

# **Introductory Tutorial for the UCSF DOCK Program (v. 6.5)**

Author: John E. Kerrigan, Ph.D.  
The Cancer Institute of New Jersey  
Robert Wood Johnson Medical School  
120 Albany Street  
New Brunswick, NJ 08903 U.S.A.  
(732) 235-4473 phone  
(732) 235-5252 fax  
[kerrigje@umdnj.edu](mailto:kerrigje@umdnj.edu)  
<http://cinjweb.umdnj.edu/~kerrigje>

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

```
mkdir proj1  
cd proj1
```

**CAUTION:** Be sure that you have enough disk quota for your docking job. The dms program, sphgen and DOCK use a fair amount of disk space. To check your quota use the following commands:

If logged on to one of the SGI's ...  
1<sup>st</sup> login to the main server to check your quota.

```
ssh username@server.umdj.edu  
quota -v
```

Exit when finished. Type "exit" then hit the "Enter" key.

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.  $E_{\text{bind}} = E_{\text{vdw}} + E_{\text{elec}}$ ). 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:** *Inhibition of Factor Xa* We will study the inhibition of coagulation factor, factor Xa, as this enzyme has a multitude of inhibitor data. The x-ray crystal structure (1FJS) will be used as a modeling template.[5]

STAGE 1. Prepare the ligand.

Obtain the PDB file, 1FJS.pdb, from the protein data bank (<http://rutgers.rcsb.org/pdb/>). Cut out the drug (residue Z34) and paste into a new file and copy over the CONECT records ranging from atoms 2241 to 2278. Save the file naming it z34.pdb.

Open the UCSF Chimera program by typing “chimera” in the unix shell. Open z34.pdb. Use the middle mouse button to translate the structure and the right mouse button to rotate the structure. Zoom in and out using the mouse wheel.

Go to: Tools > Structure Editing > AddH

Accept Defaults and click OK

Add Charges: Tools > Structure Editing > Add Charge

Accept Defaults again and click OK. A dialog called “Specify Net Charges” will appear. We will use the AM1-BCC charge method (default). Click OK.

File > Save Mol2... (Name the file “z34.mol2”)

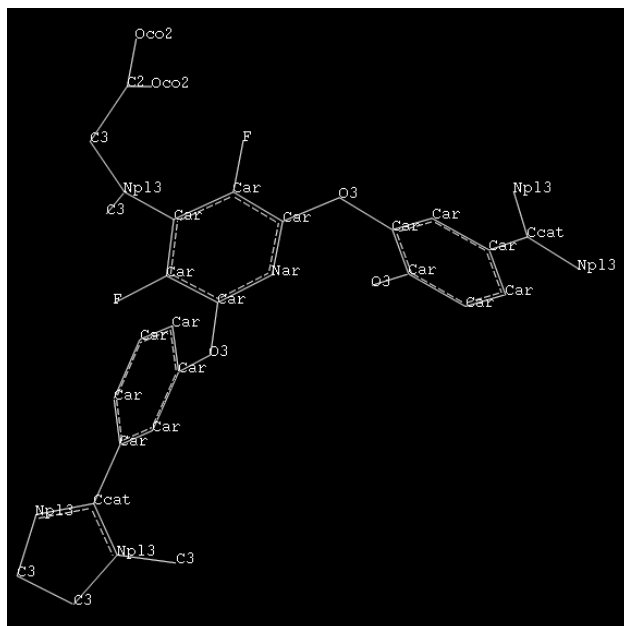
There are several atom types we must fix before proceeding.

#### Initial Atom Types

Build/Edit > Modify > Atom ... > *Option* > ONLY\_TYPE Click OK  
*Atom Expression* >

Using the Mouse: Select all atoms requiring atom type correction (see the final corrected atom types in the figure below).

Your atom type labels should appear as in the figure below.



Corrected Atom Types

Now add the hydrogen atoms.

