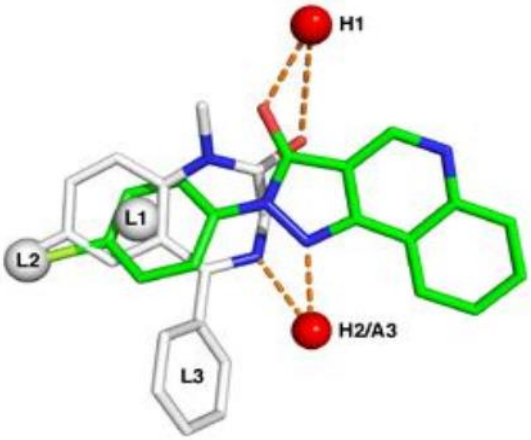
PHARMACOPHORE

A pharmacophore is an abstract description of molecular features which are necessary for molecular recognition of a ligand by a biological macromolecule. The IUPAC defines a pharmacophore to be "an ensemble of steric and electronic features that is necessary to ensure the optimal supramolecular interactions with a specific biological target and to trigger (or block) its biological response".[1] A pharmacophore model explains how structurally diverse ligands can bind to a common receptor site. Furthermore pharmacophore models can be used to identify through denovo design or virtual screening novel ligands that will bind to the same receptor

Features



An example of a pharmacophore model of thebenzodiazepine binding site on the GABAA receptor.[2]White sticks represent the carbon atoms of the benzodiazepine diazepam, while green represents carbon atoms of the nonbenzodiazepine CGS-9896. Red and blue sticks are oxygen and nitrogen atoms that are present in both structures. The red spheres labeled H1 and H2/A3 are, respectively, hydrogen bond donating and accepting sites in the receptor, while L1, L2, and L3 denote lipophilic binding sites.

Typical pharmacophore features include hydrophobic centroids, aromatic rings, hydrogen bond acceptors or donors, cations, and anions. These pharmacophoric points may be located on the ligand itself or may be projected points presumed to be located in the receptor.

The features need to match different chemical groups with similar properties, in order to identify novel ligands. Ligand-receptor interactions are typically “polar positive”, “polar negative” or

"hydrophobic". A well-defined pharmacophore model includes both hydrophobic volumes and hydrogen bond vectors.

Model Development

The process for developing a pharmacophore model generally involves the following steps:

- Select a training set of ligands – Choose a structurally diverse set of molecules that will be used for developing the pharmacophore model. As a pharmacophore model should be able to discriminate between molecules with and without bioactivity, the set of molecules should include both active and inactive compounds.
- Conformational analysis – Generate a set of low energy conformations that is likely to contain the bioactive conformation for each of the selected molecules.
- Molecular superimposition – Superimpose ("fit") all combinations of the low-energy conformations of the molecules. Similar (bioisosteric) functional groups common to all molecules in the set might be fitted (e.g., phenyl rings or carboxylic acid groups). The set of conformations (one conformation from each active molecule) that results in the best fit is presumed to be the active conformation.
- Abstraction – Transform the superimposed molecules into an abstract representation. For example, superimposed phenyl rings might be referred to more conceptually as an 'aromatic ring' pharmacophore element. Likewise, hydroxy groups could be designated as a 'hydrogen-bond donor/acceptor' pharmacophore element.
- Validation – A pharmacophore model is a hypothesis accounting for the observed biological activities of a set of molecules that bind to a common biological target. The model is only valid insofar as it is able to account for differences in biological activity of a range of molecules.

As the biological activities of new molecules become available, the pharmacophore model can be updated to further refine it.

Applications

In modern computational chemistry, pharmacophores are used to define the essential features of one or more molecules with the same biological activity. A database of diverse chemical compounds can then be searched for more molecules which share the same features arranged in the

same relative orientation. Pharmacophores are also used as the starting point for developing 3D-QSAR models.

