Useful Molecular Modelling and Drug Design Softwares and Databases

Compiled by Bhakat Soumrndranath (bhakatsoumendranath@gmail.com) - Dr Soliman's Molecular Modelling and Drug Design Lab

Databases

Chemical Structure Databases:

ZINC: a free database of commercially-available compounds for virtual screening. ZINC contains over 21 million purchasable compounds in ready-to-dock, 3D formats. ZINC is provided by the Shoichet Laboratory in the Department of Pharmaceutical Chemistry at the University of California, San Francisco (UCSF). To cite ZINC, please reference: Irwin, Sterling, Mysinger, Bolstad and Coleman, *J. Chem. Inf. Model. 2012* DOI: 10.1021/ci3001277.

ChEMBL: It is a database of bioactive drug-like small molecules, it contains 2-D structures, calculated properties (e.g. logP, Molecular Weight, Lipinski Parameters, etc.) and abstracted bioactivities (e.g. binding constants, pharmacology and ADMET data).

ChemSpider: *ChemSpider* is a free chemical structure database providing fast text and structure search access to over 28 million structures from hundreds of data sources. ChemSpider SyntheticPages, CS|SP, extends this model to cover reactions, providing quick publication, peer review and semantic enhancement of repeatable reactions. Maintained by: Royal Society of Chemistry

Drug Bank: The DrugBank database is a unique bioinformatics and cheminformatics resource that combines detailed drug (i.e. chemical, pharmacological and pharmaceutical) data with comprehensive drug target (i.e. sequence, structure, and pathway) information. The database contains 6712 drug entries including 1448 FDA-approved small molecule drugs, 131 FDA-approved biotech (protein/peptide) drugs, 86 nutraceuticals and 5079 experimental drugs and many more. Supported By: Genome Alberta & Genome Canada. This project is also supported in part by GenomeQuest, Inc.

PubChem: Database of molecules with their properties,bioassay result,3D structures and with many more informations. Copyright: National Center for Biotechnology Information (NCBI)

Approved Drugs: The Approved Drugs app contains over a thousand chemical structures and names of small molecule drugs approved by the US Food & Drug Administration (FDA). Structures and names can be browsed in a list, searched by name, filtered by structural features, and ranked by similarity to a user-drawn structure. Developed by: Molecular Materials Informatics, Inc. Check iTUNES

e-Drug3D: Currently 1595 molecular structures with a molecular weight < 2000 have been registered. Copyright: Institut de Pharmacologie Moléculaire et Cellulaire

KEGG DRUG: comprehensive drug information resource for approved drugs in Japan, USA, and Europe unified based on the chemical structures and/or the chemical components, and associated with target, metabolizing enzyme, and other molecular interaction network information. Developed by: Kyoto Encyclopedia of Genes and Genomes.

eMolecules: Free chemical structure search engine with millions of public domain structures from vendors worldwide. Copyright: eMolecules, Inc.

Database Managing/Handling:

JChem for Excel : It provides structure handling and visualizing capabilities within a Microsoft Excel® environment. Structures are fully supported within spreadsheets and can be viewed, edited, searched, resized, ordered, managed. Implementation is robust with fast loading/scrolling of many hundreds of structure rows and copy-paste throughout the Microsoft Office® suite. Copyright ChemAxon Ltd.

Bingo : Bingo is a RDBMS data cartridge that provides the industry's next-generation, fast, scalable, and efficient storage and searching solution for chemical information. Copyright **GGA Software Services LLC.**

Protein Data Bank or Database of Ligand and Protein Complexes :

RCSB-PDB: Most prominent database of structures of proteins, nucleic acids, and complex of ligand-protein etc. Copyright RCSB Protein Data Bank

Binding MOAD : a subset of the Protein Data Bank (PDB), containing every high-quality example of ligand-protein binding. Hence, we call it the Mother of All Databases (MOAD). Binding MOAD's goal is to be the largest collection of well resolved protein crystal structures with clearly identified biologically relevant ligands annotated with experimentally determined binding data extracted from literature. Reference: L Hu, ML Benson, RD Smith, MG Lerner, HA Carlson. Binding MOAD (Mother Of All Databases). Proteins 2005, 60, 333-40

PDBbind Database: is designed to provide a collection of experimentally measured binding affinity data (Kd, Ki, and IC50) exclusively for the protein-ligand complexes available in the Protein Data Bank (PDB). References: [1]Wang, R.; Fang, X.; Lu, Y.; Yang, C.-Y.; Wang, S. "The PDBbind Database: Methodologies and updates", J. Med. Chem., 2005; 48(12); 4111-4119.[2] Wang, R.; Fang, X.; Lu, Y.; Wang, S. "The PDBbind Database: Collection of Binding Affinities for Protein-Ligand Complexes with Known Three-Dimensional Structures", J. Med. Chem., 2004; 47(12); 2977-2980.

