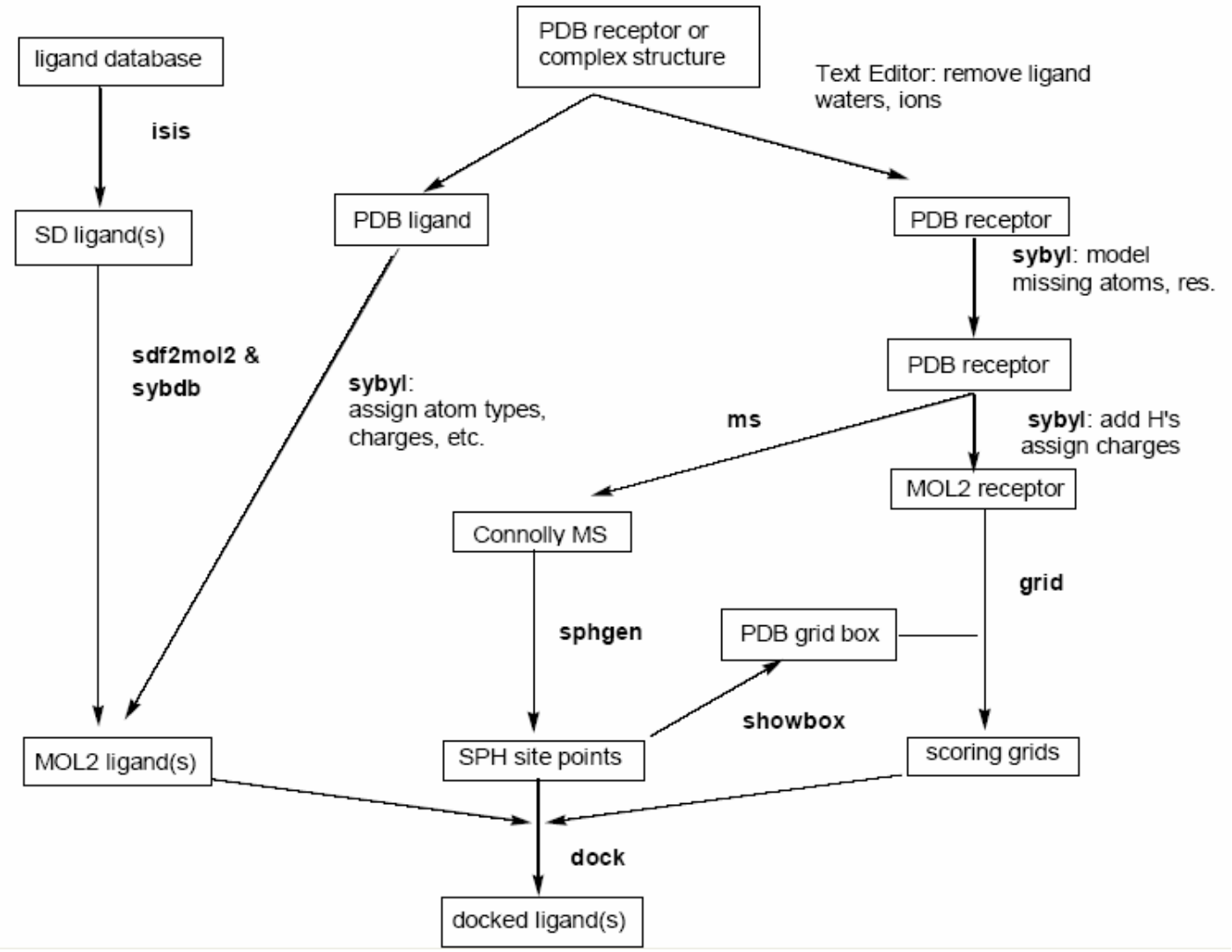
Introductory Tutorial for the UCSF DOCK Program (v. 6.0) Linux version

Authors:

Vladyslav Kholodovych, Ph.D.
John E. Kerrigan, Ph.D.

University of Medicine & Dentistry of New Jersey
Robert Wood Johnson Medical School
675 Hoes Lane
Piscataway, NJ 08854 U.S.A.
(732) 235-3229 phone
(732) 235-4473 phone
(732) 235-5252 fax
kholodvl@umdnj.edu
kerrigje@umdnj.edu
http://www2.umdnj.edu/~kholodvl
http://www2.umdnj.edu/~kerrigje

A Brief Synopsis of the DOCK Procedure.



Docking Flow Chart (adapted from the DOCK 4.0 User Guide)

Organization of Workspace: Create a project directory (use the unix *mkdir* command). Use this directory as your working directory.

mkdir dock6
cd dock6

Docking in DOCK is divided into four stages:

STAGE 1. Ligand Preparation
STAGE 2. Site Characterization
STAGE 3. Scoring Grid Calculation
STAGE 4. Docking

How DOCK works in a nutshell.

First, you begin with an x-ray crystal structure of a drug/receptor complex. The active site is identified or defined *a priori*. Points within this site known as "spheres" are used to define the volume or space within the active site pocket where the drug binds. The purpose of the spheres is to generate an unbiased grid of sphere centers that reflects the actual shape of the active site (i.e. the protein/macromolecule dictates the shape of the pocket; not the drug) using the grid program. [1, 2] Sphere centers are matched with ligand (drug) atoms to generate orientations of the ligand in the active site within the program DOCK. [1, 3] The orientation of the ligand is scored using a shape-scoring function and/or an energy function (e.g. $E_{bind} = E_{vdw} + E_{elec}$). The shape score is an empirical van der Waals attractive energy. As a final step, the orientation may be energy minimized using a rigid-body simplex minimization. [4]

Research Problem: In this tutorial you will study the docking of an inhibitor of factor Xa, an enzyme that is an important player in the blood coagulation cascade. Inhibitors of factor Xa are potential anti-coagulant drugs. The X-ray crystal structure (1EZQ) of human factor Xa complexed with **RPR128515** (3-[(3'-aminomethyl-biphenyl-4-carbonyl)-amino]-2-(3- carbamimidoyl-benzyl)-butyric acid methyl ester) will be used as a modeling template [5].

