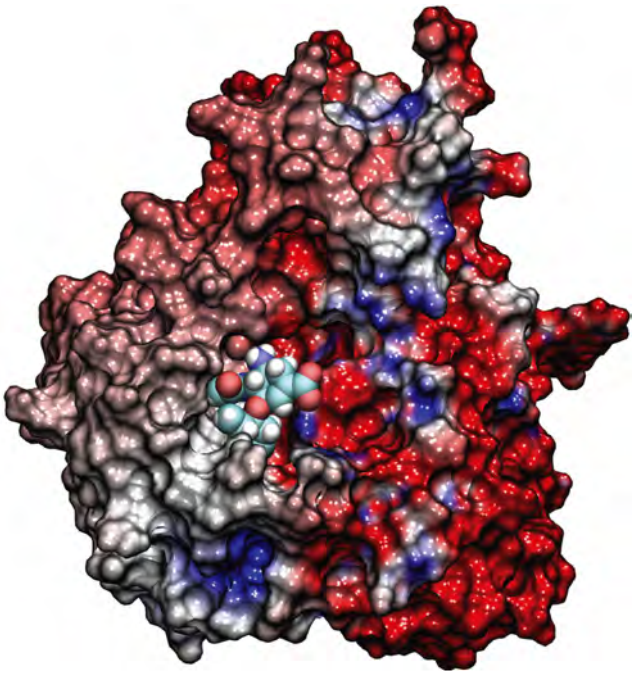
Nancy Université, Université Henri Poincaré
Centre National de la Recherche Scientifique
Équipe de dynamique des assemblages membranaires,

Centre National de la Recherche Scientifique
Laboratoire d'Ingénierie des Systèmes Macromoléculaires

University of Illinois at Urbana-Champaign
Beckman Institute for Advanced Science and Technology
Theoretical and Computational Biophysics Group

Molefacture: A tutorial to build and edit molecules



Johan Strümpfer
Peter Freddolino
Christophe Chipot

March 19, 2012

Please visit www.ks.uiuc.edu/Training/Tutorials/ to get the latest version of this tutorial, to obtain more tutorials like this one, or to join the `tutorial-l@ks.uiuc.edu` mailing list for additional help.

Abstract

`molefacture` is a VMD plugin that has been designed to facilitate the construction and parameterisation of small molecules. It additionally provides a simple interface to prepare structures and files for free energy perturbation calculations using NAMD. This tutorial serves as a primer for structure building, modification and parameterization using the `molefacture` plugin of VMD. Prior knowledge of NAMD and standard molecular dynamics simulations is assumed.

Contents

Introduction

The goal of this tutorial is to provide a guidance for:

- Building molecular structures
- Altering molecular structures

Completion of this tutorial requires:

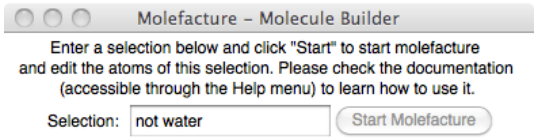
- Files contained in the archive `Molefacture_tutorial_files.zip` (see below)
- NAMD 2.8 or later (`http://ks.uiuc.edu/Research/namd`)
- VMD 1.9 or later (`http://ks.uiuc.edu/Research/vmd`)
- AmberTools 1.5 or later (`http://www.ambermd.org`)*

The advanced features of `molefacture` (geometry optimization and parameter determination) requires AmberTools. This can be obtained from the Amber website (`http://www.ambermd.org`) along with instructions for compiling AmberTools. Additional information for compiling AmberTools using OS X can be found at `http://amberonmac.blogspot.com` and compiling using Windows can be found at `http://ambermd.org/mswindows.html` .

∗ **NOTE:** The present version of AmberTools (1.5) does not contain the requisite features for `molefacture`. An alternate AmberTools source code containing the required features can be obtained from `http://www.ks.uiuc.edu/~johanstr/AmberTools-tcbg_molefmod-1.5.tar.bz2` and can be built following the same procedures as outlined in the links provided in the previous paragraph.