SCORPIO: It is a FREE online repository of protein-ligand complexes which have been structurally resolved and thermodynamically characterized

BindingDB: It is a public, web-accessible database of measured binding affinities, focusing chiefly on the interactions of protein considered to be drug-targets with small, drug-like molecules. BindingDB contains 910,836 binding data, for 6,263 protein targets and 378,980 small molecules.

AffinDB: Affinity database of Protein-Ligand Complexes. Copyright Peter Block, Christoph Sotriffer, Gerhard Klebe 2002-2013

LigandProtein Database: A collection of ligand-protein complexes, with 3D structures and experimental binding free energies. Created by: Professor Charles L. Brooks, III Department of Chemistry, Biophysics Program 930 N. University Ave; University of Michigan

Drug Target Database:

TTD. (Therapeutic Target Database): A database to provide information about the known and explored therapeutic protein and nucleic acid targets, the targeted disease, pathway information and the corresponding drugs directed at each of these targets. Also included in this database are links to relevant databases containing information about target function, sequence, 3D structure, ligand binding properties, enzyme nomenclature and drug structure, therapeutic class, clinical development status. Created by: **Dr. Chen Yuzong** Deputy Director of Center for Computational Science and Engineering; Professor in Department of Pharmacy National University of Singapore, Singapore

Molecule Pathway Database

SMPDB (The Small Molecule Pathway Database): It is an interactive, visual database containing more than 350 small molecule pathways found in humans. More than 2/3 of these pathways (>280) are not found in any other pathway database. SMPDB is designed specifically to support pathway elucidation and pathway discovery in metabolomics, transcriptomics, proteomics and systems biology. This project is supported by Genome Alberta & Genome Canada,

Chemical Structure Drawing Programmes:

Softwares:

ChemDraw. Chemical structure drawing software developed by CambridgeSoft. Available for Windows and Mac.

MarvinSketch. JAVA based chemical editor for drawing chemical structures, queries and reactions,NMR prediction and much more developed by ChemAxon.Avaiable for WINDOWS and MAC.

ACD/ChemSketch. ChemSketch is a chemical structure drawing program developed by ACD/Labs. Among other features, ChemSketch has the ability to:

Draw and view structures in 2D, or render in 3D to view from any angle

Draw reactions and reaction schemes, and calculate reactant quantities

Generate structures from InChI and SMILES strings

Generate IUPAC systematic names for molecules of up to 50 atoms and 3 ring structures

Predict logP for individual structures

Search for structures in the built-in dictionary of over 165,000 systematic, trivial, and trade names

ChemSketch uses many standard file formats for the import and export of drawings. The full list of available file formats can be found here, under the link "standard file formats." The program allows the user to draw chemical structures including organics, organometallics, polymers, and Markush structures.

Users can download a freeware version of the software on the ACD/Labs website. The full version of the software is also available for purchase.

JChem for Excel. It provides structure handling and visualizing capabilities within a Microsoft Excel® environment. Structures are fully supported within spreadsheets and can be viewed, edited, searched, resized, ordered, managed. Implementation is robust with fast loading/scrolling of many hundreds of structure rows and copy-paste throughout the Microsoft Office® suite. Copyright ChemAxon Ltd.

Accelrys Draw. It enables scientists to draw and edit complex molecules, chemical reactions and biological sequences with ease, facilitating the collaborative searching, viewing, communicating, and archiving of scientific information. Copyright: Accelrys

BKchem: It is a free software chemical drawing program. It was conceived and written by Beda Kosata and is currently maintained by Reinis Danne.

BKChem is written in Python, an interpreted and very nice programming language. This implies some of the program features:platform independence - BKChem should run on any platform that Python does.*performance - as Python is interpreted language you should not expect the performance of a native code compiled application (in present days a very cheap tradeoff for platform independence). However BKChem should be pretty usable on all modern systems.BKChem is developed on GNU/Linux. It was however successfully used under WinXP and MacOS X.

JME Molecular Editor: It is a Java applet which allows to draw / edit molecules and reactions (including generation of substructure queries) and to depict molecules directly within an HTML page. Editor can generate Dayligh SMILES or MDL mol file of created structures. The applet has been developed by Peter Ertl at Comenius University Bratislava and later enhanced at Ciba-Geigy Basel.

Online Programmes:

Marvin molecule editor and viewer. Java based chemical editor uses MARVIN applet. Developed by: ChemAxon.