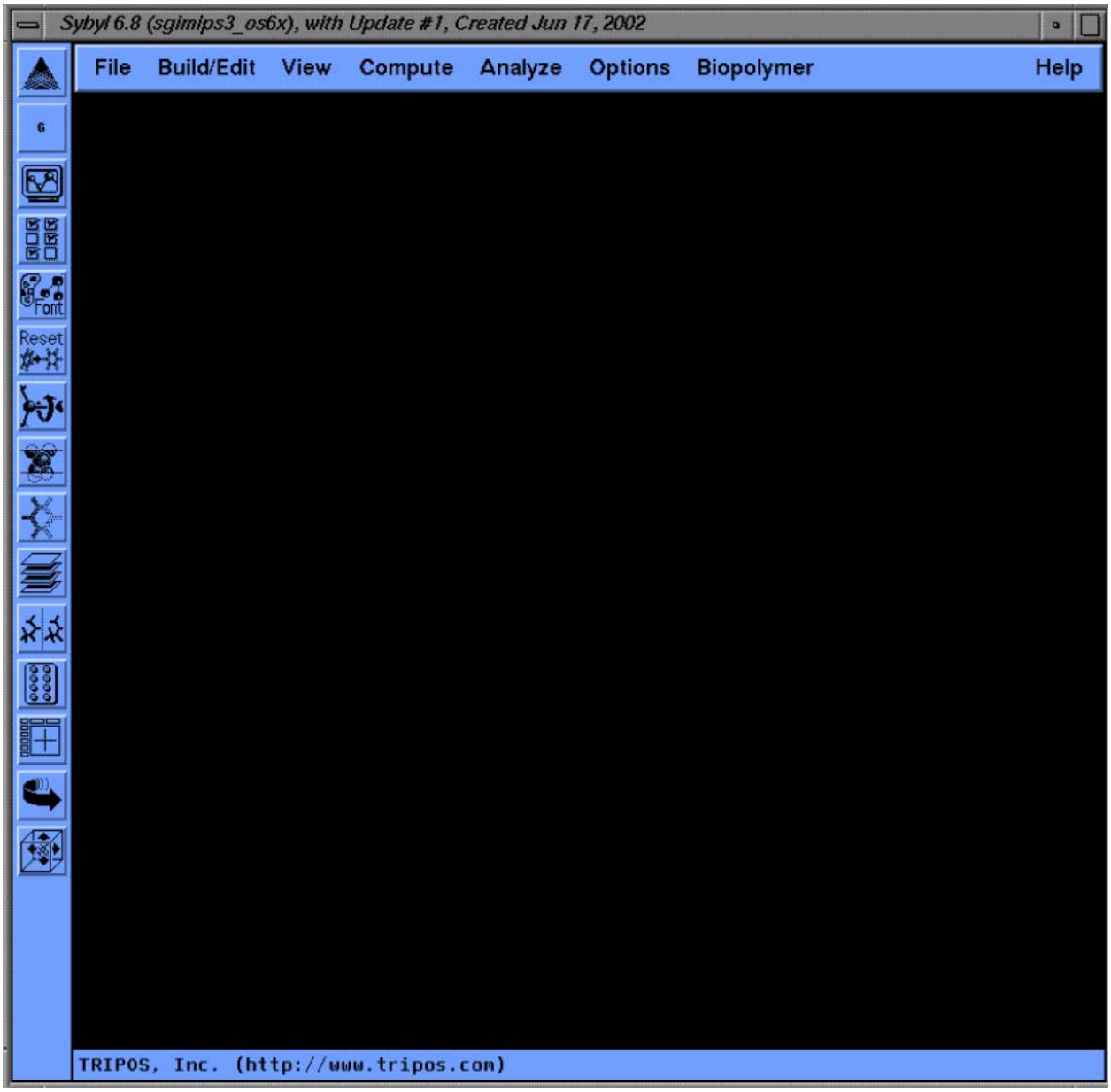
Make your project directory with the appropriate subdirectories (see previous page).

IMPORTANT!!!!

BEFORE START THE DOCK TUTORIAL, REMEMBER THAT WHEN YOU OPEN PROTEIN FILES IN SYBYL YOU MUST **NOT** CENTER MOLECULES. INSTEAD USE CENTER VIEW OPTION. ALL ORIGINAL COORDINATES SHOULD BE PRESERVED FOR FURTHER VIZUALIZATION AND DOCKING ANALYSIS!

On STAGE 1 and STAGE 2 you will save files in different formats, mol2 and pdb.
Giving name of file *.mol2 or *.pdb is not enough to save it in proper format. You should check a proper format from pull down menu in Save window.

STAGE 1. Prepare the ligand. Build the compound in Sybyl (Tripos) using the x-ray structure coordinates as a template and save to your project directory. To start Sybyl, type sybyl7.2 <enter>. The following screen appears…



Use “FTP connection” to get the PDB file 1EZQ.
File > Retrieve PDB > From RCSB…

Enter 1EZQ under PDB code: Click OK. In the Sybyl command window there will be a prompt asking for a password. Just hit Return or Enter key. The command line will tell you that the transfer was successful! In the new appeared window hit Cancel to close it. The file, **pdb1ezq.ent** will have been saved to your directory! Open this file using

File > Read > Files: **pdb1ezq.ent** (select with mouse) > Click OK > *Yes or No* “Center the Molecule”: Here is a new very useful option added to Sybyl 7.x ! It allows to center only a screen view remaining the actual coordinates of the molecule unaffected. Select **Center View** > OK

If needed, use the middle mouse button to translate the structure from the top of your display down to the center of your display.

You will notice that this structure is a hetero dimer. It has two chains A and B. However only chain A contains a ligand binding domain and a chain B is not involved in the ligand interactions. Thus we delete chain B and continue to work with a chain A - ligand complex.

Delete Chain B

Build/Edit > Delete > Atom …
Atom Expression > Change Atoms: to Monomers:
Select Chain B with the mouse by double clicking on its sequence.
Click OK

Delete the WATER Set

Build/Edit > Delete > Atom …
Atom Expression > Sets: Select WATER with the mouse > Click OK then Click OK again

Delete the calcium ion

Build/Edit > Delete > Atom …
Atom Expression > Substructures > Select the A/CA1 > Click OK then
Click OK again

Fix end groups on the protein.

Biopolymer > Prepare Structure > Fix End Groups… > *Molecule Area*: M1 > *Option* > CHARGED
Click OK

Whew! Now you should have one enzyme-inhibitor complex! Save it for further post-docking analysis:
File > Save As … > Save Molecule > File: **complex.pdb** Format: PDB Click Save

In our next step, we will extract the ligand (drug) to a new molecule area. We will delete the drug from the complex and save the enzyme as a PDB file for use later.

Extract the drug into molecule area **M2**.

Build/Edit > Extract…
Atom Expression > Substructures > A/RPR1> Click OK then
Click OK again

Select M2 in the molecule area dialog box.

Click OK

Now remove the drug from M1!

Build/Edit > Delete > Substructure
Substructure Expression > Substructures > A/RPR1 > Click OK, then
Click OK again.

The drug is now in molecule area M2.

Name the enzyme and save the PDB file of the enzyme.

Build/Edit > Name Molecule …
Molecule Area > M1: Click OK > *Name Molecule* > Factor Xa Click OK

File > Save As … > Save Molecule > (Make sure that only m1: is selected!) File: fxa_nl.pdb
Format: PDB Click Save

Remove the enzyme from display area M1:

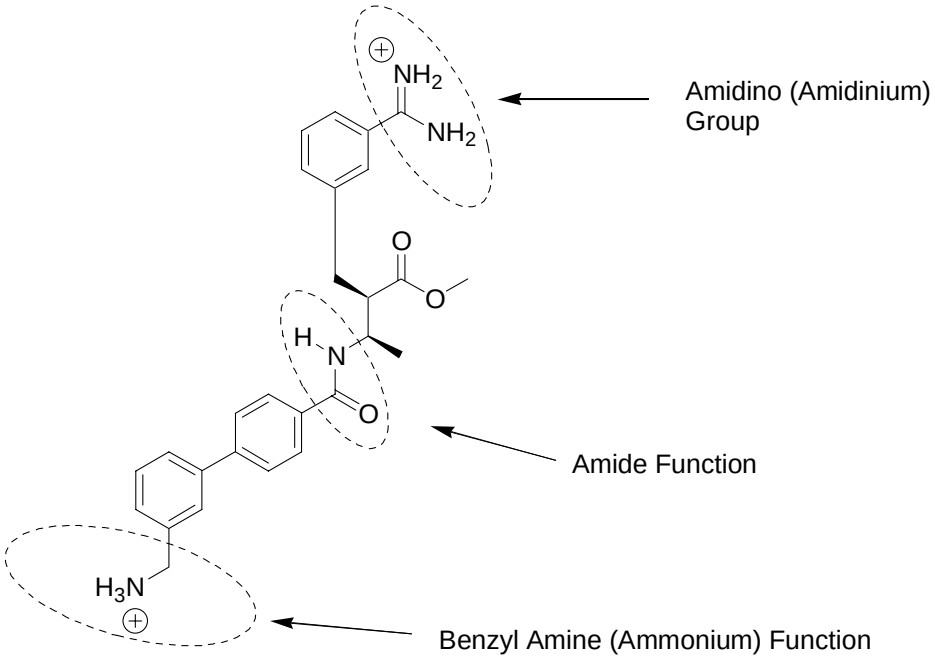
Build/Edit > Zap (Delete) Molecule > Molecule Expression > Select M1: click OK

Zoom into the drug using the middle and right mouse buttons (press simultaneously) and use the middle mouse button to translate (center). Color by atom type.