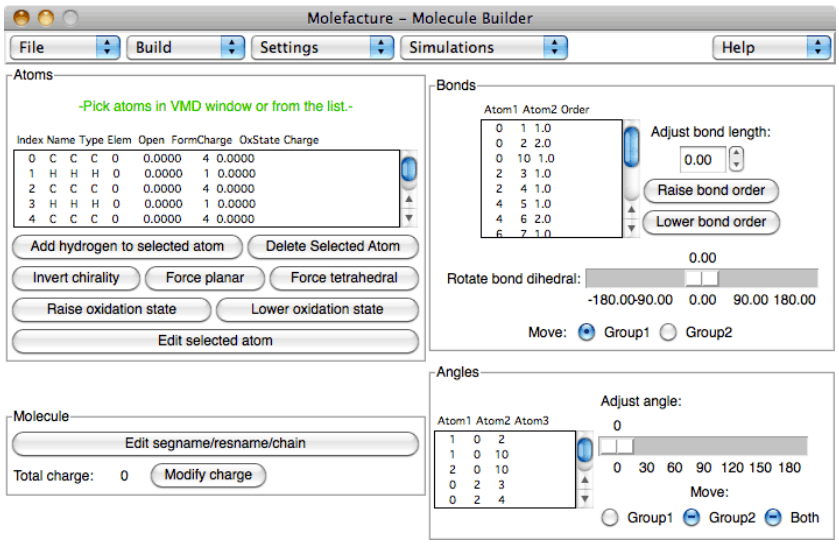
1. The `molefacture` interface

The `molefacture` plugin is found in the `Extensions->modeling` menu in VMD. When you open `molefacture` you will be presented with the start screen:



Here you can specify whether you want to edit an existing selection of atoms from the molecule in VMD currently step to `top` (e.g., by writing `not water` in the `selection` box to edit all non-water atoms), or to build a new molecule by leaving the `selection` box empty.

Clicking on the `start` button will load the main `molefacture` interface



The main interface contains 5 elements: the menubar across the top and four framed areas below the menubar: `Atoms`, `Bonds`, `Molecule` and `Angles`. In the `Atoms` frame, you are given a list of all the atoms currently being editing along with some information to go with each atom:

`Index`	The index of the atom
`Name`	The atom name
`Type`	The atom type
`Elem`	What element the atom represents
`Open`	The number unpaired electrons
`FormCharge`	The net integer charge on the atom
`OxState`	The oxidation state of the atom
`Charge`	The partial charge assigned to the atom

The various buttons in the `Atoms` frame can be used to modify the atom properties and will be exemplified in section **2.**. Atoms can be selected by either clicking on their entry in the list, or by clicking on the atom in the VMD `OpenGL` window (making sure that the "Pick" mouse mode is selected by pressing `p` on the keyboard). Atoms can be modified (element types, atom types, atom names) by clicking on the `Edit selected atoms` button after selecting an atom. Similarly, any selected atom can be deleted by clicking the `Delete Selected Atom` button. The only way to add new atoms is to select the atom to which the new atom will be bonded, and click `Add hydrogen to selected atom`.

The `Bonds` frame contains a list of bonds present in the structure. Bonds are selected either by clicking on two atoms with the mouse while holding `shift` (remember to make sure that the mouse is in "Pick" mode!), or by selecting the bond in the bond list. Bonds can be altered by raising or lowering the bond-order and adjusting the bond length. If the bond forms the center of a dihedral angle then the dihedral can be rotated by dragging the `Rotate bond dihedral` slider.

The `Angles` frame contains a list of angles as defined by the bonding in the structure. Note that angles cannot be added to the structure and only the angle terms defined by the bonding can be modified.

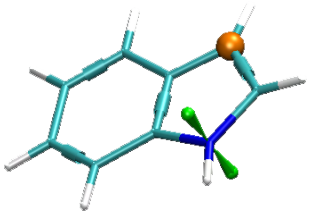
The final frame, called the `Molecule` frame, lists the total charge as calculated by the sum of the formal charges. The `Modify charge` button does not change any formal charges or assigned partial charges, but only changes what is communicated to the atom-typing program Antechamber and semi-empirical quantum mechanics minimizer SQM. The `Edit segname/resname/chain` button allows you to set the segname, resname and chain for the molecule. Note that this sets all the atoms in the structure to the names that you enter.

Finally, the molecule can be saved as either `.pdb`, `.mol2` or `.xbf` formats using the `File`→`Save` command (note: the filetype is determined from the extension you enter!). The topology file for the molecule can also be written out using `File`→`Write .top`.

