Introductory Tutorial for the AMBER Score in DOCK6:

By

Devleena Shivakumar
The Scripps Research Institute
10550 N. Torrey Pines Rd, TPC 15
La Jolla, CA 92037, USA
Phone: (858)-784-9781, (858)-784-9768
FAX: (858)-784-8896
Email: devleena@scripps.edu, case@scripps.edu

5/2/06

What is Amber score?

The generalized Born/surface area (GB/SA) continuum model for solvation free energy is a fast and accurate alternative to using explicit solvent model for molecular simulations. We have now implemented this physics-based method in the Amber scoring function in the program DOCK6. To curtail the computational cost while still maintaining the accuracy, the atoms distant from the site of ligand binding are kept frozen. In doing so the CPU time is not spent updating the energy and derivatives during the course of the simulation. The main advantage of AMBER score is – both the ligand and the active site of the protein can be flexible, allowing small structural rearrangements to reproduce the so-called "induce-fit" while performing the scoring function.

When a user calls for Amber score, the program performs minimization, and MD simulation on individual ligand, receptor, and the compound, and calculates the score as follows:

$$\mathbf{E_{binding} = E_{complex} - (E_{receptor} + E_{ligand})}$$

Where E is obtained from:

$\mathbf{E = E_{MM} + (E_{p\text{-}sol} + E_{np\text{-}sol})}$

$\mathbf{E_{MM\,=}\,E_{vdW} + E_{es} + E_{int}}$ ---- obtained from AMBER MM potentials

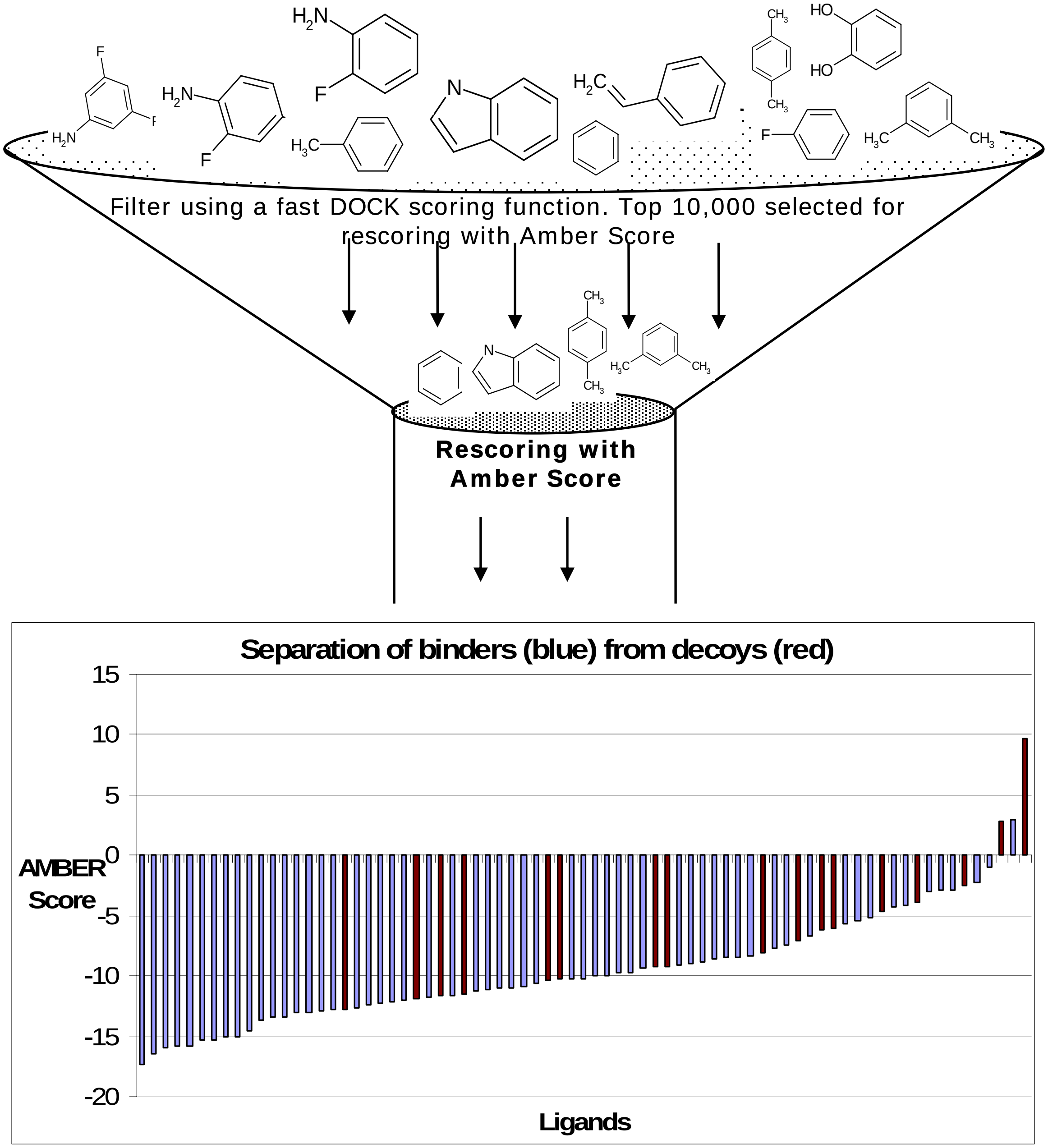
$\mathbf{E_{p\text{-}sol}}$ **---** Electrostatic part of solvation energy using GB