Lead

Early drug discovery involves several phases from target identification to preclinical development. The identification of small molecule modulators of protein function and the process of transforming these into high-content lead series are key activities in modern drug discovery.[1] The Hit-to-Lead phase is usually the follow-up of high-throughput screening (HTS). It includes the following steps:

Hit confirmation

The Hit confirmation phase will be performed during several weeks as follows:

- Re-testing: compounds that were found active against the selected target are re-tested using the same assay conditions used during the HTS.
- Dose response curve generation: several compound concentrations are tested using the same assay, an IC50 or EC50 value is then generated. Methods are being developed that may allow the reuse of the compound that generated the hit in the initial HTS step. These molecules are removed from beads and transferred to a microarray for quantitative assessment of binding affinities in a "seamless" approach that could allow for the investigation of more hits and larger libraries.
- Orthogonal testing: Confirmed hits are assayed using a different assay which is usually closer to the target physiological condition or using a different technology.
- Secondary screening: Confirmed hits are tested in a functional assay or in a cellular environment. Membrane permeability is usually a critical parameter.
- Chemical amenability: Medicinal chemists will evaluate compounds according to their synthesis feasibility and other parameters such as up-scaling or costs
- Intellectual property evaluation: Hit compound structures are quickly checked in specialized databases to define patentability
- Biophysical testing: Nuclear magnetic resonance (NMR), Isothermal Titration Calorimetry, dynamic light scattering, surface plasmon resonance, dual polarisation interferometry, microscale thermophoresis (MST) are commonly used to assess whether the compound binds effectively to the target, the stoïchiometry of binding, any associated conformational change and to identify promiscuous inhibitors.

- Hit ranking and clustering: Confirmed hit compounds are then ranked according to the various hit confirmation experiments.

Hit expansion

Following hit confirmation, several compound clusters will be chosen according to their characteristics in the previously defined tests. An Ideal compound cluster will:

- have compound members that exhibit a high affinity towards the target (less than 1 μM)
- Moderate molecular weight and lipophilicity (usually measured as cLogP). Affinity, molecular weight and lipophilicity can be linked in single parameter such as ligand efficiency and lipophilic efficiency to assess druglikeness
- show chemical tractability
- be free of Intellectual property
- not interfere with the P450 enzymes nor with the P-glycoproteins
- not bind to human serum albumin
- be soluble in water(above 100 μM)
- be stable
- have a good druglikeness
- exhibit cell membrane permeability
- show significant biological activity in a cellular assay
- not exhibit cytotoxicity
- not be metabolized rapidly
- show selectivity versus other related targets

The project team will usually select between three and six compound series to be further explored. The next step will allow to test analogous compounds to define Quantitative structure-activity relationship (QSAR). Analogs can be quickly selected from an internal library or purchased from commercially available sources. Medicinal chemists will also start synthesizing related compounds using different methods such as combinatorial chemistry, high-throughput chemistry or more classical organic chemistry synthesis.

[edit] Lead optimization phase

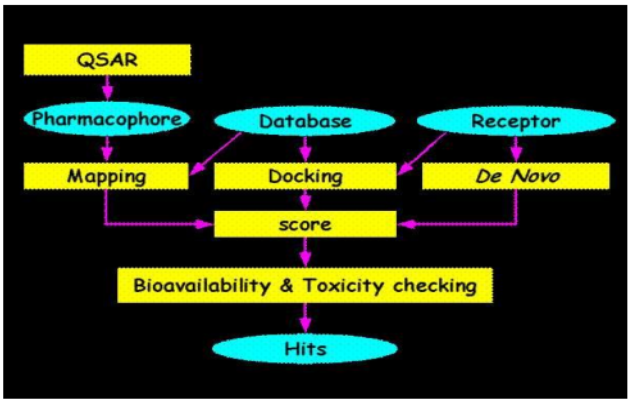
The objective of this drug discovery phase is to synthesize lead compounds, new analogs with improved potency, reduced off-target activities, and physiochemical/metabolic properties

suggestive of reasonable in vivo pharmacokinetics. This optimization is accomplished through chemical modification of the hit structure, with modifications chosen by employing structure-activity analysis (SAR) as well as structure-based design if structural information about the target is available.

