HP-SEE

OpenEye Scientific Software

www.hp-see.eu

Branko J. Drakulić
Department of Chemistry-IChTM, University of Belgrade
bdrakuli@chem.bg.ac.rs

HP-SEE

High-Performance Computing Infrastructure
for South East Europe's Research Communities

The HP-SEE initiative is co-funded by the European Commission under the FP7 Research Infrastructures contract no. 261499

Intro

HP-SEE
High-Performance Computing Infrastructure for South East Europe's Research Communities

- Founder Anthony Nicholls, physicist, 1997, SantaFe, NM

- Graduated in Oxford, work on DelPhi, author of GRASP

- OE staff list, with the brief bios http://www.eyesopen.com/staff

- Instantly OE is one of the lead software companies in the field of cheminformatics and molecular modelling

- Software distributed is the *industry standard*

- Platform-independent code(s) for *high-throughput* 2D and 3D modeling

-*Shape* and *electrostatics* as primary variables of molecular description

Licensing

HP-SEE
High-Performance Computing Infrastructure for South East Europe's Research Communities

- Free licenses are offered to academic research groups in *accredited degree-granting college or university* (as well as to 'garage thinkers').
- Head of the group must apply for the license, annual-basis
- On PARADOX (IPB) we established policy to track user license
- Details http://www.eyesopen.com/licensing-philosophy
- My opinion on the attitude of (the majority of) our scientific community on usage of the software can be found in

B.J. Drakulić, I.O. Juranić 'CompChem (RS) application in the HP-SEE project. Resources, source codes, strengths and **weaknesses**'
50th Meeting of the Serbian Chemical Society, Belgrade, June 14-15

Welcome!

Software in brief

HP-SEE
High-Performance Computing Infrastructure for South East Europe's Research Communities

- Installed: OMEGA, ROCS, EON, FRED, SZYBKI
- Command line input, output can be analyzed by native GUI - VIDA; or using 3rd party GUI's – all on **user terminal**
- Not all functions available if 3rd party GUI's is used, so depend on user need
- All programs can be run in both parallel and sequential mode
- Parallel, PVM or MPI
- On PARADOX MPI is applied
- PBS script(s)
- in following slides **basic execution** via command line shown
- **many 'fine tuning' options exist for all programs listed**

Script

HP-SEE
High-Performance Computing Infrastructure for South East Europe's Research Communities

```
#!/bin/bash
#PBS -q hpsee
#PBS -l nodes=<number of nodes>:ppn=<number of processors per node>
#PBS -l walltime=10:00:00 (for example 10 hours)
#PBS -e ${PBS_JOBID}.err
#PBS -o ${PBS_JOBID}.out

cd $PBS_O_WORKDIR
cat $PBS_NODEFILE
export OE_LICENSE=path_to_license_file
/path_to_/openeye/bin/oempirun -np <number of procesors> omega2 -in /nfs/user_name/input_file.smi -out test.oeb.gz
```

- many input file formats
- ‘oeb’ and ‘oeb.gz’ are native OE formats

OMEGA

- generate conformations, make alignment
- fast 1-2 sec / per molecule
- tens of thousands of compounds / day / processor
- input 1D (smiles), or 2D (for example mol2)
- distance bounds methods
- conformational assemblies can be used as the input for the
  ROCS (shape comparison),
  FRED (docking),
  pharmacophore based search.
- basic input:

```
$ -np 2 omega2 -in test.smi -out test.mol2
```

ROCS

- shape similarity tool
- input: conformational assembly (for example derived by OMEGA)
- input: template – molecule or pharmacophore pattern
- optimization of the best global match
- process 20-40 compounds / sec on 1 CPU
- hit ranking by both shape and pharmacophore similarity
- Tanimoto and Tversky similarity
- basic input:

```
$ -np 2 rocs -query template.mol2 -dbase test.mol2
```

- user terminal GUI version, vROCS, facilitate preparation of the input and analysis of the results

EON

- electrostatics comparison
- molecules are pre-aligned (typically by ROCS)
- electrostatic potential is calculated using Poisson-Boltzmann electrostatic tool (internal – by ZAP OE tool)
- charges of atoms are ascribed in-fly (MMFF),
- optionally charges can be provided by user
- two electrostatic similarity measures, both by Tanimoto
- I dielectric constant of surroundings 80
  II dielectric constant of surroundings 2
- both shape and electrostatic similarity is reported

EON

HP-SEE
High-Performance Computing Infrastructure for South East Europe's Research Communities

- What is needed ?
  - query file – on this molecule all molecules from database will be aligned
  - to obtain valid results use the same type of charges for the query and database molecules
  - by comparison of charges obtained on *ab initio* level (possible for query) and MM charges (typical for dbase) senseless results usually were obtained
- dbase – set of molecules for comparison
- prepare it by ROCS:

  add '-eon input true' in ROCS comm. line exc
- basic input:

  $ -np 2 eon -dbase eon_input.oeb.gz -query template.mol2

FRED

HP-SEE
High-Performance Computing Infrastructure for South East Europe's Research Communities

F.R.E.D. (Fast Rigid Exhaustive Docking)

- advantages:

