WORLD JOURNAL OF PHARMACY AND PHARMACEUTICAL SCIENCES

Volume 3, Issue 5, 963-994. Review Article ISSN 2278 – 4357

# AN OVERVIEW OF THE COMPUTER AIDED DRUG DESIGNING

**Shainda Laeeq*[1], Anup K. Sirbaiya[1], Hefazat H. Siddiqui[2], Syed Mehdi Hasan Zaidi[1]**

[1]Department of Pharmaceutical Chemistry, Faculty of Pharmacy, Integral University, Kursi Road, Lucknow-26

.[2]Department of Pharmacology, Dean of Pharmacy Faculty, Integral University, Kursi Road, Lucknow-26.

Article Received on
10 March 2014,

Revised on 05 April 2014,
Accepted on 26 April 2014

***Correspondence for Author**

**Shainda Laeeq**
Department of Pharmaceutical Chemistry, Faculty of Pharmacy, Integral University, Kursi Road, Lucknow-26

## ABSTRACT

This article is an overview on the computer aided drug designing, which can predict experimental results with reasonable accuracy and reduced time, cost and equipment. It identifies the new compound or optimize the lead compound that show significant inhibitory activity against a biological receptor. It provides specicifity of drug on the target whether it is ligand based or structure based screening. Computational molecular modeling methods attempt to predict these interactions and thus the binding affinities and conformation of protein-ligand complexes. **Introduction to computer-aided drug design (CADD):** In recent years, the field of computer-aided drug design (CADD) has grown rapidly, enhancing our understanding of complex biological processes and protein-ligand interactions. CADD can predict experimental results with reasonable accuracy and reduced time, cost and equipment. CADD continuously enhances the progress of drug discovery and refinement of therapeutic agents with many successful examples. Computational drug design has been widely used in the pharmaceutical industry to either identify new compounds or optimise lead compounds that show significant inhibitory activity against a target biological receptor. A small number of examples of these uses are included in Table 1. It is known that chemicals can bind to biological receptors and produce a specific therapeutic response. Drug design is often targeted against receptor molecules which are proteins. The ability of a ligand to bind to a specific protein is related to molecular structure, orientation and conformation. During the binding process, there are enthalpy and entropy changes in the protein-ligand system, associated with alteration of both intra- and inter-molecular structures of protein and ligand.

These conformational changes allow the ligand to bind to the protein active site in a more stable manner. In general, protein-ligand interactions of pharmaceutical interest principally involve non-covalent interactions, including hydrogen bonds, ionic interactions, hydrophobic interactions, $\pi$-$\pi$ interactions and cation-π interactions. Computational molecular modeling methods attempt to predict these interactions and thus the binding affinities and conformation of protein-ligand complexes.

**Table 1.Examples of Marketed Drugs Involving the Use of Structure-based Drug Design**

| Year | Generic Name | Brand Name | Manufacturer | Against / Inhibits |
|---|---|---|---|---|
| 1989 | Zanamivir (vonltzstein et al., 1996) | Relenza | GlaxoSmithKline | Neuraminidase |
| 1997 | Nelfinavir (Kaldoretal., 1997) | Viracept | Hoffman-La Roche | HIV protease |
| 1998 | Raltitrexed (Blackledge, 1998) | Tomudex | AstraZeneca | Thymidylate |
| 1999 | Amprenavir (Adkins &Faulds, 1998) | Agenerase | GlaxoSmithKline | HIV protease |
| 2007 | Raltegravir (Schames et al., 2004) | Isentress | Merck | HIV integrase |

**Overview**

Traditional drug discovery generally requires innovation of lead compounds by medicinal chemists. Lead compounds will then be synthesized and experimentally tested until a compound with the desired pharmacological properties has been developed. This trial-and-error process can be expensive and time consuming, the success rate depending primarily upon the knowledge and experience of the medicinal chemist. In modern drug discovery, computational methods are generally involved in identifying and modifying lead compounds. For lead discovery and lead optimization, 3D structural information on the ligand, the protein receptor, or both, is highly desirable. A commonly-used method in 3D computer-aided drug design is molecular docking.

