

-- X. Zou, 6/28/1999 --

Questions on installation of DOCK4.0.1:

=====

Q. Can I run DOCK on platforms other than SGI (e.g., SparcStations, DEC Stations, Pentium, etc.)?

A. We are releasing the beta versions for Window, Linux and hp machines (thanks Scott Pegg, David Konerding and Wolfgang Visser). We also encourage DOCK users to share their experience with us on how to transport DOCK to other platforms.

Q. How to compile DOCK on SGI IRIX 6.4 or 6.5(Octane)?

A. Versions released after March 3, 1999 should have no problem to be compiled or executed on octanes using IRIX 6.4 or 6.5. For versions released earlier, insertion of the following line on Line 54 of grid.c will solve the problem:

```
bzero(&receptor, sizeof(receptor));
```

Q. Sometimes I have got warning messages when I compile DOCK4.0. Shall I ignore them or not?

A. No warning messages appear when we compile on our local machines. Maybe some object files were left over from your previous compilation. You can ignore the warnings.

Q. I missed some of the executable files in my bin directory after installation and I had no particular errors during compilation. Where goes wrong?

A. You failed to compile DOCK in the correct sub-directory. You have to go to the "source" sub-directory of the software and type 'make install'.

## Questions on running DOCK4.0.1

=====

Q. What are the basic stages for running DOCK?

A. The docking procedure can be divided into four general stages: ligand preparation, site characterization, scoring grid calculation, and docking itself. A good starting exercise is to run the demo provided with DOCK.

Q. Could you send me some references citing the DOCK methodology just to have some ideas concerning the parameters to use?

A. 1. Ewing, TJA; Kuntz, ID.

Critical evaluation of search algorithms for automated molecular docking and database screening.

JOURNAL OF COMPUTATIONAL CHEMISTRY, 1997 JUL 15, V18 N9:1175-1189.

2. MENG EC; SHOICHET BK; KUNTZ ID.

AUTOMATED DOCKING WITH GRID-BASED ENERGY EVALUATION.

JOURNAL OF COMPUTATIONAL CHEMISTRY, 1992 MAY, V13 N4:505-524.

3. KUNTZ ID; MENG EC; SHOICHET BK.

STRUCTURE-BASED MOLECULAR DESIGN.

ACCOUNTS OF CHEMICAL RESEARCH, 1994 MAY, V27 N5:117-123.

4. The manual of DOCK4, especially,

1) the 'Overview' article in the Introduction Chapter,

2) the chapter 'Users Guide', which describes detailed procedures and optimal DOCK parameters.

Q. How to prepare a mol2 file for a single ligand (or several ligands)?

A. An inhibitor from a crystal structure in the Brookhaven Protein Data Bank is used here as an example.

Step 1. The inhibitor is separated from the protein, water molecules, ions, and etc., and saved in a pdb file, say, 'lig.pdb'.

Step 2. This pdb file is converted into the SYBYL MOL2 format by the use of 'pdb2syb', a script in the "bin" sub-directory of DOCK,

creating a mol2 file, say, 'lig\_tmp.mol2'.

Step 3. The SYBYL software is used to fix atom types, to add hydrogen atoms, to compute Gasteiger charges, to rename atoms sequentially, and to rename the substructure and molecule.

The clean ligand is saved in a mol2 file, say, 'lig.mol2'.

Step 4. This step is optional. "lig.mol2" is converted into the pdb format "lig.pdb" by the use of the accessory script "convsyb" or the command "dock -i lig.mol2 -o lig.pdb". "lig.pdb" is used for Midas display or for generation for the molecular surface of the binding site on the receptor.

For other formats of ligand files, see the Q&A section below on file conversions.

Q. How to prepare mol2 files for a ligand database?

A. Ligand database files need to be in SYBYL mol2 format. If the database is in SD-format, e.g., ACD distributed by Molecular Design Ltd. (San Leandro, CA), it can be converted into sybyl-format files by "sdf2mol2" and "sybdb" provided in the DOCK accessories. This conversion requires SYBYL to assign atom types and charges.

