SPHERE_SELECT

Andrew Magis
University of Florida

Sphere_select is my version of the DOCK suite program sphere_selector. I wrote it primarily to gain more flexibility in selecting spheres and to divorce the user from having to use strictly MOL2 format coordinate files.

Sphere_select allows the user to select spheres in several different ways. MOL2 coordinate files can be used as inputs to select spheres around a known binding ligand, as with the original sphere_selector. **Sphere_select** also reads coordinate files in PDB format, and automatically detects the file format. Spheres can also be selected around a specific atom in either the MOL2 or PDB coordinate file. This is somewhat different than selecting spheres based on a ligand. In this case, the user knows or suspects a binding site, but lacks a MOL2 ligand coordinate file with which to select spheres. In this case the original PDB (or MOL2) file of the receptor can be provided as a coordinate file, and an atom can be specified around which to select spheres. The atom number exists in column 1 of the TRIPOS MOL2 format and column 2 of the PDB file format. Finally, the user can specify a specific sphere around which to select other spheres. In this case the coordinate file is ignored. This might be useful in a situation where the user knows or suspects a binding site that has been populated with spheres, and simply picks a sphere in approximately the correct place to select a group of spheres for DOCKing calculations.

To make sphere_select, type:

```
%>make
```

To clean up executable and object files, type

```
%>make clean
```

Sphere_select accepts the following command line inputs:

-i <inputfilename>
    The name of the sphere file to be read
-r <radius>
    The radius in angstroms within which to select spheres

-o <outputfilename> *(optional)*

The output file name in which to write the selected spheres. Defaults to *selected_spheres.sph*.

-c <coordinatefilename> *(optional)*

The coordinate file used to select the spheres. Can be either MOL2 or PDB format, and can be either ligand with known binding location or original receptor file for selecting a specific atom.

-s <sphere_number> *(optional)*

This input is used to specify a specific sphere around which to select other spheres within the specified radius.

-a <atom_number> *(optional)*

When receptor coordinate file is provided, this input is used to specify a specific atom around which to select spheres within the specified radius.

-p *(optional)*

Write out a PDB sphere file as well as the sphere file for easy visualization of the selected spheres. Filename is *<outputfilename>.pdb*.

Sample Inputs

./sphere_select –i spheres.sph –c my_ligand.mol2 –r 10.0
Selects spheres from spheres.sph within 10.0 angstroms of the atoms in coordinate file my_ligand.mol2

./sphere_select –i spheres.sph –c receptor_file.pdb –a 100 –r 10.0
Selects spheres from spheres.sph within 10.0 angstroms of atom 100 in receptor_file.pdb. Receptor_file.pdb should be the same file you used to generate the molecular surface.

./sphere_select –i spheres.sph –s 10 –r 7.5 –p
Selects spheres from spheres.sph within 7.5 angstroms of sphere 10 in sphere.sph. Writes out spheres to selected_spheres.sph and automatically generates PDB file selected_spheres.sph.pdb.

Please note the instructions are only guidelines. Sphere_select can also be used to select spheres around a specific atom from any coordinate file in PDB or MOL2 format, including specific atoms of ligands with known binding modes, as long as the coordinates are relative to the coordinates of the spheres in the sphere file. If no spheres are selected, check your coordinates!

Tip: Use a visualization program such as Pymol or Chimera to find the atom or sphere around which to select other spheres. You can identifiy the amino acid or atom easily this way. If you can only identify the amino acid number, open up your coordinate file in a text editor, find the correct amino acid number, and pick an atom number from that amino acid around which to select spheres. For example, say you used Pymol to identify TYR 868 as an amino acid on the active site of your receptor. Open up the receptor PDB file and find TYR 868:

```
ATOM    209  N   TYR A 868      69.078  51.312  21.504  1.00 36.83
ATOM    210  CA  TYR A 868      68.013  50.751  20.682  1.00 38.45
ATOM    211  C   TYR A 868      67.688  49.353  21.179  1.00 39.17
ATOM    212  O   TYR A 868      68.429  48.404  20.921  1.00 38.94
ATOM    213  CB  TYR A 868      68.369  50.739  19.191  1.00 39.30
ATOM    214  CG  TYR A 868      67.193  50.351  18.316  1.00 40.81
ATOM    215  CD1 TYR A 868      66.064  51.170  18.233  1.00 42.54
ATOM    216  CD2 TYR A 868      67.198  49.163  17.585  1.00 41.54
ATOM    217  CE1 TYR A 868      64.973  50.821  17.443  1.00 42.43
ATOM    218  CE2 TYR A 868      66.108  48.803  16.787  1.00 42.80
ATOM    219  CZ  TYR A 868      65.001  49.638  16.724  1.00 42.54
ATOM    220  OH  TYR A 868      63.920  49.300  15.941  1.00 43.21
```

Now simply pick an atom from the amino acid such as # 211, and use this to select spheres around Tyrosine 868.

Here is the same amino acid from the same receptor, but this time in MOL2 format:

```
209 N          69.0780   51.3120   21.5040 N.am     26 TYR868
210 CA         68.0130   50.7510   20.6820 C.3      26 TYR868
211 C          67.6880   49.3530   21.1790 C.2      26 TYR868
212 O          68.4290   48.4040   20.9210 O.2      26 TYR868
213 CB         68.3690   50.7390   19.1910 C.3      26 TYR868
214 CG         67.1930   50.3510   18.3160 C.ar     26 TYR868
215 CD1        66.0640   51.1700   18.2330 C.ar     26 TYR868
216 CD2        67.1980   49.1630   17.5850 C.ar     26 TYR868
217 CE1        64.9730   50.8210   17.4430 C.ar     26 TYR868
218 CE2        66.1080   48.8030   16.7870 C.ar     26 TYR868
219 CZ         65.0010   49.6380   16.7240 C.ar     26 TYR868
220 OH         63.9200   49.3000   15.9410 O.3      26 TYR868
```

Again, atom 211 could be chosen. Notice how the atom number is in the first column instead of the second.

To select spheres, open the sphere file in a text editor and pick the sphere number you want:

```
cluster     1   number of spheres in cluster   50
    2  68.31599  48.55595   7.76092   3.176  598 0  0
    3  61.00529  49.00800  11.49000   2.006   26 0  0
```

```
4  62.24938  47.76450  13.37639   1.798   26 0  0
5  60.26475  49.06197  10.52271   2.550   29 0  0
8  58.37298  50.28647   8.88207   3.231   29 0  0
```

In this case, you might choose sphere 4. It is easiest to pick spheres by first converting a cluster to PDB using the DOCK program showsphere, opening them in a visualization program along with the receptor file, and finding the sphere you want. Double clicking on a sphere in Pymol will show you the sphere number.

Sphere_select has been compiled using the GNU g++ compiler and tested on Linux CentOS 4.1 and Solaris.

This software should be considered in beta testing. It may have bugs. If you find one, or if you have any questions, comments, or suggestions, feel free to contact the author at andrewmagis@gmail.com.