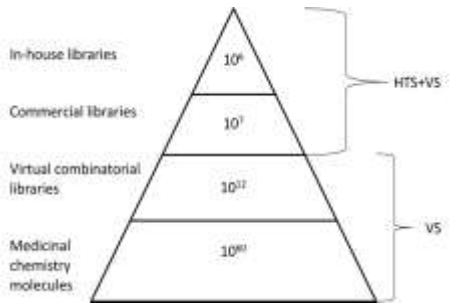
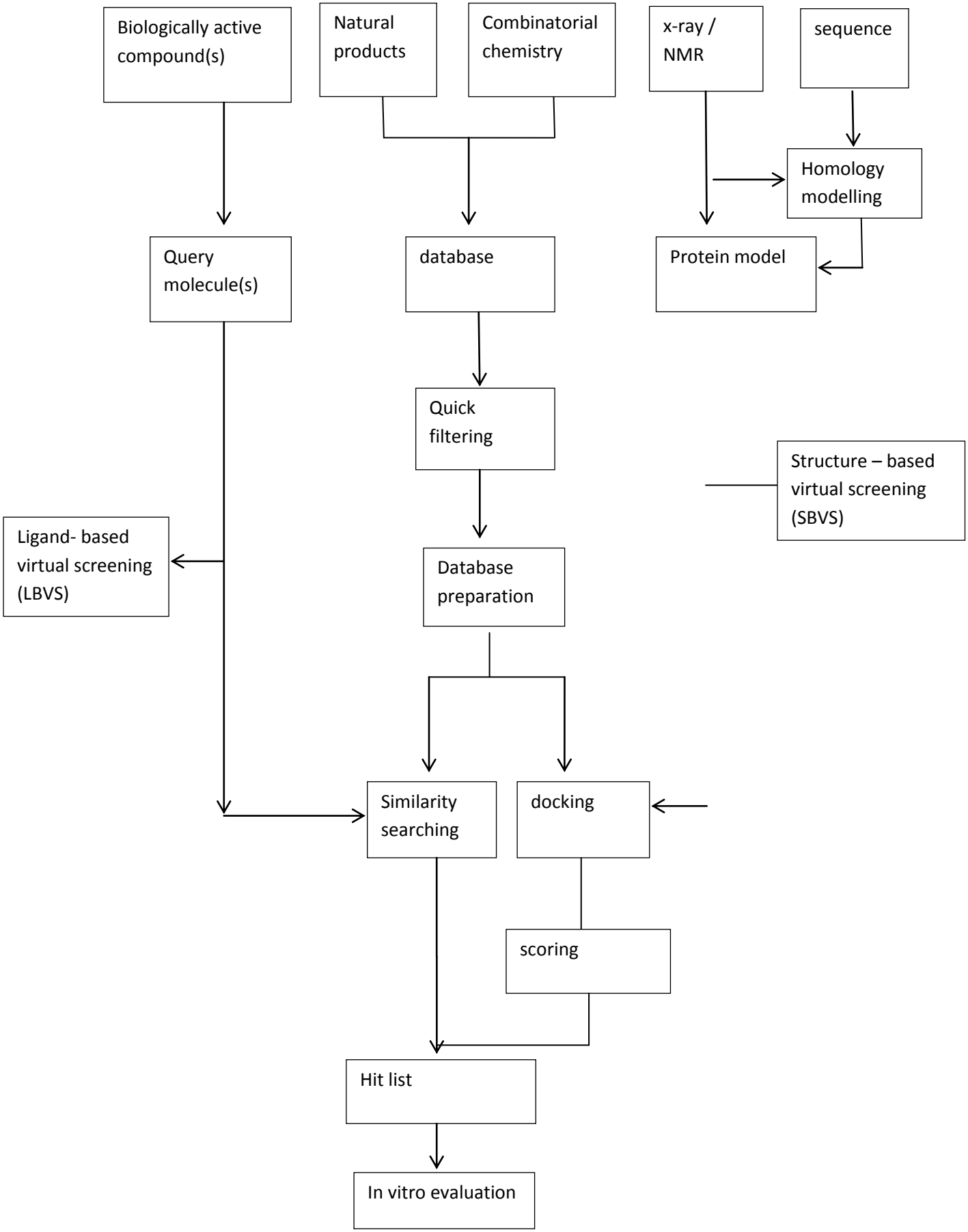



