# Open Discovery: Automated Docking of Ligands to Proteins and Molecular Simulation

**Gareth Price** (gareth.price@warwick.ac.uk); Phil Gould; Andrew Marsh

*Automated computational docking of large libraries of chemical compounds to a protein can aid in pharmaceutical drug design, and gives scientists with basic computer experience a tool to help plan wet laboratory investigations when exploring the combination of chemical and pharmacological spaces. This project aimed to: (a) produce a generalised, high-throughput protocol that can dock a library of small chemical ligands to a receptor protein, using Open Source or freely available software, (b) use this protocol to direct chemical synthesis and further analysis for a specific protein of interest (Miniproject 3). We have released the code at www.opendiscovery.org.uk.*

## 1. Similar Compound Generation

Starting with a two dimensional text representation (SMILES string) of a known ligand for a receptor, we used the ChemNProp[1] server to select a library of 50 similar compounds from PubMed (>26M structures). **Fig 1** explains the process.

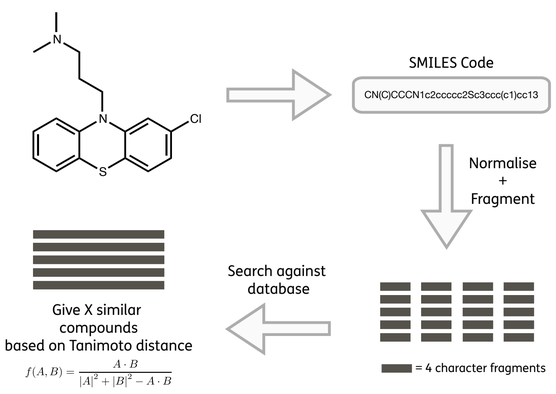



Figure 1. Creating similar chemical compounds through combinations of similar fragments.

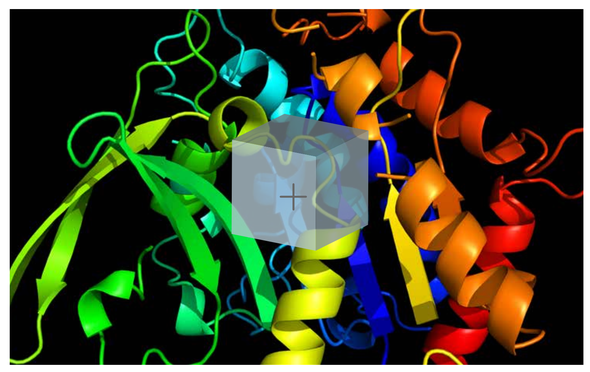


Figure 2. AutoDock Vina requires a three-dimensional box, into which the ligand is placed and tested.

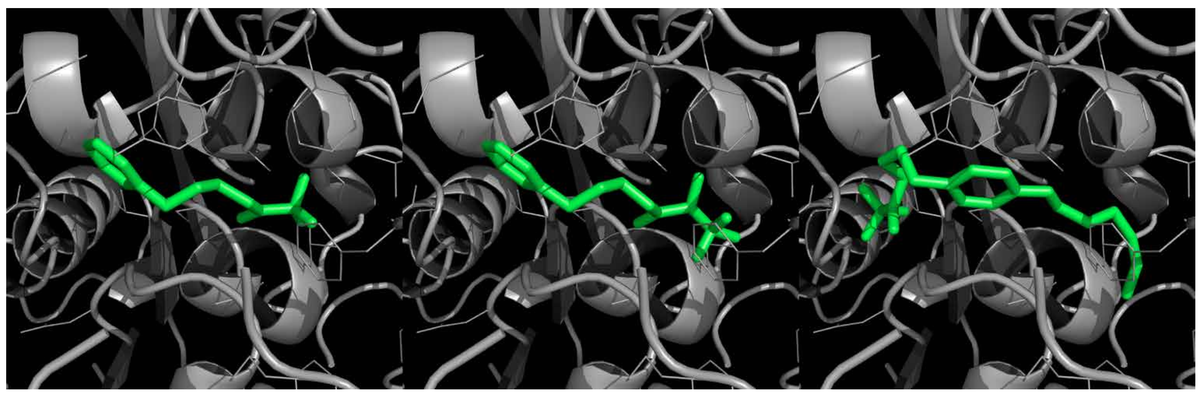


Figure 7. Three of the top ligands after docking and molecular dynamics simulations.

## 8. Conclusions

**Fig 7** shows three of the ligands that would be selected for wet lab analysis. This process has been undertaken for our protein of interest, and four have been selected for synthesis and further investigation. Due to the generality of Open Discovery, the process can be repeated for any protein and any library of chemical ligands.