DRUG DESIGNING

The shortcoming of traditional drug discovery; as well as the allure of a more deterministic approach to combating disease has led to the concept of "Rational drug design" (Kuntz 1992). Nobody could design a drug before knowing more about the disease or infectious process than past. For "rational" design, the first necessary step is the identification of a molecular target critical to a disease process or an infectious pathogen. Then the important prerequisite of "drug design" is the determination of the molecular structure of target, which makes sense of the word "rational".

In fact, the validity of "rational" or "structure-based"drug discovery rests largely on a high-resolution target structure of sufficient molecule detail to allow selectivity in the screening of compounds. Simple flowchart for drug designing shown in the figure:



Drug designing basically of two types namely ligand based approach or receptor based approach. In both the case the point of centre only differ but requirement of receptor and ligand essential in both the case. By considering these facts, the following steps and online tools shown below for drug designing.

Drug designing steps were usually divided into steps as follows:

I. LIGAND PREPARATION

II. RECEPTOR PREPARATION

III. DOCKING

IV. BINDING AFFINITY STUDIES

1. LIGAND PREPARATION:

Ligand preparation further divided into different heading namely, ligand retrieval or collection, liand conversion and ligand analysis

a. Ligands Collection:

1. DRUGBANK:http://www.drugbank.ca/

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 nearly 4800 drug entries including >1,480 FDA-approved small molecule drugs, 128 FDA-approved biotech (protein/peptide) drugs, 71 nutraceuticals and >3,200 experimental drugs. Additionally, more than 2,500 non-redundant protein (i.e. drug target) sequences are linked to these FDA approved drug entries. Each DrugCard entry contains more than 100 data fields with half of the information being devoted to drug/chemical data and the other half devoted to drug target or protein data.

2. PUBCHEM: http://pubchem.ncbi.nlm.nih.gov/

PubChem provides information on the biological activities of small molecules. It is a component of NIH's Molecular Libraries Roadmap Initiative.

3. CHEMBANK : http://chembank.broad.harvard.edu/welcome.htmIt contains part of an electronic structure collection donated by Tudor Oprea. This set originates from a compilation of ~4.5 million compounds commercially available in August 2002. These were collected from CDs offered by 10 vendors. The structures were processed into a standardized format using OpenEye's FILTER software (http://www.eyesopen.com/products/applications/filter.html). Compliance with Lipinski's Rule-of-5 was enforced (no violations allowed), and several "undesirable" chemical substructures were removed. A low-value for drug-like scores (scores > 0.2) was implemented in order to further remove chemicals that were very different from the then-accepted medicinal chemistry space.

Approximately ~2.5 million compounds passed these filters, and these were subsequently subjected to diversity selection using D-optimal design and a 2D-based descriptor system (mostly topological indices, atom counts, and LogP-type descriptors), in order to realize the final collection of ~800K compound structures.

4. LIGAND EXPO: http://ligand-expo.rcsb.org/ld-search.html

Ligand Expo (formerly Ligand Depot) provides chemical and structural information about small molecules within the structure entries of the Protein Data Bank. Tools are provided to search the PDB dictionary for chemical components, to identify structure entries containing particular small molecules, and to download the 3D structures of the small molecule components in the PDB entry. A sketch tool is also provided for building new chemical definitions from reported PDB chemical components.

5. Small molecules search by descriptors: http://www.scfbio-iitd.res.in/software/nrdbsm/drugsearch.jsp

NRDBSM database is aimed specifically at virtual high throughput screening of small molecules and their further optimization into successful lead-like candidates. The NRDBSM besides facilitating focused searches in larger databases once a hit is identified should also help in finding a small number of hits for further optimization. A Search engine is available for querying NRDBSM based on the properties mentioned.

b. Ligand conversion (format conversion):

1. Smileconvertor: http://cactus.nci.nih.gov/services/translate/

This tool was used to convert smiles into PDB, SDF, Mol formats and both in 2D and 3D formats.

2. 2Dto3D convertor: http://www.molsoft.com/2dto3d.html

ICM 2D to 3D converter convertion functionality allows construction of icm molecular objects from smiles-strings and creates optimized 3D structures using MMFF atom type assignment and force-field optimization. You can use this page to test convert your chemical structure in smiles-format to 3D and view it using our Java applet-based viewer. A simplified 3-D graphical object will be created at the Molsoft server using the ICM software, and your browser won't have to download too much data.