View > Color > By Atom Type

Set the Depth Cue to zero using the depth cue button on the left hand menu. Label atoms by type using the checkmark button on the left hand menu (your drug is in display area D2)

There are several atom types we must fix before proceeding.



Drug Structure Outlining Important functional groups.

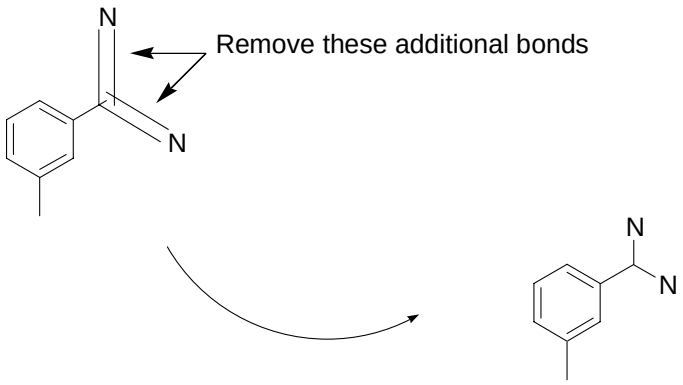
Special atom type designations.

You must modify the drug so that it appears as in the figure above. The bonds in the amidino group must be changed to single bonds (use REMOVE_BOND in Sketch) and the atom types must be changed to properly represent the amidino functional group properly. There are a few other corrections that need to be addressed.

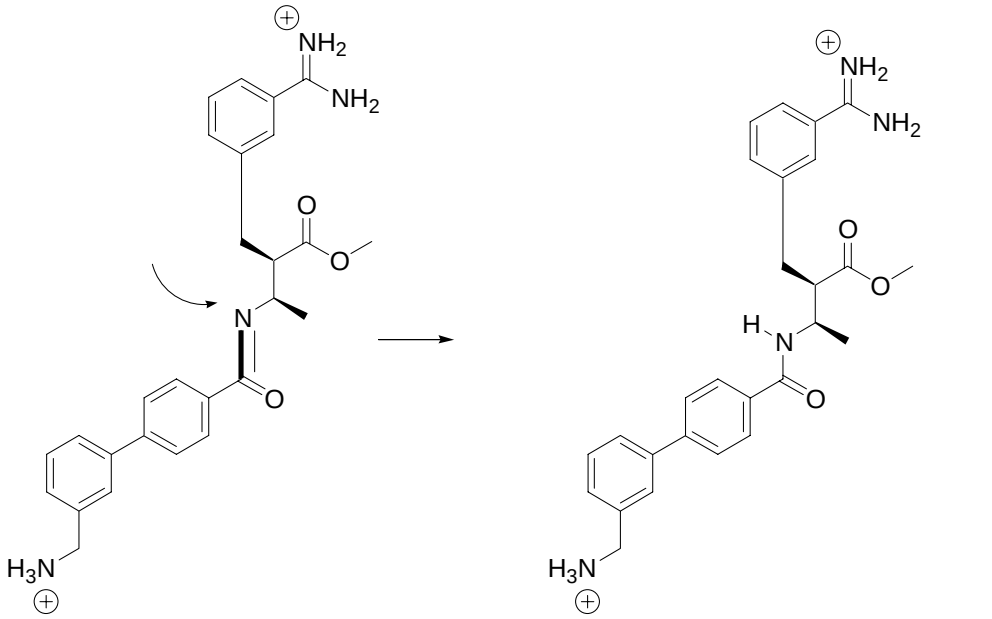
Use the Sketch utility in Sybyl to edit the molecule.
Build/Edit > Sketch Molecule… > *Molecule Area* > Select M2:<no name> then Click OK

The following dialogs appear:

Use REMOVE_BOND to change the double bonds to single bonds in the amidino group. Click on the each individual atom that makes up each bond (there will only be two per bond!).



In addition, remove the additional bond in the amide functionality.

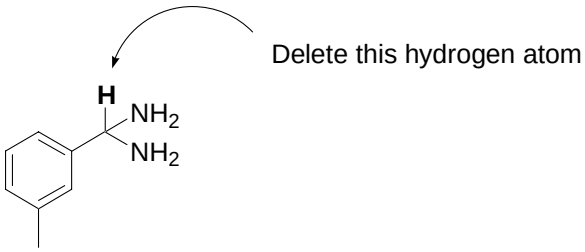


We will add hydrogens next

Use ADDH to add the hydrogen atoms.

IMPORTANT!
Before finishing go to TAILOR > CLEAN_UP Click OK and select NONE. Click OK, then Click END to get out of the menu. Click on EXIT.
Doing no clean up you preserve exactly the same conformation as in a crystal structure.

Delete the unnecessary hydrogen atom on the carbon of the amidino group.
Build/Edit > Delete > Atom
Select the atom individually with the mouse. Click OK.



Click on the label box (the button on the left-hand side with the checked boxes) and next to D2 under Atom Labels click on Type from the little pop-up menu.

Now change the atom types of the carbon atom of the amidino group from C.3 to C.cat.
Build/Edit > Modify > Atom > ONLY_TYPE Click OK
Atom Expression > Select atom using your mouse, Click OK
Option > C.cat Click OK
Click Ok again and accept the default bond types (i.e. click Ok a few more times)

Now change the nitrogen atom types on the amidino function from N.3 to N.pl3 using the same procedure.