***Fig.1:The numbers of molecules available from different sources.***


Biologically active compound(s)
Natural products
Combinatorial chemistry
x-ray / NMR
sequence
Homology modelling
Query molecule(s)
database
Protein model
Quick filtering
Structure – based virtual screening (SBVS)
Ligand- based virtual screening (LBVS)
Database preparation
Similarity searching
docking
scoring
Hit list
In vitro evaluation

**Virtual screening**

Virtual screening can be divided into two major strategies:

a. ligand-based virtual screening (LBVS) and

b. structure based or target-based virtual screening (SBVS)

Both approaches can be applied simultaneously provided that enough information is available. As with any modeling procedure, experimental data is required before predictions can be made. In LBVS, the information about other similarly bioactive compounds ("keys") is used, whereas in SBVS 3Dmodels of the target proteins ("locks") are utilized. The 3Dmodels of target proteins are either derived from X-ray crystallography and Nuclear Magnetic Resonance (NMR) experiments or homology modeling, where the existing experimental data is used to build comparative models of proteins from their amino acid sequence. The chemical libraries that are screened are usually created using combinatorial chemistry techniques or they are built from natural products, such as chemicals extracted from plants. The result of a virtual screen is a hit list that is a prioritized list of compounds suitable for biological testing (*in vitro* evaluation). It is hoped that the top of the hit list contains more bioactive compounds than could be obtained from a random selection.The large number of compounds to be screened means that virtual screening methods need to be fast in order to be truly useful for drug development. As the price of high-performance computing has plummeted due to advances in both hardware and software, virtual screening costs only a small fraction of HTS. One can also predict bioactivity for molecules that can be readily made, but do not yet exist (virtual libraries). This strategy is often applied in the lead optimization phase.

**HIGH-PERFORMANCE COMPUTING IN VIRTUAL SCREENING**

Since a large amount of data is processed in virtual screening, High-Performance Computing (HPC) is required for most real life applications. HPC is based on massively parallel computing using supercomputers and computer clusters. Most algorithms used in virtual screening are trivial to parallelize by splitting the data into smaller pieces. In the past, HPC required specialized and expensive hardware. Due to the availability of cheap multi-core processors and free operating systems like Linux, this is no longer the case.

**a)LIGAND-BASED VIRTUAL SCREENING (LBVS)**

Ligand-based virtual screening is based on "the similarity principle" that states that

similar molecules tend to have similar biological properties. Molecular similarity is a subjective concept like beauty and molecules can be “similar” in many different ways.

The aim of LBVS is usually scaffold hopping. LBVS methods can be also helpful in drug repurposing, where new targets and diseases are sought for existing drug molecules. *Scaffold hopping can be defined as the identification of isofunctional molecular structures with significantly different molecular backbones.* Although "scaffold hopping" is the most commonly used term, "leapfrogging", "scaffold searching” and "lead hopping" have also been used to describe this strategy[2].

Since peptides make very poor drug molecules for various reasons (e.g. flexibility,proteolytic stability), it is desirable to replace the peptidic scaffold of a bioactive molecule.Several successful cases have been published where peptides have been substituted by other structures.

Poor absorption, distribution, metabolism, excretion and toxicity (ADMET)properties may also be the reasons for scaffold hopping. If a lipophilic scaffold can be changed to a more polar one, this will increase the solubility of the compound, which is of tena major problem in contemporary drug discovery programs[2]. Scaffold hopping has also been used for intellectual property issues. When a “breakthrough-drug” is introduced onto the market by a pharmaceutical company, its competitors try to develop molecules with similar biological but a dissimilar chemical structure (“me-too” drugs).

Scaffold hopping is an ill-defined term and highly subjective concept. There arevarious definitions for a scaffold. The most commonly used scaffold concept is based on thework of Bemis and Murcko, where they analyzed the properties of known drugs using the Comprehensive Medicinal Chemistry (CMC) database[2]. These scaffolds are sometimes referred to as “Murcko’s scaffolds” or “molecular frameworks”[2]. The classification is based on a hierarchical description of molecules, illustrated in Fig. 2 and 3.

A molecule consists of a scaffold that has side chains, whereas a scaffold consists of a ring system and linkers. Murcko’s scaffolds have the obvious pitfall that onlycyclic scaffolds that were included in the CMC datasets can be detected. Recently, Lipkusand co-workers analyzed the scaffolds found in the CAS registry, they found out that half ofthe 24 million

organic compounds in CAS could be described by only 143 scaffolds. Other general classifications are the *maximum common substructures*, *maximum rigid fragments* and *RECAP fragments*. The problem of scaffold definition has not yet been satisfactorily solved[2].