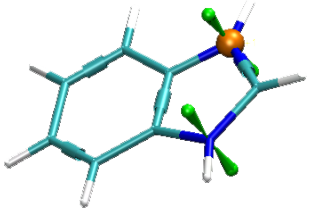
2. Building and editing structures

2.1. Building a molecular structure I

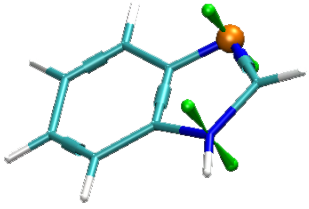
In this section, we will review the key features of what constitutes the core of `molefacture`, namely a molecular builder. The tools provided in `molefacture` are used to build a molecule. We shall build R-6-methoxy-2-((4-methoxy-3,5-dimethylpyridin-2-yl)-methylsulfinyl)-1H-benzo[d]imidazole, also known as omeprazole for short. To do so we shall start with indole as a base molecule (provided in `molefacture`) and iteratively transform it to omeprazole. This requires adding atoms and functional groups, changing bond orders, changing oxidation states and altering atom types.



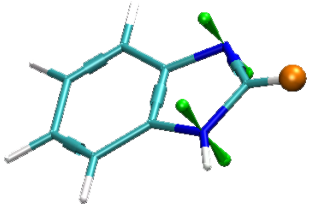
1. Open `molefacture` and load imidazole In the VMD `Main` window, click on `Extensions`→`modeling`→`Molefacture`. In the first dialogue that pops up, erase any text in the selection box and click the `Start Molefacture` button. The base molecule, imidazole, is loaded by clicking on `Build`→`New molecule from fragment`→`Indole`



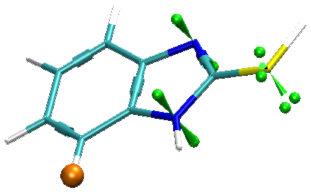
2. Transform carbon to nitrogen Select the 3rd carbon (as shown) by clicking on it (make sure you are in "pick" mode by clicking in the VMD OpenGL window and hitting the `p` button on the keyboard). Click the `Edit selected atom` button. Click on the `Element` drop down list and select `N`. Finally click `Apply` to transform carbon to nitrogen.



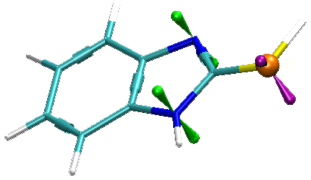
3. Delete the extra hydrogen Next we remove the extra hydrogen by selecting the atom and then click the `Delete selected atom` button.



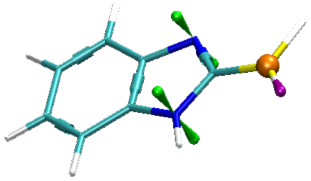
4. Optimize the structure To make sure we have a reasonable structure while building a molecule, we periodically minimize it using `SQM`. Do so by clicking on `Build→Geometry optimisation with SQM`. It is important to make sure that it has the correct total charge before doing so. Check the total charge of the molecule and if it is incorrect, fix it by correcting the bond orders.



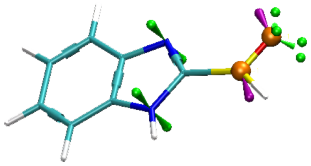
5. Add a thiol group Structures can be extended by replacing hydrogens with fragments of molecules that are found in the `Build` menu. We continue our construction by adding a thiol group to the imidazole ring. Select the hydrogen attached to the 2-carbon. Add the thiol group by clicking on `Build→Replace hydrogen with fragment→Thiol`.



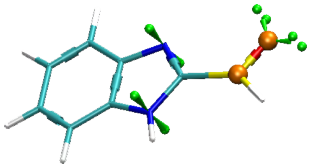
6. Raise the oxidation state of sulphur The sulphur atom in the target structure is in the +4 oxidation state, but the sulphur in the present thiol is in the +2 oxidation state. To correct this, select the sulphur and click `Raise oxidation state`. You can see that the lone pairs (green) have been replaced by empty orbitals (purple).



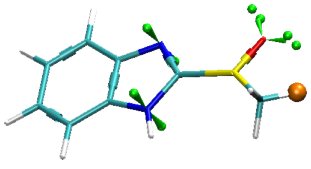
7. Add a hydrogen bound to sulphur The sulphur is double bonded to an oxygen. We construct this moeity by first adding a hydrogen: select the sulphur atom and click `Add hydrogen to selected atom`.