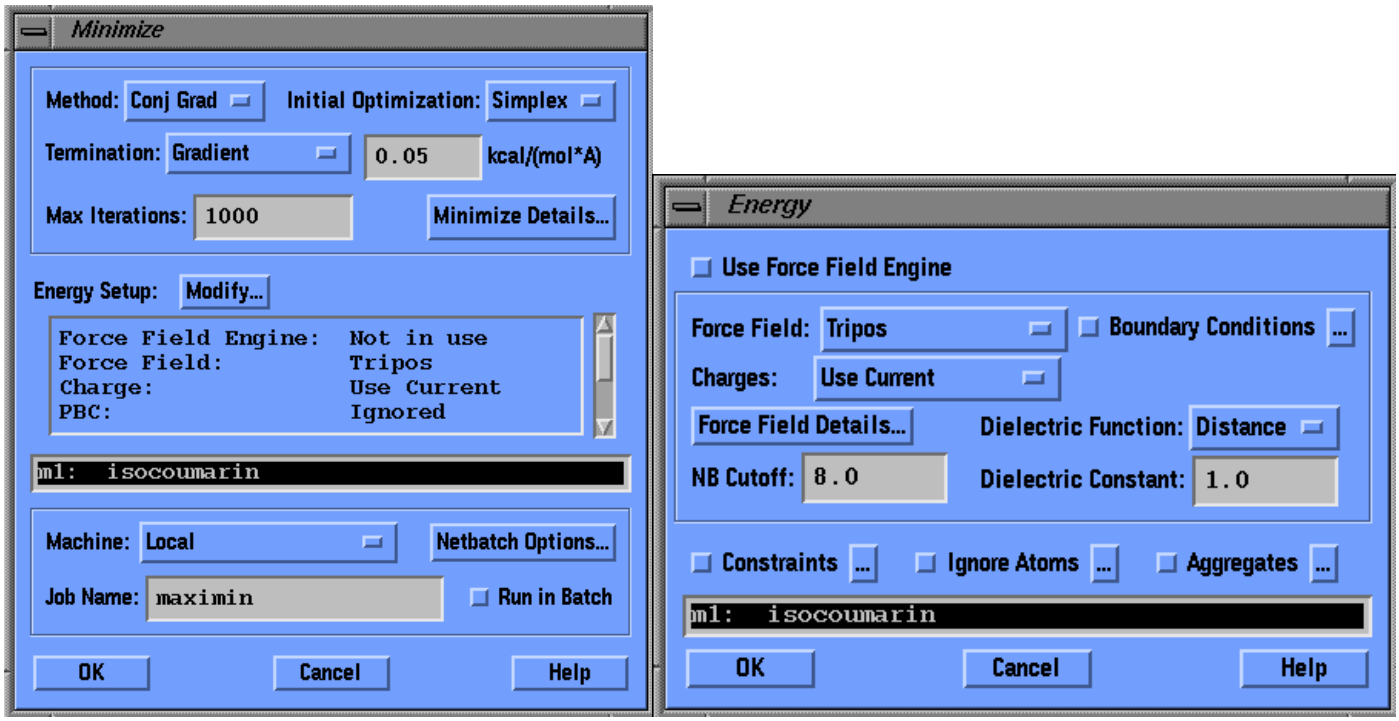
Last but not least change the benzyl amine atom type from N.3 to N.4 (see the initial figure for help)

Add hydrogen atoms to complete the structure.
Build/Edit > Add > Hydrogens
Assign charges
Compute > Charges > Gasteiger-Huckel

Click on “No” when the dialog asks you if you want to change formal charges.

Save ligand molecule at this step as rpr_xray.mol2. You will use this structure for comparison with docking results and for RMSD calculations.

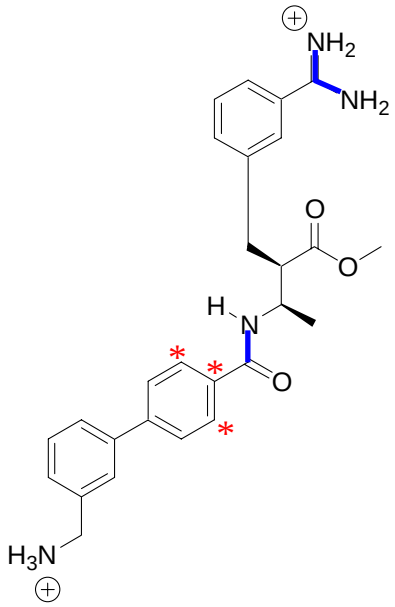
Perform the minimization using the settings indicated below.



Before you save the ligand, you must declare the Anchor atoms as a static set and declare the amide bonds as RIGID.

The ANCHOR set
Build/Edit > Define > Static Set … > *Option* > ATOM Click OK

In *Atom Expression*, select the following asterisked atoms …



Name this set ANCHOR.

The RIGID set
Build/Edit > Define > Static Set … > *Option* > BOND Click OK
In *Bond Expression*, select the highlighted blue in the figure above amide bonds …
Name this set RIGID.

Give the molecule a name before saving it.
Build/Edit > Name Molecule… > RPR Click Ok

Save the ligand as **rpr.mol2**.

STAGE 2. Site Characterization.

Remove the ligand. Build/Edit > Zap (Delete) Molecule. Open the PDB file you saved earlier, **fxa_nl.pdb**, using

File > Read > Files: **fxa_nl.pdb** (select with mouse) > Click OK > *Center View*, *Yes or No* "Center the Molecule" > Click on Center View

Use the middle mouse button to translate the structure from the top of your display down to the center of your display.

Now lets prepare the MOL2 receptor file for the grid calculation. Open the protein dictionary.

Biopolymer > Dictionary and Database Admin > Open Dictionary …

Select Protein, Click OK.

Now add hydrogen and charges:
Biopolymer > Prepare Structure > Add Hydrogens …

From pull-down menu select ALL; Water: H-bonds > Click OK

Biopolymer > Prepare Structure > Load Charges…

Biopolymer: Amber7FF02
Ligand: Gasteiger-Huckel
Uncheck Water box > Click OK

In the Warning window click YES.

Energy minimize the hydrogen atom positions only!

Compute > Minimize

Method: Powell
Initial Optimization : None
Termination: Gradient 0.01 kcal/mol•Å

Max Iterations: 1000

Energy Setup Modify…

Force Field: Tripos
Charges: Use Current
Check the Aggregates Box and Click on the ellipsis … (A dialog appears)
Click on New (An *Atom Expression* Box appears)
Click on Atom Types and select H on the right. Click OK.
Click on Invert button in the Atom Expression Box, then Click on OK.
Call the aggregate NON_HYD. Click OK
Comment String "Non-hydrogen atoms", Click OK

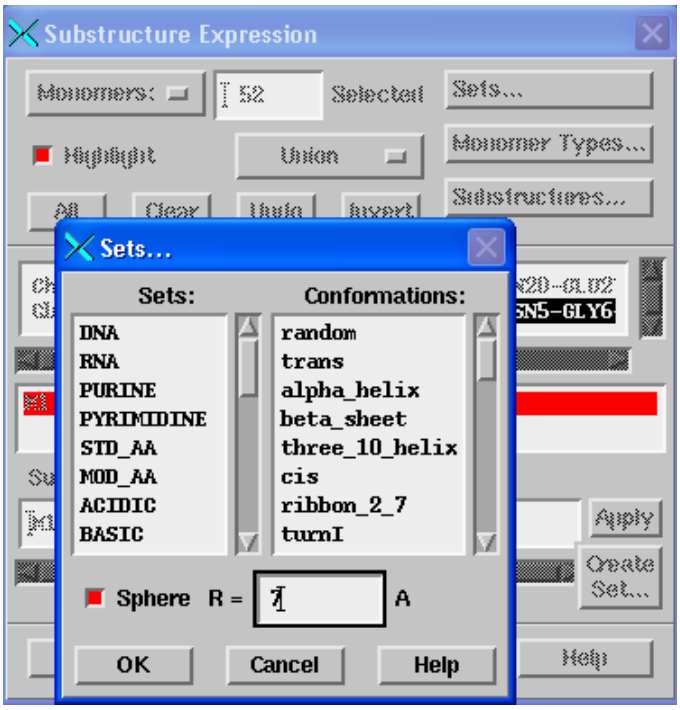
Click OK several times to get out of the dialogs and start the minimization.

Save the file as **fxa_grd.mol2** once the minimization is finished and remove (ZAP) it from display.

Make the exclude.pdb file! We need this file to exclude those residues that are remote from the binding site.