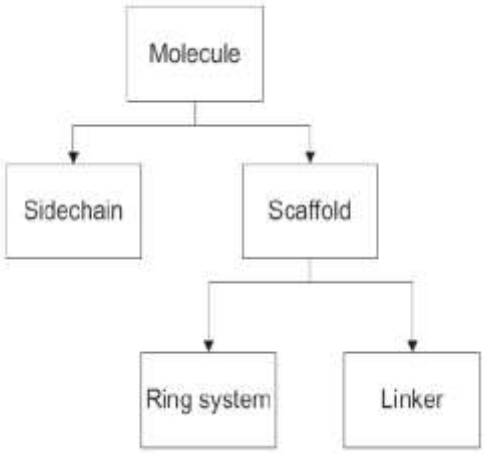



**Fig.2: Hierarchial description of molecules(adapted and modified from Bemis and Murcko 1996)**

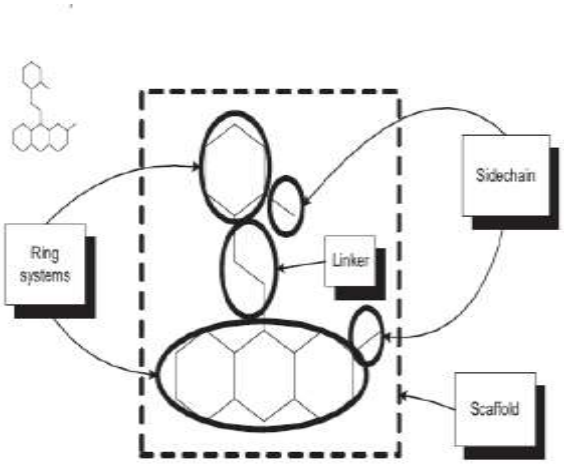



**Fig. 3: Detecting scaffolds using concepts of ring system, linkers and side chains (adapted and modified from Bemis and Murcko 1996)**

One example scaffold hopping is shown in Fig. 4, where there are the two similarly bioactive compounds that have completely different scaffolds. Hypothesis for their similar activity is based on matching three-dimensional shape of the molecules.

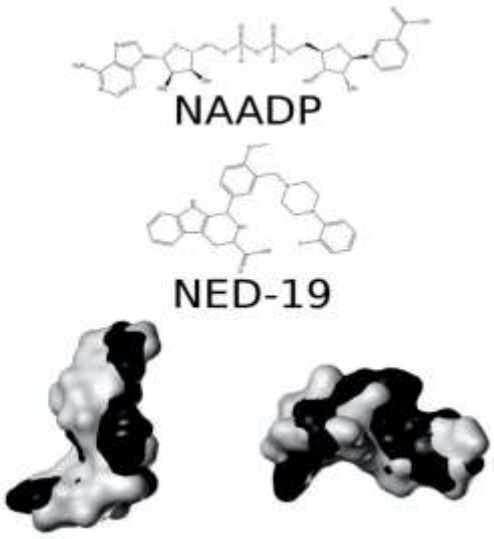



**Fig 4: Example of scaffold hopping. NAADP and NED-19 have similar bioactivity even though their scaffolds are completely different. Both molecules are similar in their 3D surfacs (black and white shapes) (Connolly 1983). Analysis is based on the findings of Naylor and co-workers (Naylor et al. 2009).**

There are many similarity methods which have been developed for LBVS. Some of the commonly used approaches are presented in Table 2.