If the user has the ACD license from MDL, we can provide converted ACD database in SYBYL mol2 format.

Q. How to prepare a mol2 file for a receptor from the Brookhaven Protein Data Bank?

A. Step 1. Keep only the receptor (protein) component in the Brookhaven pdb file and save in different a file.

Step 2. Use SYBYL (the Biopolymer module) to complete the missing residues. (Save the current molecule with heavy atoms only in a "Brookheaven" pdb file, say, "rec\_heavyatom.pdb", for later generation of the molecular surface.)

Step 3. Use SYBYL (the Biopolymer module) to add hydrogens and to load AMBER (Kollman) charges.

Step 4. Save as a mol2 file.

Q. How to generate the molecular surface of the receptor?

A. The molecular surface which will be used to produce sphere sites is

generated by the program `ms`, available from Quantum Chemistry Program Exchange (QCPE). The simplest way is to generate the molecular surface for the whole receptor. The command is

`"ms rec_heavyatom.pdb -a -p -n -o rec.ms -g rec.mslog"`, or even simpler, `"ms rec_heavyatom.pdb -n -o rec.ms"`. The flags in the `ms` command have the following meanings:

- a use all atoms, not just amino acids
- g send messages to file
- n calculate normals for surface points (necessary for "sphgen")
- o specify output file name (required)

If a user has the access to "dms" (a program provided by Midas), run `"dms rec_heavyatom.pdb -a -n -o rec.ms -g rec.mslog"`.

Q. How to generate the molecular surface of the receptor site of interest?

A. If one focuses on the binding site of the receptor, the following procedures are recommended to avoid wasting long CPU time on generation of the molecular surface for the whole protein:

Step 1. The accessory program "get\_near\_res" is used to generate a list of all residues in the receptor ("rec\_heavyatom.pdb") with any atom within a specified distance of the ligand ("lig.pdb").

Step 2. This list is compacted with the accessory program "condense" to a file, say, "rec\_near.lst2".

Step 3. "rec\_near.lst2" is suitable for input to Connolly's "ms" (UCSF version, NOT sure about the QCPE version):

`"ms rec_heavyatom.pdb -a -p -n -i rec_near.lst2 -o rec.ms -g rec.mslog"`

If a user has the access to "dms" (a program provided with Midas), run `"dms rec_heavyatom.pdb -a -n -i rec_near.lst2 -o rec.ms -g rec.mslog"`.

For details on how to run "get\_near\_res" and "condense", read the Q&A section below on DOCK accessories.

Q. How to create spheres to represent the receptor site?

A. Use the program "sphgen" in the "bin" subdirectory of DOCK. "sphgen" requires an input file, "INSPH". An example of the "INSPH" is given in the section 3.3 of the DOCK manual ("Reference Manual"/"SPHGEN"). Another example is given in the demo of DOCK (demo/3dfr/2\_site/INSPH).

Q. Why do I get different energy scoring results with different random seed numbers?

A. Yes, the choice of the random seed will slightly affect the scoring result. This is a feature of the minimization algorithm called the 'downhill simplex method', developed by Nelder and Mead, that is used by DOCK. This simple global minimum searching algorithm relies on the initial starting point, which is determined by the random seed. DOCK uses this simplex method to minimize the ligand orientation to find the best contact, chemical or energy scores. So we recommend users to try three different random seeds in order to find the "real" minimum.

For detail about the simplex method, please read the book 'Numerical Recipes: The Art of Scientific Computing' by Press, Flannery, Teukolsky, and Vetterling. (1986, Cambridge University Press, C edition pp. 292-293) Please also refer to the file simplex.c in directory 'dock4/source/dock' under your DOCK4 source code directory.

Q. When I ran database search, I found the "Energy score" is different from the sum of intramolecular component and intermolecular component. However, this does not happen for single search. Why?