Build/Edit > Add > Hydrogens

Compute > Charges > Gasteiger-Huckel

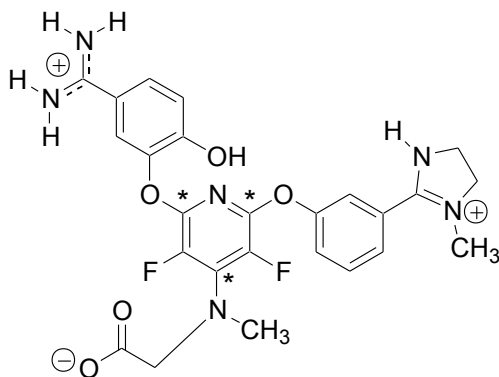
Click on “No” when the dialog asks you if you want to change formal charges.

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

The ANCHOR set

Build/Edit > Define > Static Set ... > *Option* > ATOM Click OK

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

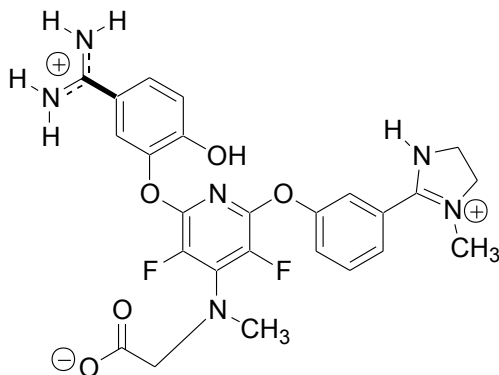


Name this set ANCHOR.

### The RIGID set

Build/Edit > Define > Static Set ... > *Option* > BOND Click OK

In *Bond Expression*, select the following highlighted bond ...



Name this set RIGID.

Using File > Save As ... Save the ligand as **zk807834.mol2**; Format: MOL2.

### STAGE 2. Site Characterization.

Before we use 1FJS.pdb we must remove the substitution code in the original PDB file. Use a simple text editor (nedit, jot or even MS Word) to perform this operation. The substitution code is residue substitutions in the protein which represent mutations or some other protein form. We do not need these additional residues. These residues are indicated by a letter following the residue ID number. For example ...

ATOM	351	CD	GLN	A	61	7.419	14.043	15.930	1.00	47.58	C
ATOM	352	OE1	GLN	A	61	6.361	14.211	16.539	1.00	51.11	O
ATOM	353	NE2	GLN	A	61	7.727	12.896	15.323	1.00	48.58	N
<b>ATOM</b>	<b>354</b>	<b>N</b>	<b>ALA</b>	<b>A</b>	<b>61A</b>	<b>8.104</b>	<b>18.641</b>	<b>12.715</b>	<b>1.00</b>	<b>35.57</b>	<b>N</b>
<b>ATOM</b>	<b>355</b>	<b>CA</b>	<b>ALA</b>	<b>A</b>	<b>61A</b>	<b>7.989</b>	<b>20.049</b>	<b>12.328</b>	<b>1.00</b>	<b>34.78</b>	<b>C</b>
<b>ATOM</b>	<b>356</b>	<b>C</b>	<b>ALA</b>	<b>A</b>	<b>61A</b>	<b>8.523</b>	<b>20.298</b>	<b>10.915</b>	<b>1.00</b>	<b>34.80</b>	<b>C</b>
<b>ATOM</b>	<b>357</b>	<b>O</b>	<b>ALA</b>	<b>A</b>	<b>61A</b>	<b>9.629</b>	<b>19.878</b>	<b>10.585</b>	<b>1.00</b>	<b>35.43</b>	<b>O</b>
<b>ATOM</b>	<b>358</b>	<b>CB</b>	<b>ALA</b>	<b>A</b>	<b>61A</b>	<b>8.719</b>	<b>20.942</b>	<b>13.335</b>	<b>1.00</b>	<b>32.76</b>	<b>C</b>
ATOM	359	N	LYS	A	62	7.728	20.968	10.078	1.00	35.29	N

Remove those residues in 1FJS.pdb with letters following their residue ID numbers. Now, remove all of the HETATM lines in the PDB file (This operation removes the drug, water and other unnecessary molecules). Save the new file as **fxa.pdb**.