- structure based primary
- combine ligand and structure based
- receptor shape (shape of the active site)
- consensus structure (use consensus of multiple scoring functions)
- MASC - Multiple Active Site Correction (score docking of the ligand in the database of common active sites)
- penalize ligand if fit well to majority of this sites - promiscuous
- Chemgauss score – Gaussian-based scoring function by OE
- Exhaustive Docking – score all possible positions of the ligand in the active site

FRED

HP-SEE
High-Performance Computing Infrastructure for South East Europe's Research Communities

- What is needed ?
  - ligand database, multiconformer
  - receptor file, that include:
    - structure of the target protein
    - location of the receptor site
    - shape potential grid (describe shape potential)
    - structure of the bound ligand (optional)
    - docking constraints (optional)
  - GUI FRED_RECEPTOR (for the user terminal) facilitate receptor preparation
  - arguments in the command line
  - basic input:

```
$ -np 2 fred -dbase ligands.oeb.gz -rec rec.oeb.gz
```

SZYBKI

- optimizes molecular structures *in situ*
- optimization with or without solvent effect
- important: optimize ligand within active site
- ligand entropy calculations in different environments
- binding energies by Poisson-Bolltzmann or Sheffield models
- refine protein part surrounding ligand (flexibility)
- (possibility) of using AM1BCC charges
- charges from input file can be also used
- basic input:

```
$ -np 2 szybki -p protein.mol2  -entropy QN  ligand.mol2
```

User view

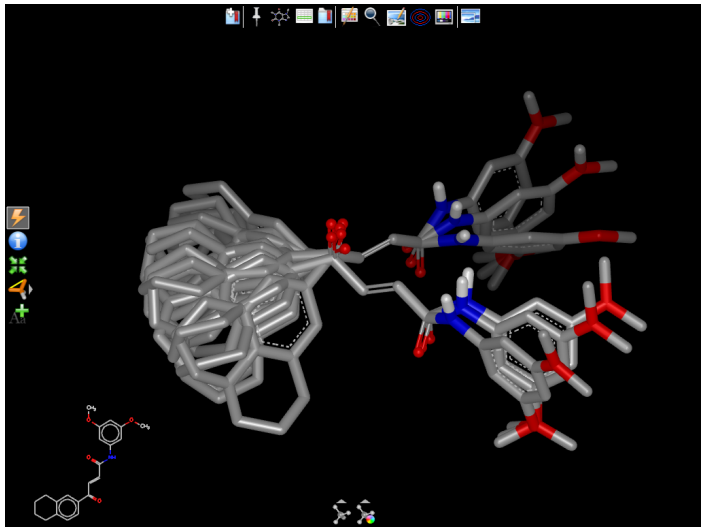
HP-SEE

High-Performance Computing Infrastructure for South East Europe's Research Communities

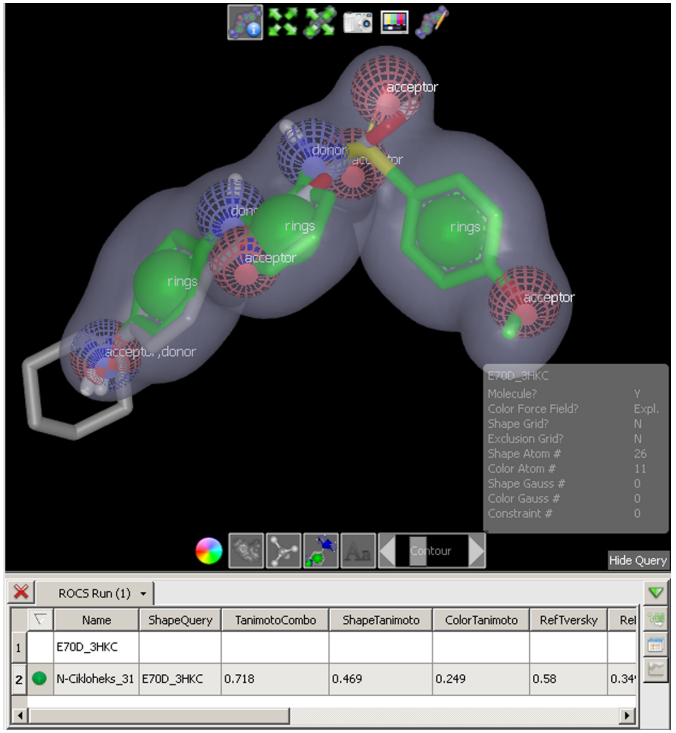
- all programs installed can be used to process large number of molecules on HPC facilities
- typically this is virtual screening
- some of programs, when used with options that allow high-quality output, are processor and memory demanding even on smaller set of compounds
- additional argument to be used with HPC
- users are encouraged to consult manuals, learn about programs functionality, and choose options that fit their needs
- version of programs that support GPU are available and will be installed when we upgrade PARADOX infrastructure
- (or incoming Blue Danube)

Some snapshots

- along with extensive text output, GUI offer comfortable work
- some common examples from our lab



- from OMEGA



- from ROCS

- shape similarity

Thank you

HP-SEE
High-Performance Computing Infrastructure for South East Europe's Research Communities

for the attention