3. CORNIA: http://www.molecular-networks.com/online_demos/index.html

This tool was used to generate 3D structures in SDF format.

4. CONVERT: http://www.molecular-networks.com/online_demos/convert_demo.html

This online tool might be utilized for the generation of 3d structures in different formats like SDF, Mol, smiles etc.,

5. PRODRG: http://davapc1.bioch.dundee.ac.uk/prodrg/index.html

This WWW PRODRG server will convert coordinates for small molecules in PDB format to the following topology formats: GROMOS, GROMACS, WHAT IF, REFMAC5, CNS, O, SHELX, HEX and MOL2. In addition, coordinates for hydrogen atoms are generated. You can now also sketch your small molecule in a simple text editor, and paste this into the window below. You will be returned all of the above topology files + a GROMOS energy

c. Ligand Analysis:

i)Molecular Descriptors:

Molecular descriptors plays crucial role in the drug identification area. So it is essential to know and have the molecular descriptors values regarding ligands. Molecular descriptors predicted through QSAR and QPAR models.

1. **MOLINSPIRATION: http://www.molinspiration.com/cgi-bin/properties**

Draw your molecule and press the [Calculate Properties] or [Predict Bioactivity] button. When the JME input is not working on your computer, try our NEW WebME Ajax editor NEW or paste a raw SMILES here. You may wish to check also our Property Prediction FAQ or more information about calculated properties and drug likeness.

2. **EDRAGAON:** http://www.vcclab.org/lab/edragon/start.html

E-DRAGON is the electronic remote version of the well known software DRAGON, which is an application for the calculation of molecular descriptors developed by the Milano Chemometrics and QSAR Research Group of Prof. R. Todeschini. These descriptors can be used to evaluate molecular

structure-activity or structure-property relationships, as well as for similarity analysis and highthroughput screening of molecule databases. DRAGON provides more than 1,600 molecular descriptors that are divided into 20 logical blocks. The user can calculate not only the simplest atom type, functional group and fragment counts, but also several topological and geometrical descriptors. The first release of DRAGON dates back to 1997. Updates and inclusions of new molecular descriptors are regularly made in order to advance research in QSAR.

3. MOLEDB: http://michem.disat.unimib.it/mole_db/

Molecular descriptors data base was used to get different molecular descriptors for different ligands and drugs which already stored in databases.

4. Model: http://jing.cz3.nus.edu.sg/cgi-bin/model/model.cgi

MODEL - **Molecular Descriptor Lab** for Computing structural and physichemical properties of molecules from their 3D structures.

5. Lipinski's rule prediction: http://www.scfbio-iitd.res.in/utility/LipinskiFilters.jsp

Lipinski rule of 5 helps in distinguishing between drug like and non drug like molecules. It predicts high probability of success or failure due to drug likeness for molecules complying with 2 or more of the following rules:

- Molecular mass less than 500 Dalton
- High lipophilicity (expressed as LogP less than 5)
- Less than 5 hydrogen bond donors
- Less than 10 hydrogen bond acceptors
- Molar refractivity should be between 40-130.

ii) ADME Prediction and Druglikliness prediction:

The success of an drug maily depends upons its ability to enter into the host and not producing any adverse effect on it. These properties were tested by ADME (Absorption, Distribution and Metabolism and Excretion) prediction tools and lead and druglikliness of the chemicals also determined in this phase of designing.

1. **ADMETools:** http://www.simcyp.com/ProductServices/FreeADMETools/

2. **ADME DATABASE:** http://modem.ucsd.edu/adme/databases/databases_extend.htm

3. **ADMETox online Tool:** http://www.pharma-algorithms.com/webboxes/

4**. Drug Likliness: http://www.molsoft.com/mprop/**

5. **Lead finding :** http://leadfinding.com/

II. RECEPTOR PREPARATION:

For drug to act, target is necessary which might be of receptor or enzyme or hormone type. Initially, the receptor should be prepared by structure modeling method or it might be download from the structure databases. After modeling or downloading from corresponding sources, it should be prepared properly for docking which might be achieved by using binding site analysis tools and target determination tools.