Start sybyl using sybyl7.0 command in your unix shell. Open the PDB file you saved earlier, **fxa.pdb**, using ...

File > Read ...

Remember to Center View only!

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. Alternatively, you may use UCSF Chimera’s built-in DMS program.

```
dms fxa.pdb -a -n -w 1.4 -v -o fxa.dms
```

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!).

INSPH

```
-----  
fxa.dms  
R  
X  
0.0  
4.0  
1.4  
fxa.sph  
-----
```

Run sphgen simply by typing “sphgen” in the unix shell as follows.

### **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.sph zk807834.mol2 10.0
```

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.

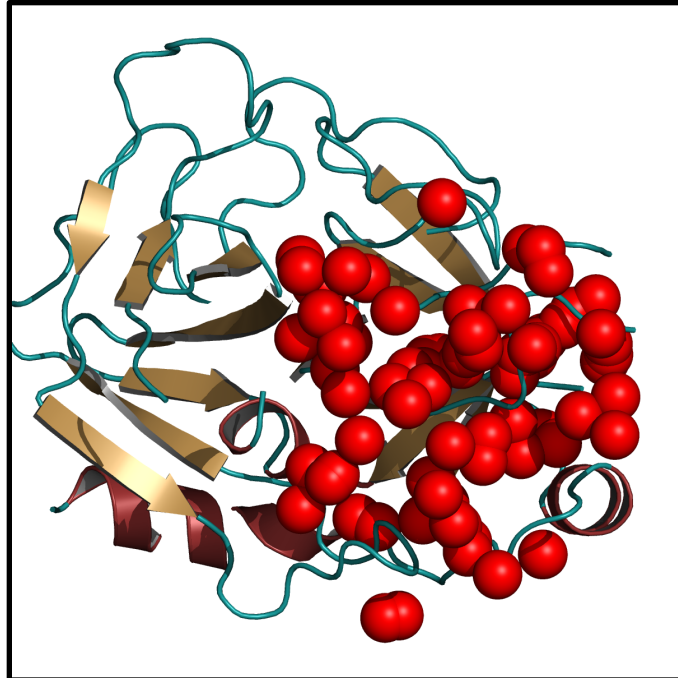
Type **showsphere** <enter>

Name of the sphere cluster file: [selected\\_spheres.sph](#)

Cluster number to process: 1

Generate surfaces? N

Name for output pdb file: [sel\\_sph.pdb](#)



Spheres selected by the sphere\_selector program in red.

Run showbox.

Type **showbox** <enter>.

*automatically construct box to enclose spheres [Y/N] ?*

Y

*extra margin to also be enclosed (angstroms)?*

5.0

*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 and use it for your grid input. See contents of grid.in which follows.

Run grid.

**nohup grid -i grid.in -o grid.out &**

## Contents 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 9999
atom_model             a
attractive_exponent    6
repulsive_exponent     9
distance_dielectric    yes
dielectric_factor      4
bump_filter            yes
bump_overlap           0.75
receptor_file          fxa.mol2
box_file               site_box.pdb
vdw_definition_file    /usr/local/dock6/parameters/vdw_AMBER_parm99.defn
score_grid_prefix      grid
```

## STAGE 4. Docking

Now run **DOCK**.

Download “dock.in” file from the course web page and use it for you dock input. See the dock.in file which follows.

Run

**nohup dock6 -i dock.in -o dock.out &**

## Contents of dock.in

```
ligand_atom_file      zk807834.mol2
limit_max_ligands     no
skip_molecule        no
read_mol_solvation    no
calculate_rmsd        no
use_database_filter   no
orient_ligand         yes
automated_matching    yes
receptor_site_file    selected_spheres.sph
max_orientations      1000
critical_points        no
chemical_matching     no
use_ligand_spheres    no
use_internal_energy   yes
internal_energy_rep_exp 12
flexible_ligand       yes
min_anchor_size       40
pruning_use_clustering yes
pruning_max_orients   100
pruning_clustering_cutoff 100
pruning_conformer_score_cutoff 25.0
use_clash_overlap     no
write_growth_tree     no
bump_filter           no
bump_grid_prefix      grid
max_bumps_anchor      12
max_bumps_growth      12
```