$\mathbf{E_{np\text{-}sol}}$ --- Non-polar part of solvation energy using SA

The user has the option to increase or decrease the number of minimization and MD simulation steps. However, it is not desirable to have higher number of steps due to the time taken for the calculations. For various protein test cases, we have found 100 minimization and 3000 MD steps to produce good results. These are set as defaults in the program.

It is highly recommended to run a DOCK calculation with a less expensive primary/secondary score to write out the topmost poses. Amber score should be used on these topmost pose for each ligand. For example, for T4 Lysozyme the DOCK score is calculated for 1 million compounds from ACD directory. Top 5000-10,000 compounds ranked by DOCK are passed through Amber score for further refinement. This is further illustrated in the cartoon below:

Millions of compounds from a database

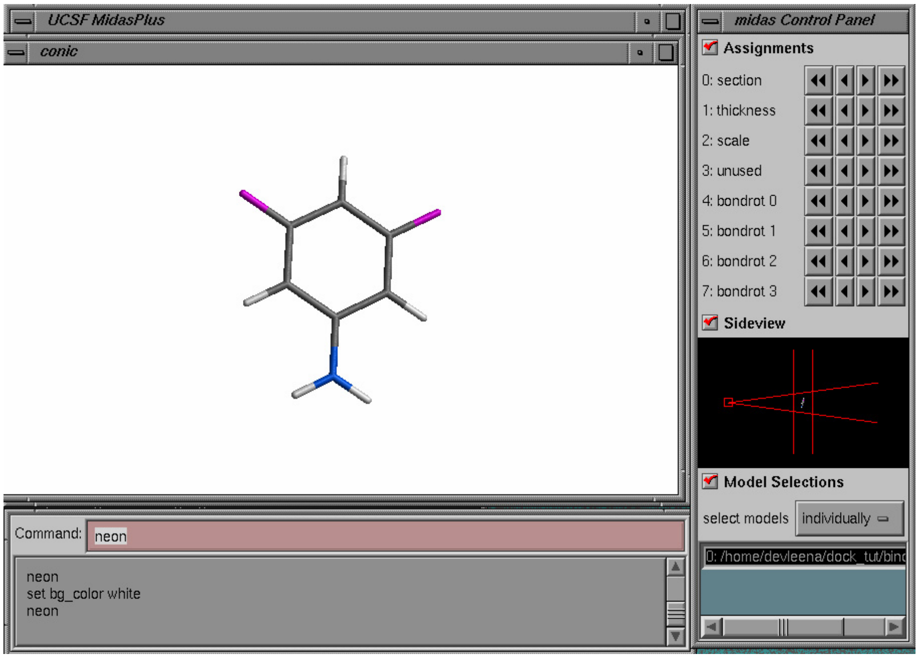


Filter using a fast DOCK scoring function. Top 10,000 selected for rescoring with Amber Score

Rescoring with Amber Score

Part I: Input files preparation.