**Table 2: Some of the commonly used LBVS approaches**

| METHOD | EXAMPLE(S) | APPROACH |
|---|---|---|
| 0D/1D | Atom counts | Generated from molecular graph |
| 2D fingerprints | MACCS | Quantitative comparison of bit strings |
| 3D descriptors | UNITY3D, NPR, USR, ESshape3D, GRIND | Generated using intramolecular distances |
| Pharmacophores | Catalyst | Common features ofactive molecules are detected |
| 3D similarity based on pair-wise alignment | ROCS/EON, BRUTUS, ShaEP,FlexS | Comparison of superimpose molecules |

**0D-2D descriptors**

The simplest ways of describing molecules are the one- and two dimensional descriptors like the number of carbon atoms ormolecular indexes based on graph theory.These kinds of descriptors are easy to calculate with modelling tools like MOE (Chemical Computing Group). Despite their simplicity, they have been shown to be surprisingly effective in virtual screening. The commonly used two-dimensional fingerprints are binary strings that encode the presence or absence of sub-structural fragments. A set of chemical features is defined and

then a bit is set to either zero (0) or one (1), depending on whether the substructure exists in the molecule or not. A fingerprint is a long bit string, which can also be expressed as an integer. An example of a two-dimensional fingerprint is shownin Figure 5, which illustrates the MACCS-fingerprint for citalopram.

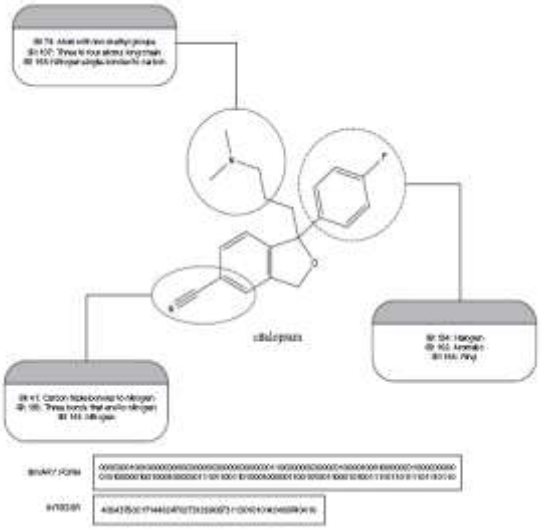


**Fig 5: Example of 2D fingerprint: MACCS structural keys for citalopram. Forclarity, only some of the defined bits are shown. Fingerprint generated with OpenBabel2.2.3 (Guha et al. 2006).**

Usually a single fingerprint is compared with a database inorder to retrieve similar compounds. However, it is also possibleto form fusion fingerprints based on multiple fingerprints from several query molecules[2].

Even though 2D fingerprints have proved to be useful tools in drug discovery projects, they suffer from several drawback[2], For example, a single atom change in a ring structure maychange the fingerprint from being nearly similar to almost completely different. Moreover, asis shown in Figure 5, two compounds that have very different topologies can non etheless adopt a similar orientation and thus could have similar biological effects.

**3D descriptors**

3D fingerprints (also known as pharmacophore keys) encode 3D relationships in a molecule as a bit string[2]. An example of such an algorithm is the UNITY 3D fingerprints. The basic idea is presented in Figure 8, where there are two different conformations of disulfiram. The combinations of features are enumerated with the distances between them. In a 3D finger

print, each bit encodes a distance between specific groups. For example, bit 0 could be "donor-donor with distance 2-2.5" and bit 1 "donor-donor with distance 2.5-3" etc. The number of features used in combinations varies from two up to nine[2]. However, the size of a fingerprint increases rapidly with the number of features used.

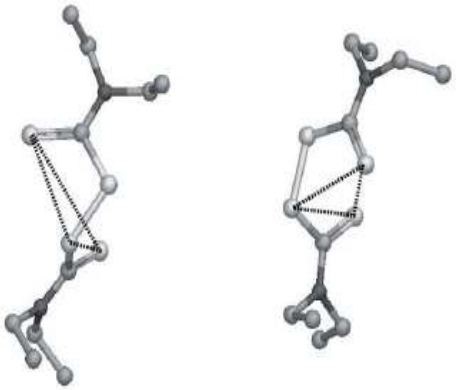


**Fig6: Two conformations of disulfiram. The three-point pharmacophoric feature isdifferent in the two conformations. Conformations generated with OPLS_2005 force fieldimplemented in MacroModel (Schrodinger Inc). Image created with Maestro (SchrodingerInc).**

The basic problem with 3D fingerprints (and with other 3D methods as well) is of course conformational sampling, since the number of possible conformations *N* increases very rapidly with the number of rotating bonds *n* (so called combinatorial explosion):

$$N = \left(\frac{360}{m}\right)^n$$

where*m* is the size of the rotational angle in degrees. For example, a molecule with six rotatable bonds has 2985984 possible conformations with a rotational angle increment of 30 degrees. It is therefore not possible to use all possible conformations in similarity calculations for most molecules.