Molinspiration WebME Molecule Editor. It allows allows creation and editing of molecules in browsers without Java support and without any plugins. The editor is based on a Web2.0 Ajax technology. WebME allows therefore web-based structure input also in institutions where Java applets are not allowed and offers complete platform compatibility. The actual molecule processing in WebME is based on reliable JMEPro editing engine running on a server.

Three Dimensional Viewing (3-D Viewing)

UCSF Chimera : is a highly extensible program for interactive visualization and analysis of molecular structures and related data, including density maps, supramolecular assemblies, sequence alignments, docking results, trajectories, and conformational ensembles. High-quality images and animations can be generated. Chimera is developed by the Resource for Biocomputing, Visualization, and Informatics, funded by the National Institutes of Health (NIGMS P41-GM103311)

PYMOL: is a **user-sponsored** molecular visualization system on an **open-source** foundation. Please support development of this open, effective, and affordable software by purchasing an incentive copy, which is pre-built and comes with maintenance and support. Distributed by DeLano Scientific LLC.

SWISS PDB VIEWER: Swiss-PdbViewer (aka DeepView) is an application that provides a user friendly interface allowing to analyze several proteins at the same time. The proteins can be superimposed in order to deduce structural alignments and compare their active sites or any other relevant parts. Amino acid mutations, H-bonds, angles and distances between atoms are easy to obtain thanks to the intuitive graphic and menu interface. Copyright Swiss Institute of Bioinformatics

Autodock/Vina Plug-In for Pymol: It contains a bunch of new features such as

- Defining binding sites and export to Autodock and VINA input files
- Doing receptor and ligand preparation automatically
- Starting docking runs with Autodock or VINA from within the plugin
- Viewing grid maps generated by autogrid in PyMOL
- Handling multiple ligands and set up virtual screening etc.

Copyright Daniel Seeliger

Computer-Aided Drug-Design Platform using PyMOL: PyMOL plugins for protein preparation (AMBER package and Reduce), molecular mechanics applications (AMBER package), and docking and scoring (AutoDock Vina and SLIDE). Copyright Purdue University

Jmol: an open-source Java viewer for chemical structures in 3D

GLmol: is a 3D molecular viewer based on WebGL and Javascript. You can embed molecular models in Web pages without using Java or plugins. GLmol is open-source software under dual license of LGPL3 or MIT license.Android version ESmol and NDKmol

ICM-Browser: Molecule Visualization, Fully Interactive 3D Slides in PowerPoint and Web, Publication Quality Images, Display Ligand Binding Pocket Surfaces, Hydrogen Bond Display, Measure Distances and Angles and much more. Copyright Molsoft LLC

YASARA: It is a molecular-graphics, -modeling and -simulation program for Windows, Linux and Mac OS X. Copyright Elmar Krieger.

RasMol: It is a program for molecular graphics visualisation originally developed by Roger Sayle

Molegro Molecular Viewer*:* Molegro Molecular Viewer is a free cross-platform application for the visualization of molecules and Molegro Virtual Docker results.Copyright CLC bio. A result of science.

Programmes for File Format Conversion:

Open Babel: is a chemical toolbox designed to speak the many languages of chemical data. It's an open, collaborative project allowing anyone to search, convert, analyze, or store data from molecular modeling, chemistry, solid-state materials, biochemistry, or related areas.Citation **J. Cheminf. 2011, 3:33**

OSRA: Optical Structure Recognition Application: is a utility designed to convert graphical representations of chemical structures, as they appear in journal articles, patent documents, textbooks, trade magazines etc., into SMILES (Simplified Molecular Input Line Entry Specification - see http://en.wikipedia.org/wiki/SMILES) or SD files - a computer recognizable molecular structure format. OSRA can read a document in any of the over 90 graphical formats parseable by ImageMagick - including GIF, JPEG, PNG, TIFF, PDF, PS etc., and generate the SMILES or SDF representation of the molecular structure images encountered within that document.

Mol2Mol: recognizes, reads and writes about 50 different file formats and subformats. It contains a simple graphic display module to inspect the currently loaded molecule. It possesses some chemical intelligence for recognizing detailed atom types, hybridization and chemical environments, which is necessary for converting simpler formats (like X-ray crystallographic files) to more advanced ones, or when hydrogen atoms are automatically to be added to the heavy atoms.

Problematic files can be corrected within Mol2mol or as ASCII files by calling directly your favourite text editor.

Visualisation and Analysis of Protein-Ligand Interaction:

Pose View: automatically generates high-quality 2D structure-diagrams of protein-ligand complexes provided as 3D-input. Such input may come directly from crystal structures or be computed for example by a docking program.