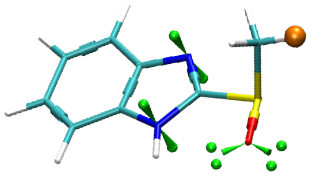
8. Transform hydrogen to oxygen. Transform any hydrogen that is attached to the sulphur atom to an oxygen using the same procedure as in step 2.



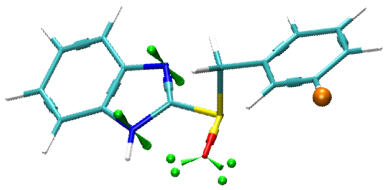
9. Increase the S–O bond order. The double bond between the oxygen and the sulphur is made as follows: First select both atoms by clicking on one then the other with the mouse while holding the shift key. Once both atoms are selected click the `Raise bond order` button.



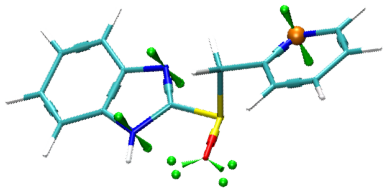
10. Add a methyl group. Use the same procedure as in step 5 to replace the remaining hydrogen on the sulphur atom with a methyl group.



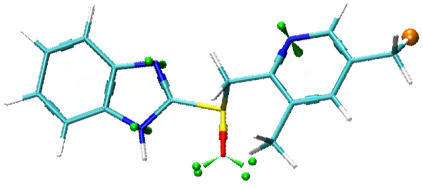
11. Optimize the structure. The next step will make a relatively large change, thus it is a good idea to optimize our structure before hand. Do so by clicking on `Build→Geometry optimisation with SQM`.



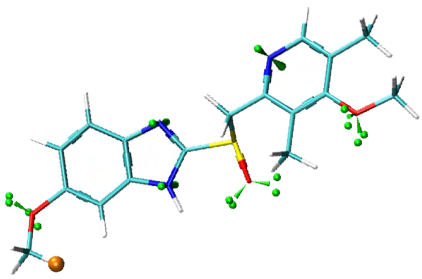
12. Add a phenyl ring. Replace the furthest hydrogen of the recently added methyl group with a phenyl ring by clicking on `Build→Replace hydrogen with fragment→Phenyl`. Optimize the structure again after the change.



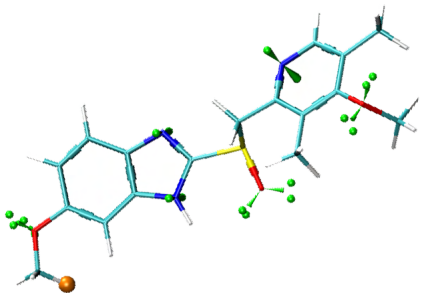
13. Change phenyl to pyridine. Next repeat steps 2 - 3 to replace a carbon of the recently added phenyl group with a nitrogen and to remove the extra hydrogen.



14. Add two methyl groups. Use `Build→Replace hydrogen with fragment→Methyl` to add methyl groups the appropriate carbons in the pyridine ring.



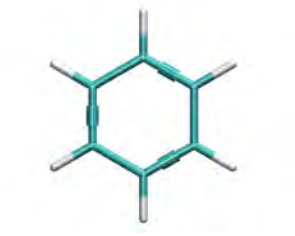
15. Add two methoxy groups. In the final step we combine the hydroxyl and methyl groups to make a methoxy group. Use `Build→Replace hydrogen with fragment→Hydroxyl` followed by `Build→Replace hydrogen with fragment→Methyl` to add the methoxy groups.



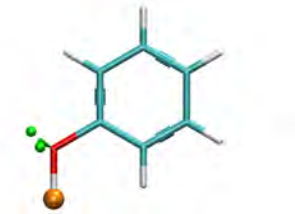
16. Optimize the structure. Finally we optimize the molecule again to get the structure of R-omeprazole.

This structure can be saved using the `File→Save` command. When saving a structure using `molefacture`, remember to add either `.pdb`, `.mol2` or `.xbgf` as the extension to the filename.

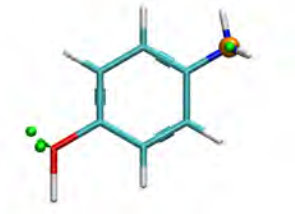
2.2. Building a molecular structure II



Choosing a new fragment. A glimpse at paracetamol suggests that the most sensible starting scaffold for the construction of the molecule is benzene. In the `Build` menu of `molefacture`, go to `New molecule from fragment` and select `Benzene`. Note the alternated simple and double bonds of the ring, corresponding to bond orders equal to 1 and 2.