i) Bindingsite Analysis:

1. **CASTp:** http://sts.bioengr.uic.edu/castp/index.php

2. **Protein Pocket:** http://sts.bioengr.uic.edu/pni/

3. **Protein Cavities:** http://luna.bioc.columbia.edu/honiglab/mark-us/cgi-bin/submit.pl

4. **MEDOCK:** http://bioinfo.mc.ntu.edu.tw/medock/step1.html

The MEDock (**M**aximum-**E**ntropy based **Dock**ing) web server is aimed at providing an efficient utility for prediction of ligand binding site. A major distinction in the design of MEDock is that its global search mechanism is based on a novel optimization algorithm that exploits the maximum entropy property of the Gaussian distribution.

5**. ODA :** http://www.molsoft.com/oda.cgi

ODA (Optimal Docking Areas) is a new method to predict protein-protein interaction sites on protein surfaces. It identifies optimal surface patches with the lowest docking desolvation energy values as calculated by atomic solvation parameters (ASP) derived from octanol/water transfer experiments and adjusted for protein-protein docking. The predictor has been benchmarked on 66 non-

homologous unbound structures, and the identified interactions points (top 10 ODA hot-spots) are correctly located in 70% of the cases (80% if we disregard NMR structures).

ii) Target Determination:

1. **TarFisDock :** http://www.dddc.ac.cn/tarfisdock/index.php

TarFisDock is a web server for identifying drug targets with docking approach. Given a small molecule which can be drug, drug candidate, natural product, or new synthetic compound, TarFisDock docks it into the protein targets in PDTD (Potential Drug Target Database), and outputs the top 2%, 5% or 10% candidates ranked by the energy score, including their binding conformations and a table of the related target information. The server is freely accessible for anonymous user. And one user's result is protected from being retrieved by another. However users are encouraged to fill in a very simple registration form for better safety and convenience. Now submit your molecular structure(in mol2 format) by clicking

2. **TTD:** http://bidd.nus.edu.sg/group/cjttd/ttd.asp

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/ligands directed at each of these targets. Also included in this database are links to relevant databases that contain information about the function, sequence, 3D structure, ligand binding properties, enzyme nomenclature and related literatures of each target. This database currently contains 1535 targets and 2107 drugs/ligands.

3. **SUPERTARGET SEARCH**: http://insilico.charite.de/supertarget/main.html#Home

4. **Binding Target Determination:** http://www.bindingdb.org/bind/vsOverview.jsp

III. DOCKING:

After collecting, preparing ligands and receptors, they should be assessed for their interaction ability with docking procedure. There were many docking tools are available online. Docking studied under two heads namely, protein-ligand docking and protein-protein docking.

i) Protein-Ligand Docking

1. **Patch dock:** http://bioinfo3d.cs.tau.ac.il/PatchDock/

2. **ParDock:** http://www.scfbio-iitd.res.in/dock/pardock.jsp

ParDOCK is an all-atom energy based Monte Carlo,rigid protein ligand docking, implemented in a fully automated, parallel processing mode which predicts the binding mode of the ligand in receptor target site.

ii) Protein-Protein Docking:

1. **CLUSPRO**: http://nrc.bu.edu/cluster/

2.**FIREDOCK:** http://bioinfo3d.cs.tau.ac.il/FireDock/

3. **CLUSPRO:** http://cluspro.bu.edu/~rb/cluspro/login/main.php

Research in the Structural Bioinformatics lab, headed by Sandor Vajda, focuses on the recognition of proteins and small molecules by protein receptors. Studying protein-protein interactions is crucial for a better understanding of processes such as metabolic control, signal transduction, and gene regulation, whereas the ability to dock small ligands to proteins is the key to rational drug and vaccine design strategies. Both problems become much more difficult if no x-ray structure of the protein is available. Accordingly, our main research areas are (1) the development of efficient protein docking algorithms, (2) docking of small ligands to proteins, primarily for the characterization of binding sites, and (3) homology modeling of proteins.