score_molecules	yes
contact_score_primary	no
contact_score_secondary	no
grid_score_primary	yes
grid_score_secondary	no
grid_score_rep_rad_scale	1
grid_score_vdw_scale	1
grid_score_es_scale	1
grid_score_grid_prefix	grid
dock3.5_score_secondary	no
continuous_score_secondary	no
gbsa_zou_score_secondary	no
gbsa_hawkins_score_secondary	no
amber_score_secondary	no
minimize_ligand	yes
minimize_anchor	yes
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_anchor_max_iterations	1000
simplex_grow_max_iterations	500
simplex_grow_tors_prem_in_iterations	0
simplex_random_seed	0
simplex_restraint_min	no
atom_model	all
vdw_defn_file	
/usr/local/dock6/parameters/vdw_AMBER_parm99.defn	
flex_defn_file	
/usr/local/dock6/parameters/flex.defn	
flex_drive_file	
/usr/local/dock6/parameters/flex_drive.tbl	
ligand_outfile_prefix	zk807_out
write_orientations	no
num_scored_conformers	10
write_conformations	no
cluster_conformations	yes
cluster_rmsd_threshold	2.0
rank_ligands	no

Here are our results ...

From the tail of dock.out (**tail -25 dock.out**):

-----  
Molecule: ZK807834

Elapsed time for docking: 63 seconds

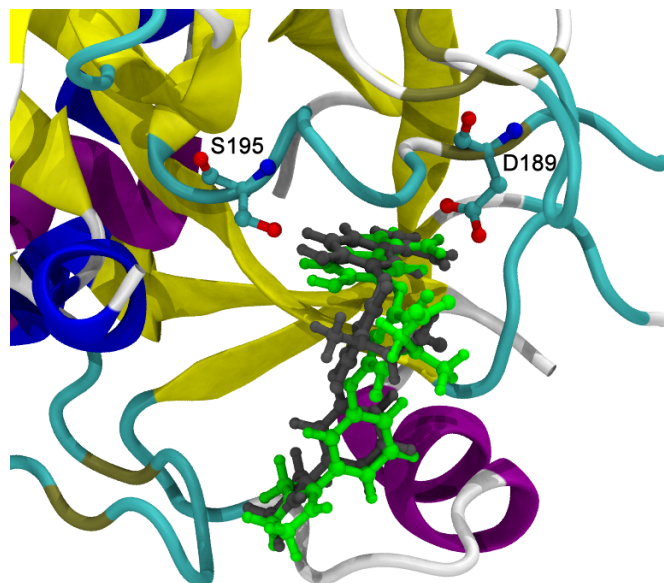
Anchors: 1  
Orientations: 1000  
Conformations: 132

Grid Score: -92.948013  
Grid\_vdw: -75.534805  
Grid\_es: -17.413210  
Int\_energy: 9.015675

1 Molecules Processed

Total elapsed time: 64 seconds

---



The structure in gray is from the x-ray crystal structure and the structure in green is the docked structure.

From the figure above we see that by and large DOCK has returned the same orientation as from the crystal structure. There are a few subtle differences in conformation, however.

Next, we will dock a small database of compounds (we used the zk807834 compound from the crystal structure as a template to build these compounds in Sybyl 7.0; hence, zk807834 also appears in this database). The database is in mol2 format. Download **pyridins.mol2** from the course webpage.