LigPlot: Generates 2-D ligand interaction maps.

Maestro Ligand Interaction: 2-D ligand interaction map using Maestro's free graphical user interface.

DOCKING:

Autodock. It is a suite of automated docking tools. It is designed to predict how small molecules, such as substrates or drug candidates, bind to a receptor of known 3D structure. Maintained by the Molecular Graphics Laboratory, The Scripps Research Institute, la Jolla.

Autodock VINA. It is a new open-source program for drug discovery, molecular docking and virtual screening, offering multi-core capability, high performance and enhanced accuracy and ease of use. It has been designed and implemented by Dr. Oleg Trott in the Molecular Graphics Lab at The Scripps Research Institute.

DOCK. the DOCK algorithm addressed rigid body docking using a geometric matching algorithm to superimpose the ligand onto a negative image of the binding pocket. Important features that improved the algorithm's ability to find the lowest-energy binding mode, including force-field based scoring, on-the-fly optimization, an improved matching algorithm for rigid body docking and an algorithm for flexible ligand docking, have been added over the years. Copyright Soichet group at the UCSF.

GOLD. Genetic Algorithm based docking program. GOLD enables you to make confident binding mode predictions, and achieve high database enrichments. GOLD reliably identifies the correct binding mode for a large range of test set cases, and has been shown to perform favourably against other docking tools in numerous independent studies. Copyright University of Sheffield, GlaxoSmithKline plc and CCDC.

Glide. It offers the full spectrum of speed and accuracy from high-throughput virtual screening of millions of compounds to extremely accurate binding mode predictions, providing consistently high enrichment at every level.. Copyright Schrödinger.

GlamDock. It is based on a Monte-Carlo with minimization (basin hopping) search in a hybrid interaction matching / internal coordinate search space. GlamDock is highly

efficient, taking from 5 seconds (fast virtual screening settings) to ~20 seconds (high quality docking settings) on average on standard 2.8GHz Intel Xeon CPUs.

GEMDOCK. Generic Evolutionary Method for molecular DOCKing GEMDOCK is a program for computing a ligand conformation and orientation relative to the active site of target protein. The tool was developed by Jinn-Moon Yang, a profesor of the Institute of Bioinformatics, National Chiao Tung University

HomDock. is a combination of the ligand based GMA molecular alignment tool and GlamDock.

ICM. ICM-Docking and chemistry module provides access to the chemical information and provides a unique set of tools for accurate individual ligand-protein docking, peptide-protein docking, and protein-protein docking, including interactive graphics tools. With the **ICM-Dock** module, you can do rapid and accurate docking simulations. Copyright MolSoft.

FlexX, Flex-Ensemble (FlexE). Incremental build based docking program. Flexible ligand. Protein flexibility through ensemble of protein structure. Copyright BioSolveIT.

FITTED (Flexibility Induced Through Targeted Evolutionary Description). It aims at improving the accuracy of existing molecule docking software program. It uses a more accurate protein models and is based on a pharmacophore-oriented docking method combined with a genetic algorithm based docking approach. The later takes advantage of more than one structure to dock compounds in virtually flexible proteins.

VLifeDock. Uses three docking approaches e.g. Grid based docking, GA docking and VLife's own GRIP docking program. WINDOWS, LINUX version is available Copyright VLife.

Molegro Virtual Docker. It is an integrated platform for predicting protein - ligand interactions. Molegro Virtual Docker handles all aspects of the docking process from preparation of the molecules to determination of the potential binding sites of the target protein, and prediction of the binding modes of the ligands.

OEDocking OEDocking is a suite of well-validated molecular docking tools and their associated workflows. Each tool is specifically designed to address its own unique application to the docking problem.OEDocking features POSIT for informed pose prediction as well as FRED andHYBRID as complementary tools for virtual screening.

Online Docking Programmes

1-Click Docking: Upload your molecule choose a target from the list and click on DOCK. Copyright **Mcule Inc.**

Swiss Dock: It a web service to predict the molecular interactions that may occur between a target protein and a small molecule.Copyright Swiss Institute of Bioinformatics

ParDOCK: Automated server for rigid docking. Copyright Prof B. Jayaram & Co-workers.