**Shape-based descriptors**

encode the shape of the molecule into numbers. The shape complementariness of the ligand to the active site is a prerequisite for the drug action, so several approaches for describing this important feature have been developed[2]. If compared to the 2D fingerprints, which describe molecules as sets of atoms, the shape-based descriptors consider molecules as volumes and surfaces[2]. However, shape based methodology has obvious serious flaws. For example,

completely different molecules like methane and fullerene would be classified as similar because they adopt a similar sphere-like conformation. The shape-based descriptors that are more relevant to the virtual screening are ES Shape 3D (implemented in MOE by Chemical ComputingGroup) and Ultrafast Shape Recognition (USR).

Since both steric and electrostatic properties are important in protein-ligand complementary, the accuracy of shape-based descriptors for virtual screening is limited. examples of a descriptor that encodes both shape and electrostatic properties are Grid- Independent descriptors (GRIND). The descriptors are derived from a collection of GRID molecular interaction fields computed using different chemical probes.

These fields are then discretized by finding "the hot spots" of interactions. The relative position of "hot spots" is then encoded into descriptors called correlograms. Principal component analysis of the correlograms is then used for the similarity calculations. The algorithm for the calculation of GRIND descriptors has evolved over the years. The most recent version of the method is implemented in Pentacle (Molecular Discovery Ltd).

**Pharmacophores**

The term pharmacophore was introduced by Paul Ehrlich in 1909.The modern iupac definition dates from 1998: *"A pharmacophore is the ensemble of steric and electronic features that is necessary to ensure the optimal supramolecular interactions with a specific biological target structure and to trigger (or to block) its biological response."*A pharmacophore is an abstract concept that describes the interaction capability of either one or a group of compounds toward a drug target instead of a real molecule or real association of functional groups.

The main advantage of pharmacophore methods is that it is possible to find very diverse compounds. The early pharmacophores were constructed manually in the 1940's with the knowledge of the bond lengths and the van derwaals radii of atoms (Figure 7). Such simple constraints could be used as a crude filtering criterion for large set of compounds to weed out clearly unsuitable molecules.