In Sybyl, open **complex.pdb**.
File > Read > Select complex.pdb > OK > Center View

Build/Edit > Delete > Substructures
Click on Substructures in the *Substructures Expression* dialog and select A/RPR1 > Click OK
You will see the ligand highlighted in the molecular view. Before proceed with the deletion click on button Sets…
In the new window select sphere radius is 7Å – it's a distance from the ligand, i.e. the neighboring atoms involved in the ligand-enzyme interactions.



Click OK then Click OK again.
Save the file as a PDB file and name it **exclude.pdb**.

Now, Exit Sybyl by typing "quit" at the Sybyl prompt.

At this point you should have 7 files:

rpr.mol2 - your optimized ligand/drug
rpr_xray.mol2 – extracted X-ray conformation of the ligand
pdb1ezq.ent - protein-ligand complex from PDB
fxa_nl.pdb - your protein without water and ligand
fxa_grd.mol2 - minimized protein structure
complex.pdb – monomer-ligand complex
exclude.pdb- protein with excluded binding site

Run the **dms** program and create the molecular surface data for the sphere calculation. Type "man dms" to obtain more information about the dms program and its run options.

dms fxa_nl.pdb **–a –n –w** 1.4 **–v –o** fxa_nl.ms

Next, we must run sphgen. In order to do so, we must have a file known as INSPH as input for the sphgen program. The contents of INSPH are as follows (Only use the information between the hashed lines!).

```
-------------------------------------------
fxa_nl.ms
R
X
0.0
4.0
1.4
fxa_nl.sph
-------------------------------------
```

Run sphgen simply by typing “sphgen” in the terminal window.

sphgen

Next we will use the coordinates of the ligand from the crystal structure to select the relevant spheres for the grid and docking computations. For this we will use a program called “**sphere_selector**”. The command line format is: “**sphere_selector** *file.sph ligand.mol2 #.#*” where #.# is the number of angstroms out from the ligand that you want to include spheres.

sphere_selector fxa_nl.sph rpr.mol2 6.0

The output from this operation is always a file named “selected_spheres.sph”. Use the **showsphere** program to make a pdb file of the selected spheres.

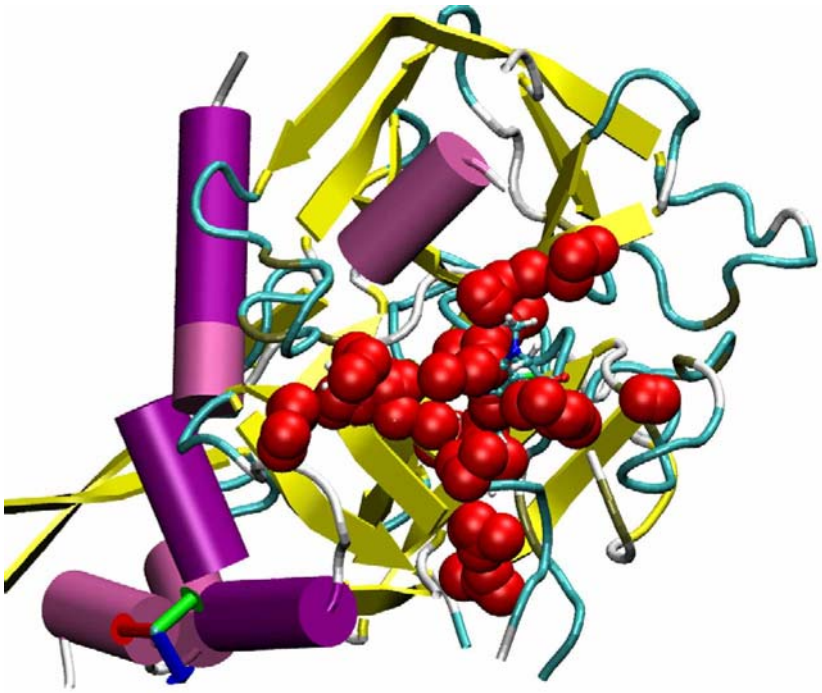
Type **showsphere** <enter>

Name of the sphere cluster file: selected_spheres.sph

Cluster number to process: 1

Generate surfaces? N

Name for output pbd file: sel_sph.pdb



Spheres selected by the sphere_selector program in red, as shown in VMD.

Next step is a creation of a boundary box in which we will dock our molecules. We want to keep it tight as close as possible to the original ligand binding groove.

Run showbox.
Type **showbox** <enter>.

automatically construct box to enclose spheres [Y/N] ?
Y
extra margin to also be enclosed (angstroms)?
2
sphere file-
selected_spheres.sph
cluster number
1
output filename?
site_box.pdb

STAGE 3. Scoring Grid Calculation

Download grid.in file from the course web page (http://www2.umdnj.edu/~kholodvl) and use it for you grid input. *For the listing of file content and the explanation of different parameters in this file consult the DOCK manual and an appendix of this tutorial.*

Run grid.

grid –i grid.in **–o** grid.out > grid.out &
Before proceeding to the next step check the grid.out file for errors!!

tail -20 grid.out If there is no error messages at the bottom of the file proceed to Docking

STAGE 4. Docking

Now run **DOCK**.

Download dock.in file from the course web page (http://www2.umdnj.edu/~kholodvl) and use it for you dock input. *For the listing of file content and the explanation of different parameters in this file consult the DOCK manual and an appendix of this tutorial.*

Run
dock6 –i dock.in **–o** dock.out > dock.out **&**

(Note: dock.out is an output to screen only. To save the output to file we use "**>**" operator

The whole run should be done approximately in 25-30 minutes.

Check the dock.out file for errors.
tail -20 dock.out
If there is a problem with growing ligand or a small size of a grid box, we need to increase the size of a box by repeating a step of generating box (**showbox** command) and increase extra margin to 4Å. Then regenerate grid and run dock again.

ANALYSIS OF THE RESULTS

First let us look at the favored structure (lowest energy dock based upon DOCK's scoring function) produced by DOCK.

rpr_ranked.mol2 - contains the lowest E dock result. This file contains the best docking conformations of the ligand. Look for RMSD values in the header for each conformer in this file
rpr_conformers.mol2 - contains the conformers tested.
rpr_orients.mol2 - contains the orientations tested.

Our output from this run produced a docked structure with a E_{bind} of –28.78 kcal/mol. (Note: your output might be different!)