### **dock\_db.in**

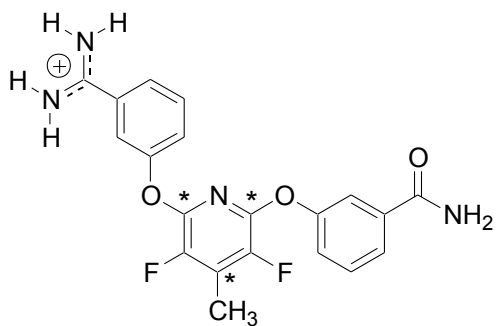
```
ligand_atom_file      pyridins.mol2
limit_max_ligands    no
skip_molecule       no
read_mol_solvation   no
calculate_rmsd       no
use_database_filter  yes
dbfilter_max_heavy_atoms 999
dbfilter_min_heavy_atoms 0
dbfilter_max_rot_bonds 999
dbfilter_min_rot_bonds 0
dbfilter_max_molwt   9999.0
dbfilter_min_molwt   0.0
dbfilter_max_formal_charge 10.0
dbfilter_min_formal_charge -10.0
orient_ligand        yes
automated_matching   yes
receptor_site_file   selected_spheres.sph
max_orientations     1000
critical_points       no
chemical_matching     no
use_ligand_spheres   no
```

use_internal_energy	yes
internal_energy_rep_exp	9
flexible_ligand	yes
min_anchor_size	5
pruning_use_clustering	yes
pruning_max_orients	100
pruning_clustering_cutoff	100
pruning_conformer_score_cutoff	100.0
use_clash_overlap	no
write_growth_tree	no
bump_filter	no
score_molecules	yes
contact_score_primary	no
contact_score_secondary	no
grid_score_primary	yes
grid_score_secondary	no
grid_score_rep_rad_scale	1
grid_score_vdw_scale	1
grid_score_es_scale	1
grid_score_grid_prefix	grid
dock3.5_score_secondary	no
continuous_score_secondary	no
gbsa_zou_score_secondary	no
gbsa_hawkins_score_secondary	no
amber_score_secondary	no
minimize_ligand	yes
minimize_anchor	yes
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_anchor_max_iterations	1000
simplex_grow_max_iterations	500
simplex_grow_tors_premin_iterations	0
simplex_random_seed	0
simplex_restraint_min	no
atom_model	all
vdw_defn_file	/Users/kerrigje/software/dock6/parameters/vdw_AMBER_parm99.defn
flex_defn_file	/Users/kerrigje/software/dock6/parameters/flex.defn
flex_drive_file	/Users/kerrigje/software/dock6/parameters/flex_drive.tbl
ligand_outfile_prefix	pyridins_out
write_orientations	no
num_scored_conformers	1
rank_ligands	yes
max_ranked_ligands	20000

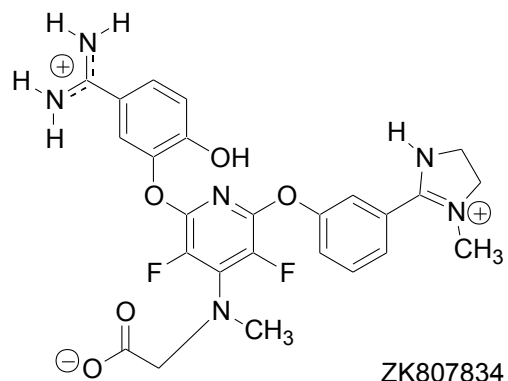
The *num\_scored\_conformers* parameter is set to 1; thereby, instructing DOCK to save only the lowest energy hit for each docking to *pyridins\_out*. This is important as we will only re-score our lowest energy hits with the GB/SA scoring function later. Run the database using ...

**nohup dock6 -i dock\_db.in -o dock\_db.out &**

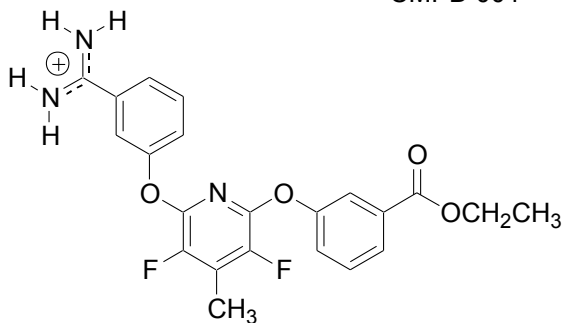
The database docking gives slightly different results for the zk807834 compound. This run was performed on a 2.2 GHz Linux workstation. Here are the structures of the compounds in the database:



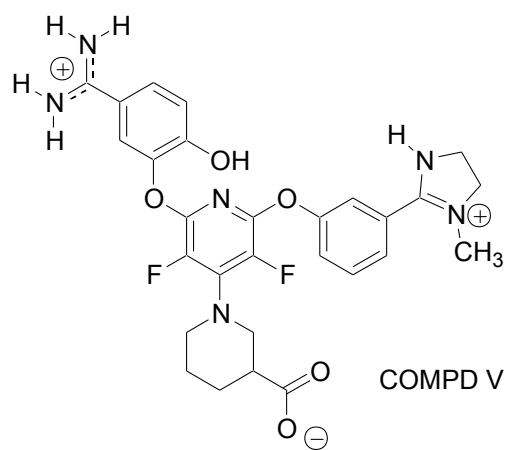
CMPD 001



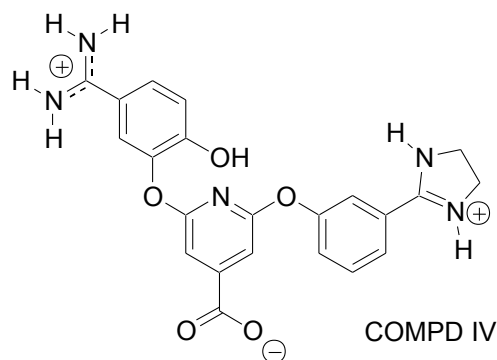
ZK807834



CMPD 002



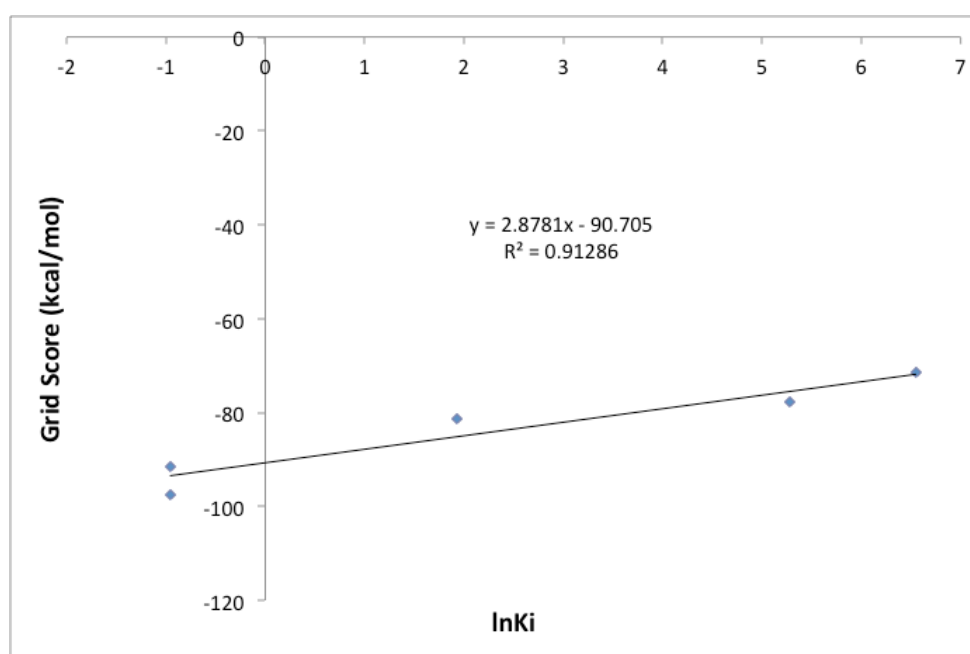
COMPD V



COMPD IV

The pyridins\_out\_scored.mol2 database can be viewed using chimera (see <http://www.cgl.ucsf.edu/chimera/>) or Sybyl.