4. **Vakser Lab:** http://vakser.bioinformatics.ku.edu/resources/gramm/grammx/

This is the Web interface to our current protein docking software made available to the public. This software is different from the original GRAMM, except that both packages use FFT for the global search of the best rigid body conformations.

5. **3DGarden:** http://www.sbg.bio.ic.ac.uk/3dgarden/index.cgi

3DGarden is an integrated software suite for performing protein-protein docking. For any pair of protein structures specified by the user, 3DGarden's primary function is to generate an ensemble of putative complexed structures and rank them. The highest-ranking candidates constitute predictions for the structure of the complex. 3DGarden cannot be used to decide whether or not a particular

pair of proteins interacts. 3DGarden cannot currently be used for docking DNA/RNA structures with proteins or with other DNA/RNA.

IV. BINDING AFFINITY STUDIES:

1. **DRUGSCORE: http://pc1664.pharmazie.uni-marburg.de/drugscore/**

DrugScore^{ONLINE} is a web-based user interface for the knowledge-based scoring functions DrugScore^{CSD} and DrugScore^{PDB}. DrugScore^{ONLINE} enables you to score protein-ligand complexes of your interest and to visualize the per-atom score contributions as illustrated in the figures shown below. Blue spheres denote favorable interactions whereas red spheres stand for disfavorable ones. The sizes of the spheres correlate with the contributing per-atom scores.

2**. BAPPL :** http://www.scfbio-iitd.res.in/software/drugdesign/bappl.jsp

Binding Affinity Prediction of Protein-Ligand (BAPPL) server computes the binding free energy of a non-metallo protein-ligand complex using an all atom energy based empirical scoring function.

3.**AFFINITY DB**: http://www.agklebe.de/affinity

AffinDB is a database of affinity data for structurally resolved protein–ligand complexes from the Protein Data Bank (PDB). It is freely accessible at http://www.agklebe.de/affinity. Affinity data are collected from the scientific literature, both from primary sources describing the original experimental work of affinity determination and from secondary references which report affinity values determined by others. AffinDB currently contains over 730 affinity entries covering more than 450 different protein–ligand complexes. Besides the affinity value, PDB summary information and additional data are provided, including the experimental conditions of the affinity measurement (if available in the corresponding reference); 2D drawing, SMILES code and molecular weight of the ligand; links to other databases, and bibliographic information. AffinDB can be queried by PDB code or by any combination of affinity range, temperature and pH value of the measurement, ligand molecular weight, and publication data (author, journal and year). Search results can be saved as tabular reports in text files. The database is supposed to be a valuable resource for researchers interested in biomolecular recognition and the development of tools for correlating structural data with affinities, as needed, for example, in structure-based drug design.

4. **LIGAND-PROTEIN DB**: http://lpdb.chem.lsa.umich.edu/

In computational structure-based drug design, the scoring functions are the cornerstones to the success of design/discovery. Many approaches have been explored to improve their reliability and accuracy, leading to three families of scoring functions: force-field-based, knowledge-based, and empirical. The last family is the most widely used in association with docking algorithms because of its speed, even though such empirical scoring functions produce far too many false positives to be fully reliable. In this work, we describe a World Wide Web accessible database that gathers the structural information from known complexes of the PDB with experimental binding data. This database, the Ligand-Protein DataBase (LPDB), is designed to allow the selection of complexes based on various properties of receptors and ligands for the design and parametrization of new scoring functions or to assess and improve existing ones. Moreover, for each complex, a continuum of ligand positions ranging from the crystallographic position to points on the surface of the protein receptor allows an assessment of the energetic behavior of particular scoring functions.

5. **PEARLS: http://ang.cz3.nus.edu.sg/cgi-bin/prog/rune.pl**

OTHER SITES

DENOVO DOCKING-GWIDD http://gwidd.bioinformatics.ku.edu/home

GWIDD is a comprehensive resource for genome-wide structural modeling of protein-protein interactions. It contains interaction information for multiple organisms. The structures of the participating proteins are modeled or crystallographic coordinates are retrieved, if available, and docked by GRAMM-X. The resource is not restricted to interactions in the GWIDD database - other sequences or structures may be entered at various stages.

CLINICAL TRIALS: http://www.centerwatch.com/