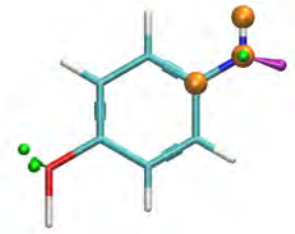


Editing the molecule. The next step consists in adding the functional groups characteristic of paracetamol. Starting with the hydroxyl moiety, pick one hydrogen atom of benzene and go to the `Replace hydrogen with fragment` submenu of the `Build` menu. Select `Hydroxyl` therein.

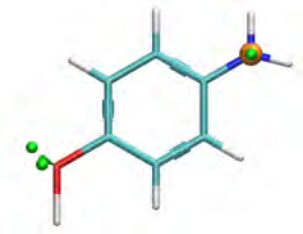


Pursuing with the functional groups of paracetamol, we will build the amide moiety in the *para* position with respect to the hydroxyl group. To do so, select the hydrogen atom in *para* of the —OH moiety and repeat the previous step, selecting the `Amino` option in the `Replace hydrogen with fragment` submenu.

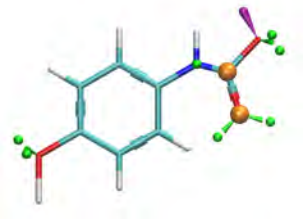
It should be noted that the —NH_2 added to the aromatic ring possesses an sp_3–like geometry, incompatible with the planar structure of the amide moiety. While we could proceed and let the energy minimizer handle incorrect geometries, we propose to delete one of the hydrogen atoms of the amino fragment.



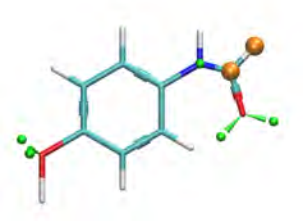
Let us then select the three adjacent atoms forming the C–N–H valence angle, and modify the latter by means of the `Adjust angle` slider in the lower right quadrant of the `molefacture` graphical user interface. It should be noted here that valence of the nitrogen atom is now incomplete and, therefore, ought to be corrected



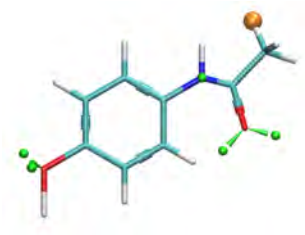
This step is easily completed by picking the nitrogen atom and clicking the `Add hydrogen to selected atom` button. Guaranteeing that the central nitrogen atom and its three neighbors be coplanar can be achieved by means of the `Force planar` option. The latter is recommended whenever an sp_3 element needs to be altered into its sp_2 form.



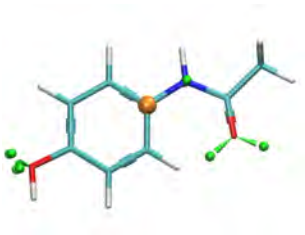
To obtain an amide group, the quickest route consists in replacing the newly created hydrogen atom by a carboxyl moiety. To do so, select the `Carboxylate` option in the `Replace hydrogen with fragment` submenu. Evidently, the —O^- fragment is superfluous and ought to be transformed.



Editing a chemical element. Let us select the singly bonded oxygen atom and invoke the `Edit selected atom` option of the graphical user interface. Let us modify therein the chemical element to hydrogen, because replacement with fragment can only be achieved when the selected atom is a hydrogen.



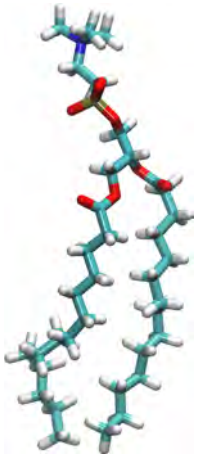
Repeat the penultimate step and modify the previously created hydrogen atom by a —CH_3 group. Towards this end, select the `Methyl` option in the `Replace hydrogen with fragment` submenu. Construction of the paracetamol is now complete. The molecule is ready for energy minimization.



