

Preparing Molecules for DOCKing

The purpose of this document is to describe the steps required to prepare molecules as input for a DOCK run that attempts to predict the orientation of a ligand in an active site. This tutorial uses L-Arabinose-Binding Protein Complex bound to L-Arabinose (pdb code 1ABE) as an example system. However, these techniques should be transferable to any protein-ligand system. All files discussed in this tutorial are linked at the appropriate places and also can be found in the "demo/1_struct" folder distributed with DOCK 5.2 and later.

A variety of packaged programs and script libraries are available to perform the necessary modifications to the structure files described below (see http://dock.compbio.ucsf.edu/DOCK_Links/index.htm for more information). We have not compared the performance or accuracy of any of these programs and, therefore, do not recommend any of the packages over any other.

STEP 1: Examine the pdb file

The first step in any docking project is selecting the file that will be used for the structure of the target. In this case, the file we will be using is [1ABE.pdb](#). A visualization of this file can be seen below:

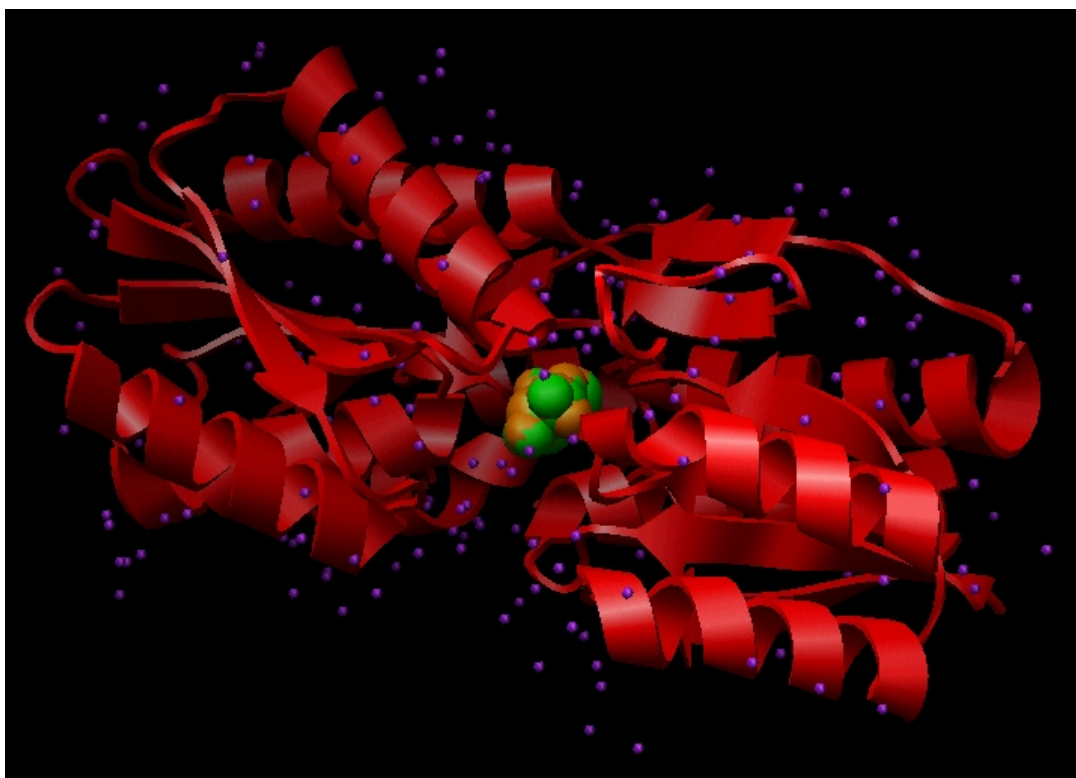


Image generated using Chimera (<http://www.cgl.ucsf.edu/chimera>)

This file contains cartesian coordinates for the protein (red ribbon), crystallographic waters (purple), and two conformations of the ligand L-arabinose (green and orange). Each of these components must be dealt with separately before DOCK can be used.

STEP 2: Prepare the ligand file

Here, we will only prepare L-arabinose A included in the pdb file for simplicity. Selection of conformation A over conformation B in this case was arbitrary.

- a) Create file containing only ligand atoms and connectivity ([lig.pdb](#)).
- b) Add hydrogens to the ligand ([lig_H.pdb](#)). It is critical that the atom type, bond order, and protonation state of the ligand are checked at this point for accuracy. Mistakes in any of these values will effect later charge calculations and docking results.
- c) Calculate charges for the ligand and save in mol2 format ([lig_charged.mol2](#)).

STEP 3: Prepare the receptor file

- a) Create new file containing receptor and all other molecules thought to be required for binding ([rec.pdb](#)). In this case, the ligand interacts with the protein only. However, in other structures, items you should consider including might be ions, water molecules critical for binding, and small molecule cofactors. Inclusion of these components are highly system dependent and decisions should be made on a case by case basis. Importantly, though, if other molecules are included at this point, they should be added to the rec.pdb file and be considered part of the receptor for the rest of the tutorial. All other components of the pdb file should be removed.
- b) Modify residues as necessary ([rec_fixed.pdb](#)). Almost every protein target needs some level of modeling before it can be used in docking. Once again, choices of what to model and what to ignore are highly system dependent and should be considered on a case by case basis. Below is a list of issues our lab has commonly come across:

- Residues listed in protein sequence that are incomplete or missing**
- Protonation states of amino acid (ie histidine) that vary based on chemical environment**

-The biological unit of the protein is a dimer or trimer, even though only the monomer has been included in the structure file

-Unnatural amino acids have been incorporated into the receptor structure

In this example, the terminal ends of the protein were highly flexible and did not produce any crystal density. In the structure file, only the backbone and alpha-carbon atoms were included for these residues. We have thus mutated both Asn 2 and Lys 306 to Gly.

c) Add hydrogens, calculate charges, and save in mol2 format ([rec_charged.mol2](#)).