## 7. Molecular Dynamics Simulation

The parameterised protein-ligand complex, is put in a box of water (**Fig 6**), minimised and equilibrated. Inspection of the active site reveals which interactions are significant: a chemist's "eye" is still important!

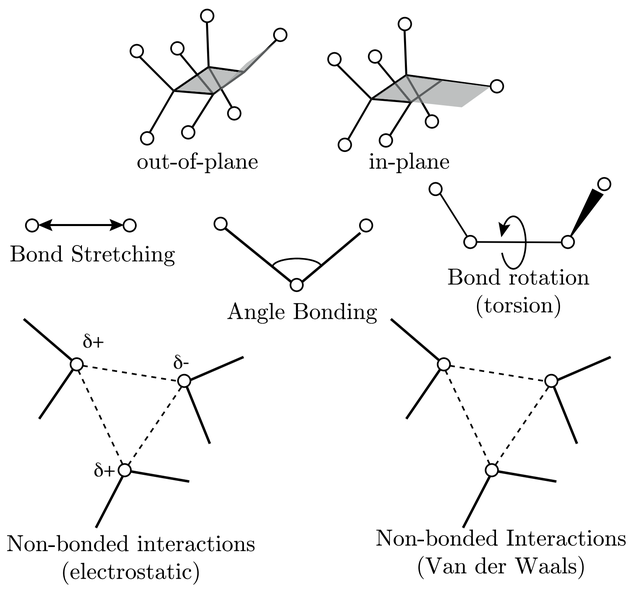
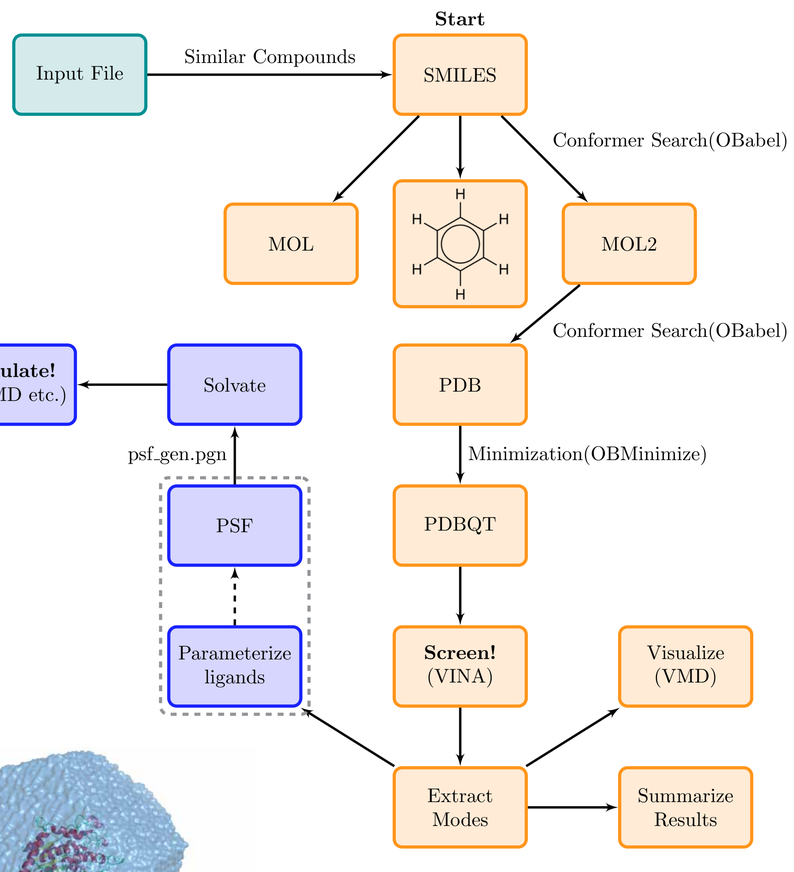



Figure 5. These are the main parameters in forcefields that are not present for the ligands.

## 6. Ligand Parameterisation for Molecular Dynamics

Additional confidence in the docking runs would come from molecular dynamics simulations using an explicitly solvated, fully flexible receptor. Parameters such as bond lengths, angles and dihedrals (**Fig 5**) for all components are required, however the forcefields typically used in the simulations only describe amino acids, sugars and lipids. Therefore parameterisation of the ligands is necessary. Although there are several free web services that claim to do this, these do not fit into the workflow and in our experience do not solve subsequent steps. *Ab initio* quantum mechanical calculations were also tested, but it is challenging to extract the required information from the output and will prove computationally intensive for large compound libraries

The freely available `AmberTools` (part of the commercially available Amber MD[5] package) was thus used to parameterise the ligands in an automated manner. We highlight the need for a better way to bridge the gap between docking and simulation in a submitted paper[6].