A. Because we don't know what the minimum intramolecular score for a molecule is (without doing complete conformation samplings), we cannot make comparisons of intramolecular score between DIFFERENT molecules. However, because this "ground-state" intramolecular score is the same for all conformations of a single molecule, we can compare intramolecular scores for different conformations of the SAME molecule. Therefore, in DOCK4,

1. for single search:

energy score = intermolecular component + intramolecular component

2. for database search,

1) all the conformations of a single molecule are compared using intra + inter, and only the conformation with the best score is saved for comparison to other molecules.

2) all of these "best" conformations (1 from each molecule) are compared to one another based ONLY on the intermolecular score.

energy score = intermolecular component

For both single search and database search, we always have

intermolecular component + intramolecular component

= van der Waals component + electrostatic component

Please refer to the section 2.2.3 of the DOCK manual  
("Users Guide"/Advanced Techniques"/"Scoring") for detail.

Q. Can you explain the chemical definitions in the flex.defn file such as  
"definition .3 [ 3 H ]" and "defintion C.2 ( O.2 ) ( C.3 )"?

A. For an explanation of the nomenclature used in flex.defn please refer to  
Table 7-9 in the section 3.6 of the dock4 manual  
("Reference Manual"/"Parameter Files").

Q. In flexible docking, what are the recommended values for  
"maximum\_orientations" (named as "total\_orientations" in DOCK4.0) and  
"configurations\_per\_cycle" (named as "peripheral\_seeds" in DOCK4.0)  
for single ligand and database search respectively?

A. There are no exact numbers. Try 500 (single), 100 (database) for  
"maximum\_orientations" and 50 (single), 10 (database) for  
"configurations\_per\_cycle". Choose the largest values that you can afford.

Q. Why ligand\_energy\_file contains always a maximum of 10 molecules even if  
rank\_orientation\_total parameter is set to 100 ?

A. Because you have to set the "configurations\_per\_cycle" = or > 100 too.  
"configurations\_per\_cycle" determines # of conformations/orientations  
to seed to the next level of the orientation search in the anchor-first  
algorithm. If "configurations\_per\_cycle" = 10, only 10 orientations are  
saved (even though you set rank\_orientation\_total =100) .  
Please refer to Section 2.3.2 ("Users Guide/Advanced Techniques/Conformation  
Search") and Section 3.1, Table 3B of the manual for detail.

Q. Is chemical scoring generally better than other scorings?

A. Chemical scoring is under development. You should use only at your own risk. No claim is made for its accuracy.

Questions on DOCK accessories:

=====

Q. How to use "get\_near\_res"?

A. Type "get\_near\_res". Input the following parameters as asked:

lig.pdb            (ligand PDB file)  
rec\_heavyatom.pdb (receptor PDB file which contains no hydrogen atoms)  
7                  (distance cutoff in Angstroms)  
1                  (residues with ANY atom w/in cutoff of reference)  
y                  (allow hydrogens atoms as close contacts)  
rec\_near          (prefix of output file)

The program creates two output files, rec\_near.lst and rec\_near.pdb.  
"rec\_near.lst" is later used by the program "condense".

Q. How to use "condense"?

A. Type "condense". Input the following parameters as asked:

rec\_near.lst (file to be condensed, i.e., output of "get\_near\_res")  
rec\_near.lst2 (output)

The output file "rec\_near.lst2" is used as a 'ms -i' file for the UCSF  
ms program.

Q. When I ran "autoMS", it failed to work unless I changed Line 112  
in the script file from "\$MS\_executable" to  
"\$MS\_executable fort.1 fort.2 fort.3". Any comment?

A. This may depend on the version of qcpe\_ms one uses. We do not encounter  
this problem on our local machines. But for QCPE\_429, the modification  
seems to be requisite according to observation of two DOCK users

(Takahashi Mitsuo Ajinomoto and Ed Moret).