	$\ln K_i$	Grid Score ( $E_{int}$ ) in kcal/mol
COMPD_V	-0.96	-97.44
ZK807834	-0.96	-91.41
COMPD_IV	1.93	-81.46
COMPD002	5.28	-77.62
COMPD001	6.55	-71.52



The  $K_i$  data for compounds 001 and 002 were obtained from a review paper.[6] All other  $K_i$  data were obtained from the structure paper.[5] The compounds in the series fall into line in this plot very well. Normally, we would sample more than just five compounds to establish a trend. This example is just illustrative of the application of the DOCK method and to caution one about the use of the simple scoring function ( $E_{int} = E_{vdw} + E_{elec}$ ) used in DOCK.

### Hawkins GBSA

Now, we are ready to rescore the database docking using the Hawkins GB/SA scoring function. [7, 8] Here is our input file ...

#### dock\_hct\_gbsa.in

```
ligand_atom_file      pyridins_out_ranked.mol2
limit_max_ligands     no
skip_molecule        no
read_mol_solvation    no
```

calculate_rmsd	no
use_database_filter	no
orient_ligand	no
use_internal_energy	no
flexible_ligand	no
bump_filter	no
score_molecules	yes
contact_score_primary	no
contact_score_secondary	no
grid_score_primary	yes
grid_score_secondary	no
grid_score_rep_rad_scale	1
grid_score_vdw_scale	1
grid_score_es_scale	1
grid_score_grid_prefix	grid
dock3.5_score_secondary	no
continuous_score_secondary	no
descriptor_score_secondary	no
gbsa_zou_score_secondary	no
gbsa_hawkins_score_secondary	yes
gbsa_hawkins_score_rec_filename	fxa.mol2
gbsa_hawkins_score_solvent_dielectric	78.5
gbsa_hawkins_use_salt_screen	yes
gbsa_hawkins_score_salt_conc(M)	0.15
gbsa_hawkins_score_gb_offset	0.09
gbsa_hawkins_score_cont_vdw_and_es	yes
gbsa_hawkins_score_vdw_att_exp	6
gbsa_hawkins_score_vdw_rep_exp	9
grid_score_rep_rad_scale	1
minimize_ligand	no
simplex_max_iterations	1
simplex_tors_premin_iterations	0
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_secondary_minimize_pose	yes
use_advanced_secondary_simplex_parameters	no
simplex_secondary_max_iterations	1
simplex_random_seed	0
simplex_restraint_min	no
atom_model	all
vdw_defn_file	
/Users/kerrigje/software/dock6/parameters/vdw_AMBER_parm99.defn	
flex_defn_file	
/Users/kerrigje/software/dock6/parameters/flex.defn	
flex_drive_file	
/Users/kerrigje/software/dock6/parameters/flex_drive.tbl	
ligand_outfile_prefix	pyridins_gbsa
write_orientations	no
num_primary_scored_conformers_rescored	1
num_secondary_scored_conformers	1
rank_primary_ligands	no
rank_secondary_ligands	no

Run using

**nohup dock6 -i dock\_hct\_gbsa.in -o dock\_hct\_gbsa.out &**

Our re-scored values were disappointing

Cmpd	log $K_i$	GB/SA Score (kcal/mol)
CMPD001	6.55	-80.77
CMPD002	5.28	-87.86
CMPDIV	1.93	-60.05

CMPDV	-0.96	-92.66
ZK807834	-0.96	-77.96

## Zou GBSA

For the Zou GBSA method, we need to prepare two sets of grids: the surface area grid (sa\_grid); the generalized Born grid (gb\_grid).[9] These grids are prepared in separate directories: nchemgrid\_SA and nchemgrid\_GB. We need to prepare a receptor pdb file (rec.pdb) with hydrogen atoms added (use the fxa.mol2 file and save as a PDB from Chimera). We need an input file named “INCHEM” in each directory. See section 3.3 of UCSF DOCK Manual for an explanation of these input files.