Partial listing of the end of dock.out

```
##############################################
Molecule:  RPR
 Anchors:           1
 Orientations:      100
 Conformations:     211

 Grid Score:      -28.779739
       vdw:       -10.796704
        es:       -17.983036
##############################################
```

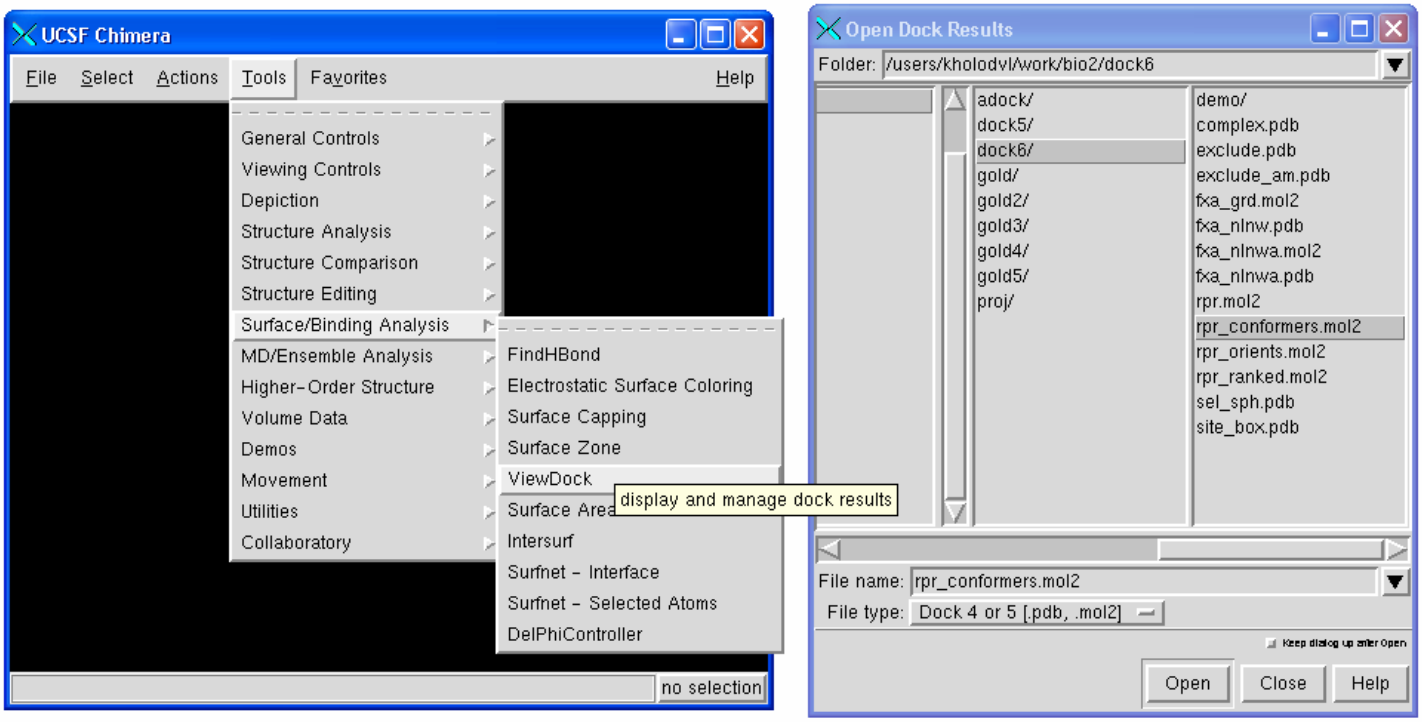
In the tutorial you do not need to run DOCK twice with different **max_orientation** parameter in dock.in configuration file (see Appendix). But you might need to experiment with this parameter for your project assignment.

To analyze conformers and orientations file we use a program CHIMERA.
Start chimera by typing
chimera

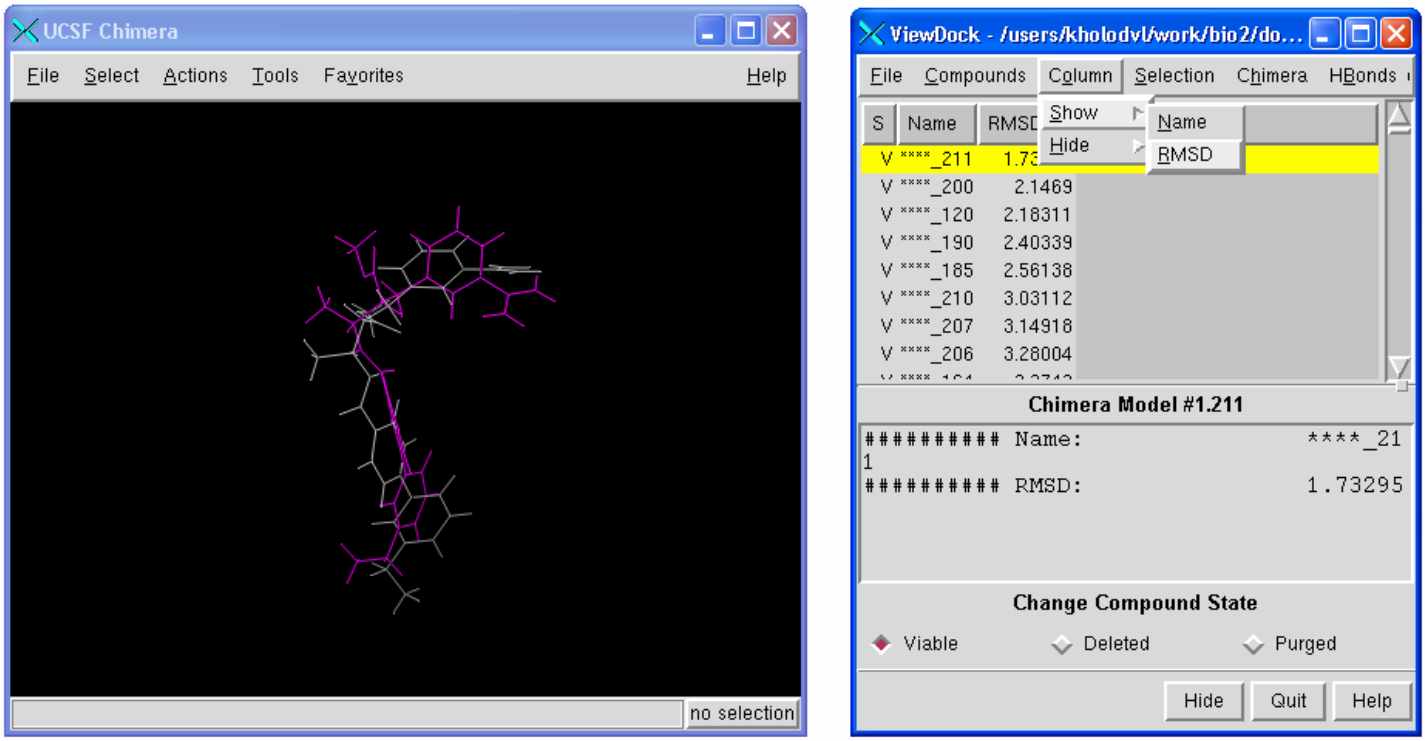
open a reference molecule extracted from the crystal structure rpr_xray.mol2
File > Open> rpr_xray.mol2
From the top menu choose
Tools > Surface/Binding analysis> ViewDock

Select rpr_conformers.mol2 file which contains all generated and processed conformers.