Q. The sybdb script needs two modifications to correctly handle substructures:

- 1) The four `\$delsubs` should be replaced by `\"$delsubs"`.
- 2) Line 152 should be changed from `'if %gt(\$nsub 1)'` to `'if %gt(\$nsub 2)'`.

A. Yes, these are the two bugs in sybdb that have been fixed. (Thanks to Ed Moret and Jason McDermott.)

Questions on file conversions:

=====

Q. How to make a ptr file from a mol2 database?

A. Use the command `"dock -i database.mol2 -o database.ptr"`.

Q. Do I need SYBYL to convert a pdb file to mol2 file? Can I use the script `"pdb2syb"` or `"idtosyb"` provided by DOCK accessories instead?

A. Yes, the easiest way is to use SYBYL to convert a pdb file to a SYBYL mol2 file. Please refer to p18-22 in the manual for detail. Or you can write a script of your own to do the conversion. Examples of the SYBYL mol2 format are provided in the demo sub-directory.

The scripts `"pdb2syb"` and `"idtosyb"` prepare files for DOCK3.5. The converted mol2 format cannot be recognized by DOCK4. These scripts help assign atom types and bonds for ligand molecules.

Q. Can I use the mol2 format from INSIGHTII?

A. Yes.

Q. When I use `"sdf2mol2"` and `"sybdb"` to convert a sd format database file into a SYBYL mol2 database file, the names of the molecules get lost, and instead, it writes out the line that tells number of atoms as a name. What's wrong?

A. The problem is due to the format of your sd file. The sdf2mol2 command is yielded from a fortran program (sdf2mol2.f in the sub-directory /source/database/sdf2mol2 under your software directory), and therefore requires very strict input format, unlike C programs.

There are two ways to solve your problem:

Method 1: Add a couple of lines started with "> <MOLNAME>" in your sd file. Here is one example. DO NOT remove the FIRST blank line. For several molecules, just append them together.

..... cut here (an example of a sd file) .....

1234

ligandname

13 13 0 0 1

0.0000 0.0000 0.0000 C 0 0 0 0 0

0.8400 0.6200 1.0800 C 0 0 0 0 0

0.7200 0.2800 2.2400 O 0 0 0 0 0

1.7400 1.5800 0.7200 N 0 0 0 0 0

2.4200 1.4800 -0.5800 C 0 0 3 0 0

3.7600 2.1000 -0.4800 N 0 0 0 0 0

6.1800 2.5800 -1.5000 Cl 0 0 0 0 0

4.6400 1.6800 -1.5600 C 0 0 3 0 0

5.1400 -0.3600 -2.8400 N 0 0 0 0 0

4.8800 0.1400 -1.5000 C 0 0 3 0 0

3.7000 -0.6400 -0.8600 C 0 0 3 0 0

3.9800 -1.0000 0.8600 Cl 0 0 0 0 0

2.4200 0.0800 -1.0400 N 0 0 0 0 0

1 2 1 0 0 0

2 3 2 0 0 0

2 4 1 0 0 0

4 5 1 0 0 0

5 6 1 0 0 0

7 8 1 0 0 0

8 6 1 0 0 0

9 10 1 0 0 0

10 8 1 0 0 0

10 11 1 0 0 0

11 12 1 0 0 0

11 13 1 0 0 0

13 5 1 0 0 0

> <MOLNAME>

1234

> <MDLNUMBER>

ligandname

\$\$\$\$

..... cut here (end of the sd file) .....

Method 2: If you do not want to change the format of your sd file, you can edit the fortran program readSDF.f in the sub-directory source/accessories under your software directory. "readSDF.f", which reads in sd files, is a subroutine for "sdf2mol2.f".

Comment: Remember to remove all H atoms before you run the dbprocess unless you want to assign the H atoms in your own way. "sybdb" will automatically add H atoms and assign the correct formal charges (or protonation) assuming pH=7.