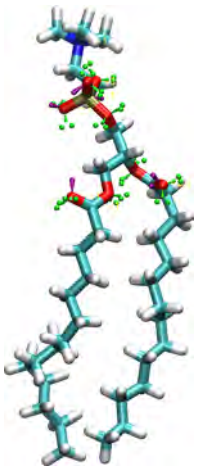
Select in the `Build` menu the `Geometry optimization with SQM` option. This will prompt the quantum-mechanical program SQM to start an energy-minimization at the semi-empirical level defined in the `SQM Settings` submenu of the `Settings` menu. Note the changes in the molecular geometry, in particular the rotation of the methyl group.

2.3. Editing of molecular structures

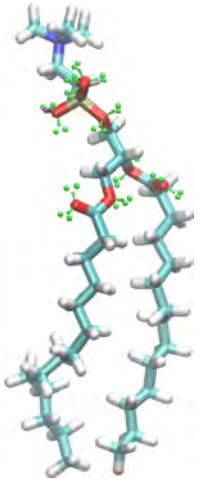
In this section, an overview of the basic editing features of `molefacture` will be provided. To illustrate these features, use will be made of the example of 1,2-dimyristoyl-sn-glycero-3-phosphocholine, better known as dimyristoylphosphatidylcholine or DMPC, which will be altered into 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine, often referred to as palmitoyloleoylphosphatidylcholine or POPC. Transforming between compounds is a common exercise in computer-assisted molecular modeling and is generally preferable over constructing the desired species from scratch. In the proposed example, the alteration consists in extending by two methylene groups the sn1 chain and by four methylene groups the sn2 chain, whilst creating in the middle of the latter an insaturation.



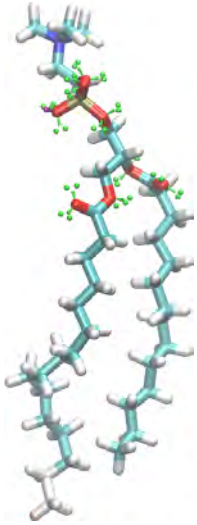
Loading the molecule and invoking `molefacture`. The starting point of the transformation is DMPC, the coordinates of which are given in `dmpc.pdb` in the `02_dmpc` directory. Once the molecule loaded in VMD, invoke `molefacture` to edit the lipid molecule. Upon initialisation a dialog window will open that allows you to specify which part of a molecule you wish to edit. Since we will be modifying the whole DMPC structure in `molefacture`, enter `all` for the selection of atoms and press `open`.



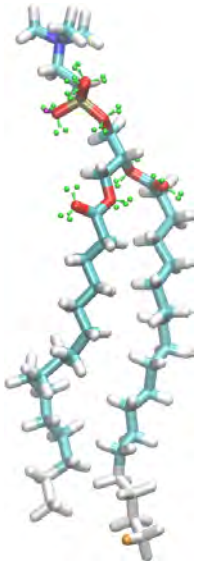
Updating the bond order. When invoked over the entire molecule, `molefacture` will display attributes for a number of atom types, e.g. lone pairs on oxygen atoms. Since PDB files do not include any information regarding bond orders, erroneous bonding is to be expected, reflected in the total charge of the molecule.



Here, the C—O moiety of the two ester linkages ought to contain a double bond. In addition, one of the P—O bonds in the phosphate group should be a double bond. Note that the second P—O bond contributes to the net -1 charge of the fragment, while the central nitrogen atom of the choline group bears a net $+1$ charge, resulting in a zwitterionic species.

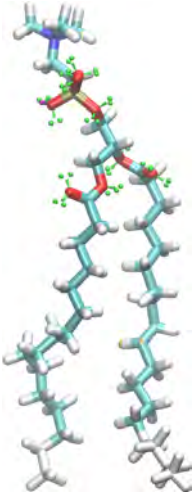


Editing the molecule. Select the last hydrogen atom of the sn1 chain and replace it by a methyl group. Repeat this operation to tailor the aliphatic chain to the correct length, namely myristoyl (C14), i.e. to palmitoyl (C16). Orientation of this fragment can be fine tuned by selecting adjacent carbon atoms and modifying the dihedral angle.



Select the last hydrogen atom of the sn2 chain and replace it by a methyl group. Repeat this operation to tailor the aliphatic chain to the correct length, namely myristoyl (C14), i.e. to oleyl (C18). Orientation of this fragment can be fine tuned by selecting adjacent carbon atoms and modifying the dihedral angle.

Select the two central methylene groups of the oleyl chain, i.e. the ninth and tenth carbon atoms starting from the ester functional group, and delete one hydrogen atom per methylene group. Correct the position of the remaining hydrogen atoms by modifying the relevant valence angle. Increase the bond order to form the central double bond of the sn2 chain.

