OpenDiscovery: Automated Docking of Ligands to Proteins and Molecular Simulation

Gareth Price
Computational MiniProject
Aims + Achievements

• Produce a high-throughput protocol to screen a library of chemical compounds against a protein

• Use this protocol to produce some putative ligands for chemical synthesis and screening
  • (Miniproject III)

• Paper submitted to ACS JCE
  • “The use of freely available and Open Source tools for in silico screening in Chemical Biology”

• Ground work for OpenDiscovery:
  • Soon will be found at: www.opendiscovery.org.uk
Broad Outline

- Prepare Files
- Screen
- Simulate
- Parameterise
Generating Similar Compounds

SMILES Code

Normalise + Fragment

Search against database

Give X similar compounds based on Tanimoto distance

Stepping out of the current chemical “space”
oBabel is very useful at converting between chemical file formats
- and is the most popular tool of choice
- Also allows a conformer search, so that the energy minimum can be found
  - random, weighted or systematic (exhaustive) search
- Installs with OBMinimize, an *in vacuo* minimisation tool
Vina actually requires PDBQT
- i.e. with atomic coordinates (PDB), partial charges (Q) and atom types (T)

This was achieved by using AutoDock Tools

MGLTools (creators of Vina) provide a python script to do this non-interactively
- automate using python/shell scripting

PDBQT files for both receptor and ligand

Choice between rotatable, non-rotatable and unrotatable bonds
Set this to 1 to use Angstrom scaling
Broad Outline

1. Input File
2. Similar Compounds
3. SMILES
4. Conformer Search (OBabel)
5. MOL
6. Conformer Search (OBabel)
7. MOL2
8. PDB
9. Minimization (OBMinimize)
10. PDBQT
11. Solvate
12. Simulate! (NAMD etc.)
13. PSF
14. psf_gen.pgn
15. Parameterize ligands
16. Screen! (VINA)
17. Visualize (VMD)
18. Extract Modes
19. Summarize Results
20. Visualize (VMD)
Running Vina

- receptor and ligand files
- output PDBQT with 1+ modes
- box coordinates and dimensions
- exhaustiveness \( \propto \) amount of time devoted to finding accurate conformation and position

```
vina
  --receptor protein.pdbqt
  --ligand ligand.pdbqt
  --out all.pdbqt
  --center_x 14.45
  --center_y -24.772
  --center_z 15.079
  --size_x 10
  --size_y 10
  --size_z 10
  --log log.txt
  --exhaustiveness 600
```
Results

• Binding Affinity Energies

• One PDBQT file that has 1+ modes
  • i.e. a single file with the significant conformations + positions
  • can’t open in VMD or Chimera
    • VMD refuses and Chimera opens all modes all at once
  • but PyMol does work!

• Use Scripting/PyMol
  • (again!)... to extract modes
- A relatively new method of **clustering docking results**
- based on contact analysis
Broad Outline

- Input File
- Similar Compounds
- SMILES
- Conformer Search (OBabel)
- MOL
- MOL2
- Conformer Search (OBabel)
- PDB
- Minimization (OBMinimize)
- PDBQT
- Parameterize ligands
- PSF
- psf_gen.pgn
- Simulate! (NAMD etc.)
- Solvate
- Screen! (VINA)
- Extract Modes
- Visualize (VMD)
- Summarize Results
- Screen
- Visualize
- Summarize
- Results
A big problem

• Most forcefields only include definitions for amino acid residues
  • i.e. only bonds etc. found in proteins
• Our chemical compound ligands are not proteins
• So we need a way to “atom-type” each atoms to a peptide version
  • …or derive new parameters from ab initio calculations
Parameters

- Bond Stretching
- Angle Bonding
- Bond rotation (torsion)
- Non-bonded interactions (electrostatic)
- Non-bonded Interactions (Van der Waals)

Non-bonded interactions can be classified as either electrostatic or Van der Waals forces.
Gaussian, SwissParam and MATCH

- Gaussian is a program that can perform ab initio quantum mechanics calculations

- 3 problems:
  1. Not all the results will be correct
  2. It doesn’t give all values (energies, impropers)
  3. It takes ages and extracting information is very manual
Amber + Amber Tools

• Amber is available for a nominal fee (400$), but their tools are free
  • antechamber and LeAP are very useful
  • includes solvation
  • parameter files can also be used in other programs too

• Amber is a viable option but it is a long, convoluted process that requires a user who knows what they are doing
  • compilation, testing and set up of Amber/Tools is non-trivial – we need something better (highlighted in our paper)
Sample Results
Further Work and Acknowledgements

• Combine the screening (ODScreen) and parameterisation/simulation (ODParam) with cheminformatics
  • do we know anything about these compounds already?
  • do we want to know them? (if there is no information - probably no patent!)

• Implement other analysis tools
  • other than AuPoSOM

• I would like to thank MOAC, EPSRC, Dr. Andrew Marsh and Dr. Phil Gould for their support and guidance throughout the project
  • Also, Dr. Dan Rathbone (Aston), Dr. Alex Simperler (Imperial), Dr. David Bray and Prof. Mark Rodger (Warwick)
Software Used

• AutoDock Vina*
• AutoDockTools*
• AutoDockScripter (own software) *
• OBabel*
• MacPyMol*
• UCSF Chimera*
• Avogadro*
• NAMD*
• Gaussian
• GaussView
• ChemNProp*
• AuPoSOM*
• Amber + Amber Tools *
• GROMACS*
• Spartan10