Web alert

Computational drug design

Chemistry & Biology August 1999, 6:R227-R228

Most phases of the drug design process have benefited to some extent from new technologies. The early steps of computational design: drug target identification, initial drug molecule identification, and chemical elaboration on the lead compound have been significantly improved by advances in algorithms, large amounts of data and better technology. There are many online resources that relate to drug design techniques; here we focus on a few websites relevant to the 'lead identification' phase of the computational drug design process. Advances in biochemistry, molecular biology and combinatorial chemistry, together with new techniques such as high-throughput screening have had a dramatic impact on the field but they will not be covered in detail here.

Structure-based design: protein structures

Knowledge of a three-dimensional structure for a drug target can greatly facilitate the drug design process. A previous web alert (*Chem. Biol.* **5**, R149) describes a number of websites where protein structures can be obtained, the most important of which is the Protein Data Bank at the Research Collaboratory for Structural Bioinformatics (PDB–RCSB). Frequently, however, there is no known structure for a drug target. In these cases, homology modeling after the identification of a homologous sequence can point to what the structure might look like. A popular method for identifying similar protein sequences is the BLAST program from the National Center for Biotechnology Information (BLAST–NCBI). The alignment can be restricted to sequences found in the PDB (i.e. those with known structures). Generating homology models is an active area of research. Most prediction tools recommend substantial user input and interaction when generating the model. The CMS Molecular Biology Resource is an excellent index that has links to prediction tools and many other sites related to drug design.

Structure-based design: small molecules

Many drug design efforts rely on screening available compounds. One publicly available database is at the National Cancer Institute's Developmental Therapeutics Program (DTP–NCI). Here the three-dimensional structures of approximately 140,000 small molecules can be downloaded (an additional 240,000 are available in SMILES format). It is also possible to order some of the compounds in experimental quantities for testing. One of the largest structure-searchable database of commercially available chemicals is the Available Chemicals Directory from MDL Information Systems (ACD–MDL), which lists compounds commonly available from a number of different chemical companies. Another useful database is the Cambridge Structural Database (CSD), which contains the crystal structures of about 190,000 small molecules.

Small molecule databases

Most groups attempting to screen large databases of compounds for potential drug leads try to filter out molecules with properties known to make poor drug molecules (e.g., those with high ClogP values). Daylight Chemical Information Systems has an online demonstration of products, which are also available for purchase, that can be used to calculate some of these properties. Although filtering databases can greatly reduce their size, databases of 100,000–1,000,000 compounds are not uncommon. It is important in most drug design efforts to search and sort through these large chemical databases, using both traditional search parameters and more complex features such as structural motifs or functional groups. Two popular tools specifically designed for organizing databases of small molecules include ISIS/Base from MDL and Thor/Merlin from Daylight.

Molecular docking

A common tool used to evaluate complementarity between potential lead molecules and their targets is molecular docking. A widely used program is DOCK, available from the Kuntz lab at University of California at San Francisco (UCSF).

Web sites and URLs

PDB–RCSB	http://www.rcsb.org/
BLAST–NCBI	http://www.ncbi.nlm.nih.gov/BLAST/
The CMS Molecular Biology Resource	http://www.unl.edu/stc-95/ResTools/cmshp.html
Developmental Therapeutics Program of the National Cancer Institute	http://dtp.nci.nih.gov/
MDL Information Systems, Inc.	http://www.mdli.com/
Cambridge Crystallographic Data Centre	http://www.ccdc.cam.ac.uk/
Daylight Chemical Information Systems, Inc.	http://www.daylight.com/
DOCK	http://www.cmpharm.ucsf.edu/kuntz/dock.html
FlexX	http://cartan.gmd.de/flex-bin/FlexX
AutoDock	http://www.scripps.edu/pub/olson-web/doc/autodock/index.html
Molecular Simulations, Inc.	http://www.msi.com/download/index.html
UCSF MidasPlus	http://www.cgl.ucsf.edu/Outreach/midasplus/
PubMed	http://www.ncbi.nlm.nih.gov/PubMed/
Institute for Scientific Information	http://www.isinet.com/

This software can be downloaded and is free for academic institutions. The code is distributed with DOCK, allowing for local development — we use a variation of this program in our own research. Another popular molecular docking algorithm is FlexX. This algorithm can be used through a web interface and can be downloaded. Although the web is useful for demonstration purposes, screening compound databases is more efficient if you download and install the software, as most molecular docking calculations are CPU intensive. A molecular mechanics based method, AutoDock, available from the Olson lab at Scripps, is another well-tested algorithm. AutoDock and FlexX are also free for academic institutions.

Visualization tools

Frequently, these complex algorithms for evaluating complementarity are used as a guide to direct the user to better molecules. In our work, we have found that visual inspection of the top few hundred leads proves to be invaluable. There are many programs available for this visual inspection (see *Chem. Biol.* **6**, R25). One of the best visualization tools for the Windows and Mac platforms is WebLab from Molecular Simulations Inc (MSI). For visualization on SGI workstations, we have found Midas from UCSF to be very useful.

There are many online resources that can be used to help identify new drug leads. It is important to remember that traditional methods still contribute to the drug design process. In particular, scientific journals are increasingly filled with references about new techniques for drug design. PubMed, the large, publicly available database of medically related journals is useful but does not index some of the journals containing information about drug design (e.g. *Journal of Computational Chemistry* and *Journal of the American Chemical Society*). The Science Citation Index distributed by the Institute for Scientific Information (ISI) contains a broader list of journals related to drug design. Although this resource is not free, it is invaluable for keeping up with the current methods and technology.

David M. Lorber, Shoichet Lab, Department of Molecular Pharmacology and Biological Chemistry, Northwestern University Medical School, Chicago, IL 60611, USA;
e-mail: dlorber@nwu.edu;
http://pixie.pharm.nwu.edu