DNA Ligand Docking It is an all-atom energy based Monte Carlo DNA ligand docking, implemented in a fully automated, parallel processing mode which predicts the binding mode of the ligand in the minor groove of DNA. The input is a DNA sequence and drug PDB file. The output will a docked structure alongwith the binding affinity of the docked structures. Copyright Swiss Institute of Bioinformatics

DockingServer It offers a web-based, easy to use interface that handles all aspects of molecular docking from ligand and protein set-up.Copyright Virtua Drug

Rosetta FlexPepDock is a high-resolution peptide docking (refinement) protocol, implemented within the Rosetta framework. The input for this server is a PDB file of a complex between a protein receptor (first chain) and an estimated conformation for a peptide (second chain). FlexPepDock was shown to be able to accurately refine the peptide structure starting from up to 5.5A RMSD of the native conformation, allowing full flexibility to the peptide and side-chain flexibility to the receptor.

PatchDock is an algorithm for molecular docking. The input is two molecules of any type: proteins, DNA, peptides, drugs. The output is a list of potential complexes sorted by shape complementarity criteria.

DOCK Blaster, a public access service for structure-based ligand discovery

INVDOCK, has been developed for computer-automated identification of potential protein and nucleic acid (RNA or DNA) targets of a small molecule (such as a drug, newly designed drug candidate, natural product or other chemical compound).

CLICK HERE for more DOCKING tools

Softwares for Molecular Dynamics

1. **AMBER**: AMBER is a biomolecular package which deals with mainly two things a. a set of force field b. a biomolecular simulation package with MM and QM/MM simulation capability.

2. **Desmond**: Desmond is a high performance molecular dynamics package developed by D.E. Shaw research. Its source code is is available without cost for academic use.

3. **Gromacs**: Free and open source molecular dynamics package capable of to simulate biomolecular systems e.g. protein lipid etc.

4. **COSMOS**: Hybrid QM/MM molecular dynamics, NMR structure calculations etc.

5. **CHARMM**: Widely used force fields and molecular dynamics simulation package developed by Martin Karplus and his group at Harvard.

6. **LAMMPS**: Large-scale Atomic/Molecular Massively Parallel Simulator, a molecular dynamics simulation package which uses MPI for parallel simulation.

7. **Materials Studio**: Commercial package to simulate and modelling materials.

8. **SCIGRESS**: Commercial molecular dynamics package for performing MM, DFT, semiemperical methods, linear scaling SCF, conformational analysis etc.

9. **TeraChem**: High performance ab-initio molecular dynamics and DFT software package with GPU acceleration.

10. **NAMD**: Its parallel efficiency molecular dynamics simulation package often used to simulate large biomolecular systems.

TOOLS FOR TARGET PREDICTION

Related Softwares

MolScore-Antivirals is an expert system, which can detect molecules with antiviral activity. The expert system analyses the probability of a compound to become an antiviral drug and is defined as a value between 0 and 1.Copyright PharmaInformatic Boomgaarden

MolScore-Antibiotics discriminates between antibiotics and non-antibiotics. The MolScore-Antibiotics of a compound measures the probability of having antibiotic activity and is defined as a value between 0 and 1. Copyright PharmaInformatic Boomgaarden

ONLINE TOOLS

TarFisDock : a web server for identifying drug targets with docking approach

ReverseScreen3D is a reverse virtual screening tool that searches against a biologically-relevant and automatically-updated subset of ligands extracted from the RCSB Protein Data Bank in order to identify potential target proteins that are likely to bind a given compound.

BIOLOGICAL ACTIVITY PREDICTION

PASS (Prediction of Activity Spectra for Substances): It estimates the probable biological activity profiles for compounds under study based on their structural formulae presented in MOLfile or SDfile format. General list of predictable biological activities consists of over 4,000 terms including pharmacotherapeutic effects (e.g., antiarrhythmic), biochemical mechanisms (e.g., cyclooxygenase 1 inhibitor), toxicity (e.g., carcinogenic), metabolism (e.g., CYP3A4 inhibition), gene expression

regulation (e.g., VEGF expression inhibition), transporter-related activities (e.g., P-glycoprotein substrate). PASS prediction is based on the knowledge base about structure-activity relationships for more than 260,000 compounds with known biological activities. Average accuracy of prediction estimated in leave-one-out cross-validation procedure for the whole PASS training set is about 95%.

MOST USER FRIENDLY ADME, TOXICITY PREDICTION TOOL

preADMET : Web based application for prediction of different ADME and toxicity parameters. Works with Internet Explorer and Netscape browser. PC version is also available

SYBYL-X : It has the capabilities for small molecule modeling and simulation, macromolecular modeling and simulation, cheminformatics, lead identification, and lead optimization, all wrapped up in an easy to use, cost-effective interface. You can perform 3D-QSAR, Ligand Based Virtual Screening,Cheminformatics,Docking with SYBYL. Copyright Tripos,L.P