1) *Start with the output mol2 file from a previous DOCK run [lig.mol2].*



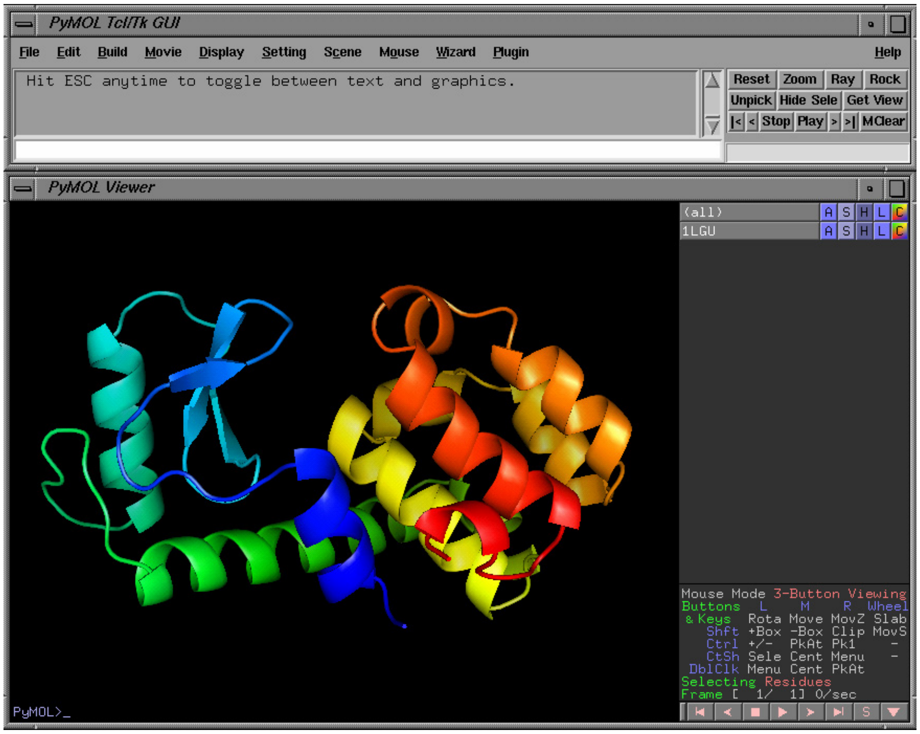
Command: neon

```
neon
set bg_color white
neon
```

2) *Receptor without cofactors. (1lgu.pdb)*

(a) Clean PDB:

Remove all the ligand, ions and crystal water molecules from the receptor pdb file. If you know that certain water molecules, ions play catalytic or structural role, use your scientific judgment to decide whether to keep them in the PDB file.



```
Hit ESC anytime to toggle between text and graphics.
```

```
Mouse Mode 3-Button Viewing
Buttons  L    M    R  Wheel
& Keys  Rota Move MovZ Slab
  Shft  +Box -Box Clip MovS
  Ctrl  +/-  PkAt Pk1   -
  CtSh  Sele Cent Menu  -
 DblClk Menu Cent PkAt
Selecting Residues
Frame [  1/   1] 0/sec
```

```
PyMOL>_
```

[Structure of T4 Lysozyme, PDB: 1LGU)]

(b) Determine the protonation state of the histidine and other titratable residues in the receptor. Care should be taken to assign the appropriate protonation state, especially if the residue is at or near the active site or within the flexible region while scoring calculations. Use experimental data from the literature, or your chemical intuition to assign the protonation states for these residues. [Hint: Check for hydrogen bonding residues nearby to see whether the His or Asp should be protonated.] Or, you can use softwares to do this job. Some examples:

i. PDB2PQR [http://pdb2pqr.sourceforge.net/] - Python software package that automates the PDB file preparation and protonation state assignments.

ii. H^{++} [http://biophysics.cs.vt.edu/H++/] is a tool to estimate pKa's of protein side chains, and to automate the process of assigning protonation states for molecular dynamics simulations.

(c) After assigning the protonation states, make sure that your receptor PDB file has residue names according to the AMBER readable format. Check the name of the residues to make sure that they have correct names:

Group or residue	**Residue Name, Alias**
Acetyl beginning group	ACE
Amine ending group	NHE
N-methylamine ending group	NME
Alanine	ALA
Arginine	ARG
Asparagine	ASN
Aspartic acid	ASP
Aspartic acid--protonated	ASH
Cysteine	CYS
Cysteine--deprotonated	CYM
Cystine, S--S crosslink	CYX
Glutamic acid	GLU
Glutamic acid--protonated	GLH
Glutamine	GLN
Glycine	GLY
Histidine, delta H	HID
Histidine, epsilon H	HIE
Histidine, protonated	HIP
Isoleucine	ILE
Leucine	LEU
Lysine	LYS
Methionine	MET
Phenylalanine	PHE
Proline	PRO
Serine	SER
Threonine	THR
Tryptophan	TRP
Tyrosine	TYR
Valine	VAL

3) ***Prepare AMBER readable input file***s for each ligand, receptor and the corresponding complex. This is done with the help of a perl script that is provided in the bin directory – prepare_amber.pl

Find out whether perl is installed in your machine.

```
$ which perl
/usr/bin/perl
```

If you cannot find perl on your machine, please install a copy.