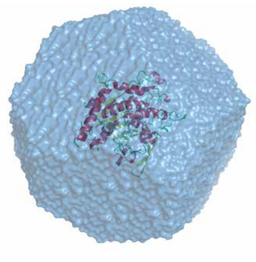



Figure 6. The protein is placed in a box of water, ready for molecular simulation.

## 2. Open Babel - The Chemist's File Converter

Open Babel[2] was used to convert between the many file types that are used within the protocol. It allows a conformer search, to find the energy minimum and then performs *in vacuo* minimisation of the compounds. A PDBQT file is produced, using a `python` script provided by the authors of the docking software, AutoDock Vina[3], which adds partial charges and atom types to the PDB format.

## 3. Receptors and Grid Box

Receptors are the proteins that the ligands are docked into. These need explicit hydrogens and must be in the PDBQT format.

The grid box (**Fig 2**) refers to the three-dimensional area of the receptor where the ligands will bind to (i.e. the active site). If not known, this can be explored using other methods by a grid box the size of the protein.

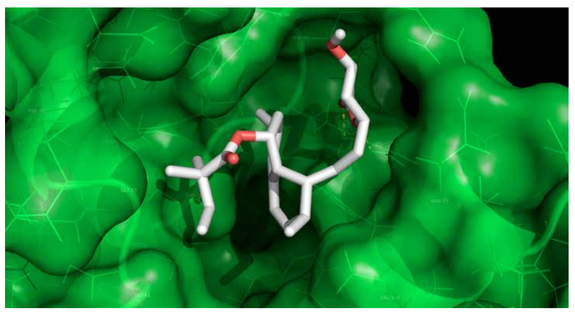


Figure 3. Example binding mode of a ligand to its protein receptor.

## 4. Ligand Docking

AutoDock Vina requires little configuration, includes a rotamer search of the ligands, flexible side-chains and is computationally inexpensive. It is run on the command line, but a GUI was produced for single-ligand runs (AutoDockScripter, available at www.gprice.co.uk/ads). Results like **Fig 3** can easily be created using software such as PyMOL.

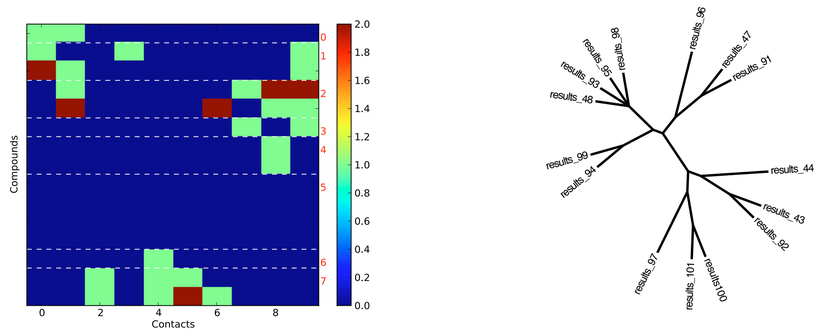



Figure 4. Example clustering of ligands using AuPoSOM. Here, 15 ligands are clustered into 8 leafs, with 3 having high contacts.

## 5. AuPoSOM

AuPoSOM[4] is post-processing analysis using a neural network to produce a self-organising map that clusters ligands on the basis of the contacts that the ligands make with the receptor protein (**Fig 4**).

**Acknowledgements**

I gratefully thank the ESPRC for funding and Dr. Andrew Marsh and Dr. Phillip Gould for their guidance throughout the project.

**References**

1. ChemNProp http://chemnprop.irbbarcelona.org (accessed Apr 2013)
2. The Open Babel Package, version 2.3.1 http://openbabel.org (accessed Apr 2013)
3. AutoDock Vina is available at http://autodock.scripps.edu (accessed Apr 2013)
4. Bertho, G.; Mantsyzov; Bouvier; Evrard-Todeschi Contact-based ligand-clustering approach for the identification of active compounds in virtual screening. AABC 2012, 5, 61.
5. D.A Case et al. AMBER 12, University of California, San Francisco. 2012
6. Price, GW.; Gould, PS.; Marsh, A.; The use of freely available and Open Source tools for in silico screening in Chemical Biology. Currently submitted to ACS JCE

THE UNIVERSITY OF WARWICK

MoAC

EPSRC Pioneering research and skills