For nchemgrid\_SA

INCHEM

```
../rec.pdb
../parameters/prot.table.ambcrg.ambH
../parameters/vdw.parms.amb
../site_box.pdb
0.3
1.4
2
8.0
sa_grid
```

For nchemgrid\_GB

INCHEM

```
../rec.pdb
cavity.pdb
../parameters/prot.table.ambcrg.ambH
../parameters/vdw.parms.amb
../site_box.pdb
0.3
2
1
8.0 8.0
78.3 78.3
2.3 2.8
gb_grid
1
```

Note: the “cavity.pdb” file is infrequently used and is an empty file. Create using the “touch” unix command.

**touch cavity.pdb**

**dock\_zou\_gbsa.in**

```
ligand_atom_file                pyridins_out_ranked.mol2
limit_max_ligands               no
skip_molecule                 no
read_mol_solvation             no
calculate_rmsd                 no
use_database_filter            no
orient_ligand                  no
use_internal_energy            no
flexible_ligand                no
bump_filter                    no
score_molecules                yes
contact_score_primary          no
contact_score_secondary        no
grid_score_primary             yes
grid_score_secondary           no
grid_score_rep_rad_scale       1
grid_score_vdw_scale           1
grid_score_es_scale            1
grid_score_grid_prefix         grid
dock3.5_score_secondary        no
continuous_score_secondary     no
descriptor_score_secondary     no
gbsa_zou_score_secondary       yes
gbsa_zou_gb_grid_prefix        ./nchemgrid_GB/gb_grid
gbsa_zou_sa_grid_prefix        ./nchemgrid_SA/sa_grid
gbsa_zou_vdw_grid_prefix       grid
gbsa_zou_screen_file           ./parameters/screen.in
gbsa_zou_solvent_dielectric    78.3
minimize_ligand                no
atom_model                     all
vdw_defn_file                  /Users/kerrigje/software/dock6/parameters/vdw_AMBER_parm99.defn
flex_defn_file                 /Users/kerrigje/software/dock6/parameters/flex.defn
flex_drive_file                /Users/kerrigje/software/dock6/parameters/flex_drive.tbl
ligand_outfile_prefix          pyridins_zou
write_orientations             no
num_primary_scored_conformers_rescored 1
num_secondary_scored_conformers 1
rank_primary_ligands           no
rank_secondary_ligands         no
```

Run using

**nohup dock6 -i dock\_zou\_gbsa.in -o dock\_zou\_gbsa.out &**

Our results are ...

Cmpd	log $K_i$	GB/SA Score (kcal/mol)
CMPD001	6.55	-13.99
CMPD002	5.28	-10.51
CMPDIV	1.93	-20.58
CMPDV	-0.96	-14.76
ZK807834	-0.96	-15.19

## References:

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, *Orientalional sampling and rigid-body minimization in molecular docking*. Proteins, 1993. **17**(3): p. 266-278.
5. Adler, M., D. Davey, G. Phillips, S. Kim, J. Jancarik, G. Rumennik, D. Light, and M. Whitlow, *Preparation, Characterization and the Crystal Structure of the Inhibitor Zk807834 (Ci-1031) Complexed with Factor Xa*. Biochemistry, 2000. **39**(41): p. 12534-12542.
6. Kontogiorgis, C. and D. Hadjipavlou-Litina, *Current Trends in Quantitative Structure Activity Relationships on FXa inhibitors: Evaluation and Comparative Analysis*. Med. Res. Rev., 2004. **24**(6): p. 687-747.
7. Hawkins, G.D., C.J. Cramer, and D.G. Truhlar, *Pairwise solute descreening of solute charges from a dielectric medium*. Chem. Phys. Lett., 1995. **246**: p. 122-129.
8. Hawkins, G.D., C.J. Cramer, and D.G. Truhlar, *Parametrized models of aqueous free energies of solvation based on pairwise descreening of solute atomic charges from a dielectric medium*. J. Phys. Chem., 1996. **100**: p. 19824-19839.
9. Liu, H., I.D. Kuntz, and X. Zou, *Pairwise GB/SA Scoring Function for Structure-based Drug Design*. J. Phys. Chem. B, 2004. **108**: p. 5453-5462.