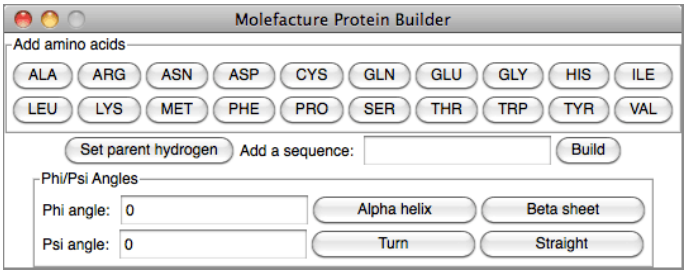


Select the two carbon atoms forming the ethylene fragment of the oleyl chain. Modify the torsional angle so that the chain adopts an E or *entgegen* conformation. The resulting compound is POPC. Its structure can be compared to that supplied in the PDB file `popc.pdb`. Note that additional trans–gauche defects may be added by modifying the relevant torsions.

It ought to be noted that the new fragment introduced in the original molecule are appended to the list of atoms, thereby modifying the expected order in which they should appear. In other words, extension of both the sn1 and the sn2 chains will result in a subset of atoms added at the end of the DMPC in a non-sequential order. Furthermore, newly introduced atoms will be devoid of a type. Their description is limited to the chemical element.

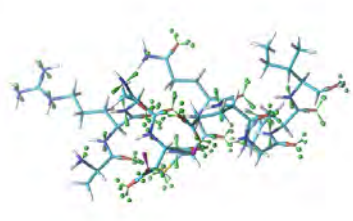
2.4. Building peptides

To build small peptides in `molefacture` use the `Protein Builder`, found in the `Build` menu.

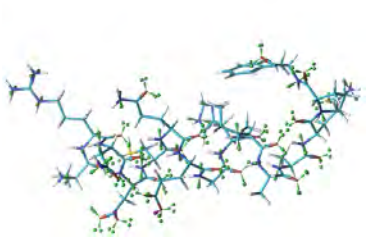


There are two ways to build small peptides, either by clicking on each residue in turn, or by entering the protein sequence using one letter codes in the `Add a sequence` box. In this section we will use both methods to construct an alpha connected, via a turn, to a 3-10 helix.

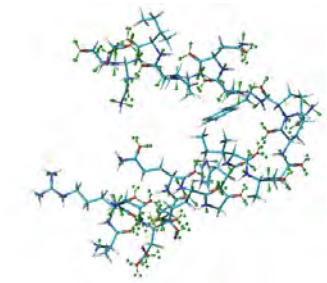
Start the `Protein Builder`. Open `molefacture` with an empty selection or click `File`→`New` in the `molefacture` window to start editing a new molecule. Then click on `Build`→`Protein Builder`.



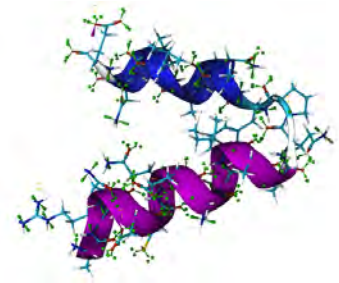
α helix. To make sure to build an α helix, first click the `Alpha helix` button. Next click on ten different residues, here we clicked on all the residues in the top row: `ALA-ARG-ASN-ASP-CYS-GLN-GLY-GLU-HIS-ILE`.



Inter-helix turn. Next, enter `-60` and `-30` for the Phi and Psi entries, respectively, and click on `ALA`. Next, click on the `Straight` button, followed by `GLN` and then `PRO`. Finally click on on the `Turn` button, followed by `PHE`.



3-10 helix. To build the 3-10 helix, enter -49 and -26 for Phi and Psi angles. Next enter `ANALINE` into the `Add a sequence` box. Then finally click on `build` to construct the helix.



Save. You can now close the `Protein Builder` and save your molecule from `molefacture`. Note that `CHARMM` atom types and residue names are assigned and standard protonation states are chosen for the residues and termini. There is no need to save a topology file when using the `Protein Builder` in `molefacture` as all the residues are defined in the `CHARMM` 22 topology file.

.

Acknowledgements

Development of this tutorial was supported by the National Institutes of Health (P41-RR005969 - Resource for Macromolecular Modeling and Bioinformatics).