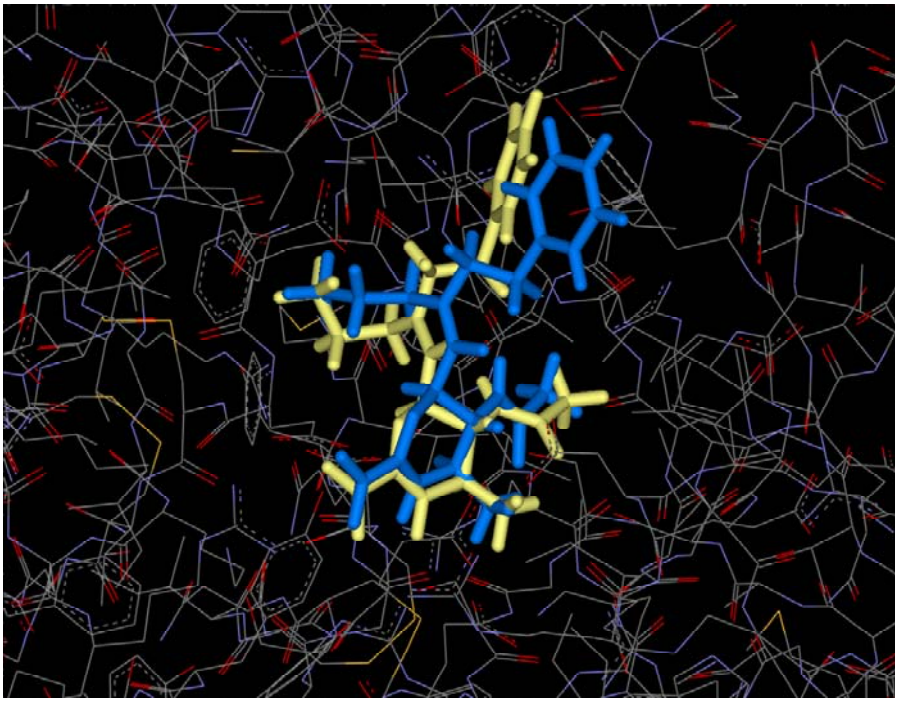
Refer to figure below.



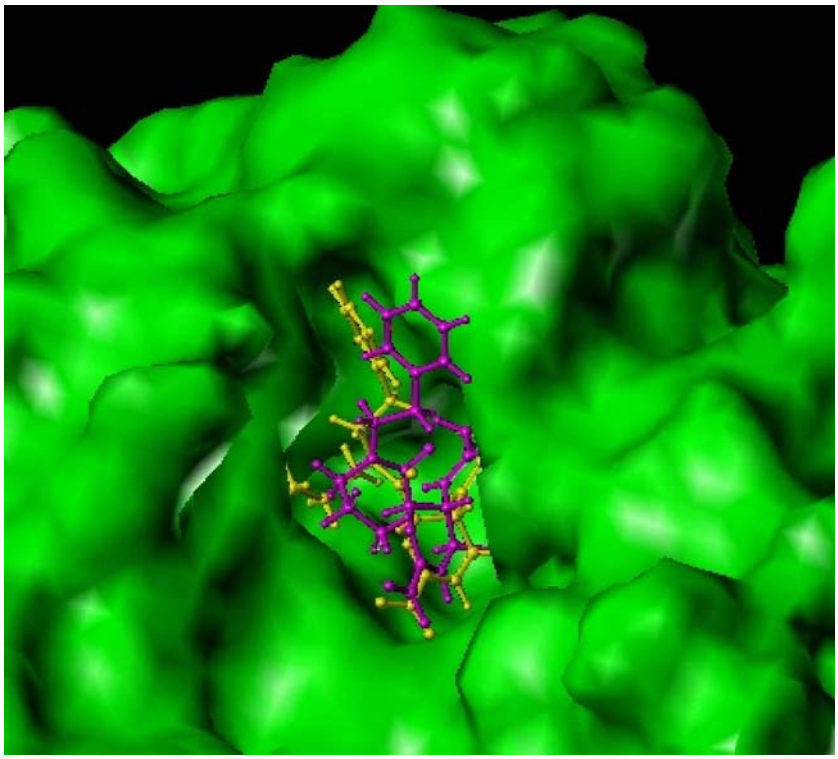
In the ViewDock window display the RMSD values:
Column > Show> RMSD



Sort results based on RMSD by clicking on RMSD column header. Select with a mouse one of the conformer from the list. In the chimera main window analyze both molecules – a crystal structure and a docked one. Notice all similarities and differences. The conformer with the best RMSD value in this column corresponds to the best fitted conformation found by DOCK. Notice it's not necessarily the same conformation stored in the rpr_ranked.mol2 file! **Write down RMSD for the two best fittings and save them as mol2 files**.



Favored docking orientation produced from simple $E = E_{vdw} + E_{elec}$ scoring function in DOCK colored in blue in comparison to the crystal structure colored in cream (light tan). Viewed in ViewerPro (Accelrys). This dock is in good accord with experimental data.



Similar to above results viewed in Sybyl. Favored docking orientation of the ligand obtained in DOCK colored in magenta in comparison to the crystal structure colored in yellow are presented in the binding groove of the protein generated as Connolly surface in green.

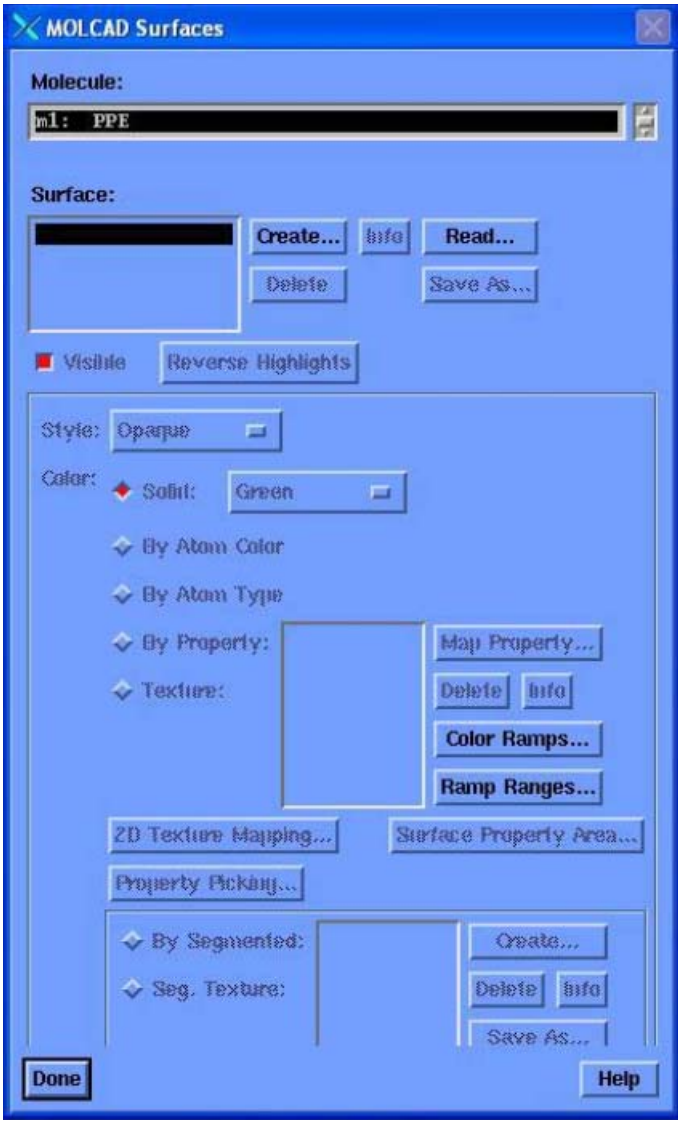
Visualization and analysis of docking results using Sybyl

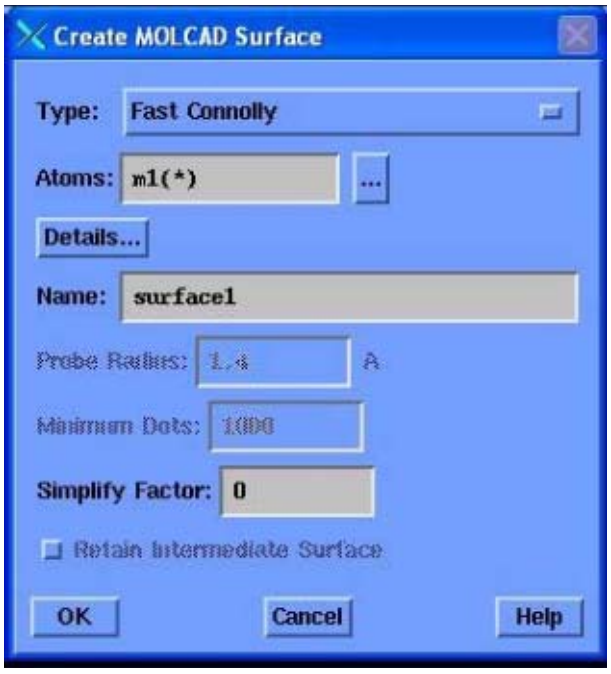
Generate molecular surface of protein.