The command line for using prepare_amber.pl is:

```
prepare_amber.pl lig.mol2 1lgu.pdb
```

Output files:

Files associated with Ligand:
lig.amber.score.mol2
lig.1.mol2
lig.1.amber.pdb
lig.1.gaff.mol2
lig.1.prmtop
lig.1.frcmod
lig.1.inpcrd

Files associated with receptor
1lgu.prmtop
1lgu.amber.pdb
1lgu.inpcrd

Files associated with complex:
1lgu.lig.1.prmtop
1lgu.lig.1.amber.pdb
1lgu.lig.1.inpcrd

prepare_amber.pl also has the capability to split a file containing multiple mol2 into individual mol2 files that are then read by the program.
Since in this example, there was only one ligand in lig.mol2, the output was lig.1.mol2. Had there been 2 ligands in the mol2 file, the output prefix will be: lig.1.mol2, lig.2.mol2 …

The following is done by the script prepare_amber.pl:

(i) Adds hydrogens to protein & ligand

(ii) Generate a mol2 file with suffix amber.score.mol2 that will be read into the DOCK run (lig.amber.score.mol2).

(iii) Run antechamber program to determine semi-empirical charges (am1-bcc) for the ligand.

(iv) Creates parameter file for ligand using GAFF forcefield (prmtop and frcmod) and writes a mol2 file with GAFF atom types (gaff.mol2)

(v) Read in the PDB file for the receptor; add hydrogens if not present; add amber force field atom types and charges. Generate parameter and coordinate file.

(vi) Combine each ligand with the receptor to generate the parameter and coordinate files for each complex.

4) Run DOCK6

Prepare an input file for DOCK6 run. For ligand_atom_file, use the output file with the suffix _.amber.score.mol2 generated from prepare_amber.pl (see (ii) above)

The following options are amber score specific option:

```
amber_score_primary                                          yes
amber_score_secondary                                        yes
receptor_file_prefix                                         1lgu
amber_score_movable_region                                   ligand
amber_score_gb_model                                         5
amber_score_md_steps                                         1
amber_score_minimization_cycles                              1
amber_score_nonbonded_cutoff                                 18.0
amber_score_temperature                                      300.0
amber_score_verbose                                          no
```

For receptor_file_prefix, use the prefix of the receptor PDB file. For example in this case it is 1lgu for our pdb file 1lgu.pdb
Choose amber_score_movable_region as ligand. This defines the region that is allowed to move while scoring. To select other options, please read the manual.

****
dock.in file

```
ligand_atom_file                               lig.amber_score.mol2
ligand_outfile_prefix                                        output
limit_max_ligands                                            no
read_mol_solvation                                           no
write_orientations                                           no
write_conformations                                          no
skip_molecule                                                no
calculate_rmsd                                               no
rank_ligands                                                 no
num_scored_conformers_written                                1
orient_ligand                                                no
flexible_ligand                                              no
bump_filter                                                  no
score_molecules                                              yes
contact_score_primary                                        no
contact_score_secondary                                      no
grid_score_primary                                           no
grid_score_secondary                                         no
chemgrid_score_primary                                       no
chemgrid_score_secondary                                     no
continuous_score_primary                                     no
continuous_score_secondary                                   no
gbsa_zou_score_primary                                       no
```

```
gbsa_zou_score_secondary                                     no
gbsa_hawkins_score_primary                                   no
gbsa_hawkins_score_secondary                                 no
amber_score_primary                                          yes
amber_score_secondary                                        yes
receptor_file_prefix                                         1lgu
amber_score_movable_region                                   ligand
amber_score_gb_model                                         5
amber_score_md_steps                                         1
amber_score_minimization_cycles                              1
amber_score_nonbonded_cutoff                                 18.0
amber_score_temperature                                      300.0
amber_score_verbose                                          no
```