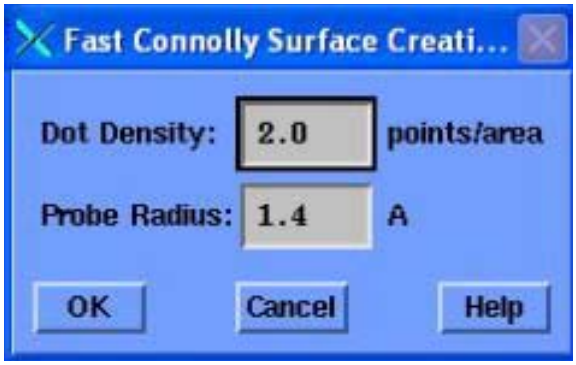
Run Sybyl

Sybyl7.2

Open protein structure fxa_nl.pdb.
DO NOT CENTER MOLECULE!
Go to
View->MolCad Surfaces->Molecular Surfaces







MOLCAD Surfaces
molecule: M1: Factor Xa
click on Create to create new surface

In Create MOLCAD Surface

Type: Fast Connolly
Atoms: M1(*) – all atoms
Name: surface 1
Simplify factor: 0

Click on Details

In Fast Connolly Surface Creation
Dot Density: 2.0 points/area
Probe Radius: 1.4 Å
Click OK to close this window and OK to start generate molecular surface.

If the surface does not show up choose the following parameters in still open MOLCAD Surfaces window
Check visible box
Style: opaque
Color : Solid green
Check “Apply immediately” box and click on Apply button
Click DONE.

You can change the visualization parameters (color, style) for already created MS any time going to View->MOLCAD Surfaces->Molecular Surfaces

Now you need to hide all atoms of protein to speed up work with surface, i.e. rotation, translation, zoom in/out.
View->Undisplay Atoms-> All->OK

Open best docking ligand files
rpr_ranked.mol2 or two best fitting mol2 files saved from Chimera.

If you remembered not to center the protein molecule when you open it in Sybyl you should have ligand visible in the binding pocket of protein surface similar to the figure above.
You may need to rotate the whole complex to see it.

For comparison open x-ray extracted ligand rpr_xray.mol2

You can also open sphere cluster (selected_spheres.pdb) and box (site_box.pdb) files to see how the spheres and docking box aligned with the binding groove.

Bibliography

1. Shoichet, B.K., D.L. Bodian, and I.D. Kuntz, *Molecular docking using shape descriptors.* J. Comp. Chem., 1992. **13**(3): p. 380-397.
2. Meng, E.C., B.K. Shoichet, and I.D. Kuntz, *Automated docking with grid-based energy evaluation.* J. Comp. Chem., 1992. **13**: p. 505-524.
3. Kuntz, I.D., J.M. Blaney, S.J. Oatley, R. Langridge, and T.E. Ferrin, *A geometric approach to macromolecule-ligand interactions.* J. Mol. Biol., 1982. **161**: p. 269-288.
4. Meng, E.C., D.A. Gschwend, J.M. Blaney, and I.D. Kuntz, *Orientational sampling and rigid-body minimization in molecular docking.* Proteins, 1993. **17**(3): p. 266-278.
5. Maignan S., Guilloteau J.P., Pouzieux S., Choi-Sledeski Y.M., Becker M.R., Klein S.I., Ewing W.R., Pauls H.W., Spada A.P., Mikol V., *Crystal Structures Of Human Factor Xa Complexed With Potent Inhibitors* // J.Med.Chem., 2000. **43**: p. 3226 2000

Appendix

Full listing of grid.in

```
##########################################################################
compute_grids                  yes
grid_spacing                   0.3
output_molecule                no
contact_score                  yes
contact_cutoff_distance        4.5
chemical_score                 no
energy_score                   yes
energy_cutoff_distance         5
atom_model                     a
attractive_exponent            6
repulsive_exponent             12
distance_dielectric            yes
dielectric_factor              4
bump_filter                    yes
bump_overlap                   0.5
receptor_file                  fxa_grd.mol2
box_file                       site_box.pdb
vdw_definition_file            /usr/local/dock/parameters/vdw_AMBER_parm99.defn
score_grid_prefix              grid
##########################################################################
```

The **vdw_AMBER.defn** file content is derived from the AMBER95 force field (Cornell, et al. *J. Am. Chem. Soc.* **1995**, *117*(19), 5179-5197.).

Full listing of dock.in

```
##########################################################################################
ligand_atom_file                                                  ./rpr.mol2
ligand_outfile_prefix                                              rpr
limit_max_ligands                                                  no
read_mol_solvation                                                 no
write_orientations                                                 yes
write_conformations                                                yes
skip_molecule                                                      no
calculate_rmsd                                                     yes
use_rmsd_reference_mol                                             yes
rmsd_reference_filename                                            ./rpr_xray.mol2
rank_ligands                                                       yes
max_ranked_ligands                                                 20
scored_conformer_output_override                                   no
orient_ligand                                                      yes
automated_matching                                                 yes
receptor_site_file                                                 ./selected_spheres.sph
max_orientations                                                   100
critical_points                                                    no
chemical_matching                                                  no
use_ligand_spheres                                                 no
flexible_ligand                                                    yes
min_anchor_size                                                    50
num_anchor_orients_for_growth                                      100
number_confs_for_next_growth                                       100
use_internal_energy                                                yes
internal_energy_att_exp                                            6
internal_energy_rep_exp                                            12
internal_energy_dielectric                                         4.0
use_clash_overlap                                                  no
bump_filter                                                        no
score_molecules                                                    yes
contact_score_primary                                              no
contact_score_secondary                                            no
grid_score_primary                                                 yes
grid_score_secondary                                               yes
grid_score_vdw_scale                                               1
grid_score_es_scale                                                1
grid_score_grid_prefix                                             ./grid
minimize_ligand                                                    yes
minimize_anchor                                                    no
minimize_flexible_growth                                           yes
use_advanced_simplex_parameters                                    no
simplex_max_cycles                                                 1
simplex_score_converge                                             0.1
simplex_cycle_converge                                             1.0
simplex_trans_step                                                 1.0
simplex_rot_step                                                   0.1
simplex_tors_step                                                  10.0
simplex_grow_max_iterations                                        10
simplex_final_min_add_internal                                     no
simplex_secondary_minimize_pose                                    no
simplex_random_seed                                                0
atom_model                                                         all
vdw_defn_file                      /usr/local/dock/parameters/vdw_AMBER_parm99.defn
flex_defn_file                               /usr/local/dock/parameters/flex.defn
flex_drive_file                      /usr/local/dock/parameters/flex_drive.tbl
##########################################################################################
```