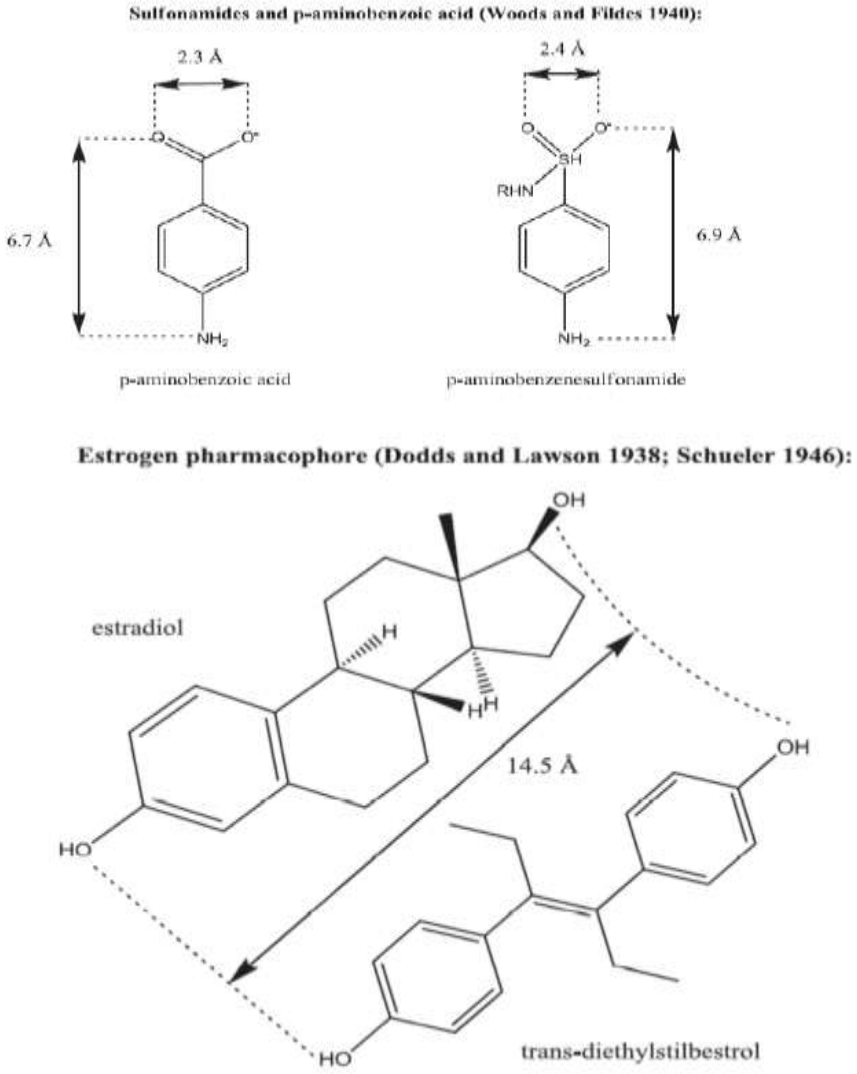



**Fig 7: Two early pharmacophores with example molecules (adapted from Wermuth 2006)**

The work flow for general pharmacophore modeling is presented in Fig 8. Several compounds that have similar biological activities are needed to form a hypothesis. Some methods also allow incorporation of activity data. An important assumption is that all compounds in the pharmacophore have a similar binding mode and thus they can be superimposed. After compounds are superimposed, common features of the molecules can be detected.

A pharmacophore can almost always be generated, but it must be validated by using an external data set before use. After a reasonable pharmacophore is formed, the virtual screening step itself is fast.

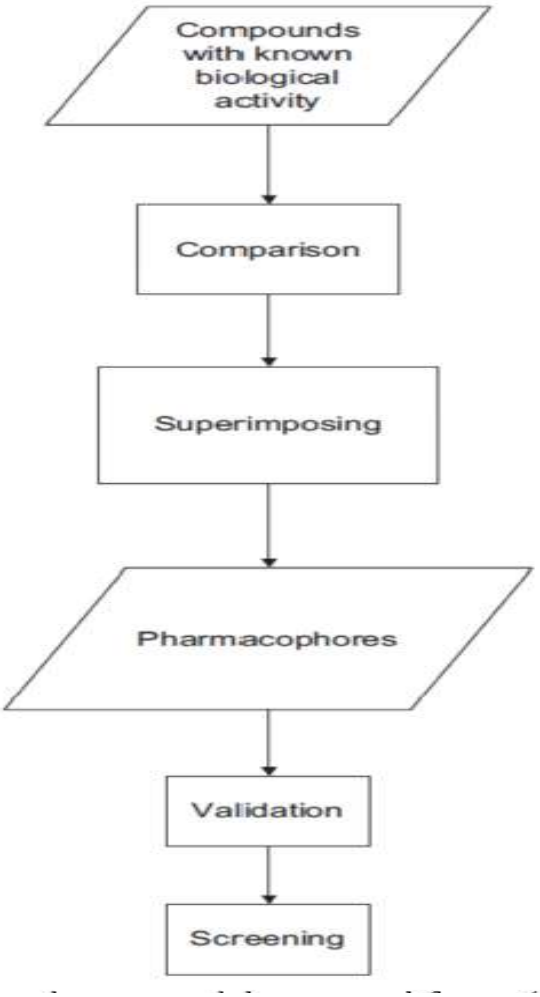



**Fig 8: General pharmacophoremodeling workflow (adapted and modified from Poptodorov et al. 2006)**

To some extent, pharmacophores have been neglected and the development of new methods has been extremely slow[2]. This might be due to the strong emphasis on SBVS in recent years. Since SBVS methods have not been as successful as was originally anticipated, there has been increasing interest in using the pharmacophore approach[2].

Geometry- and feature-based pharmacophore methods usually consider compounds as sets of connected features like hydrophobic and H-bond acceptors/donors. These features are important for selective binding of drug molecules as they describe hydrogen bonding, electrostatics and hydrophobic interactions[2.]

The most widely used geometry and feature-based pharmacophore elucidation method is Catalyst from Accelrys, which is currently a part of the Discovery Studio package.Catalyst is an integrated set of algorithms for conformation generation (ConFirm), molecular

superimposition (HipHop), pharmacophore generation (HypoGen) and database searching (Info).

HipHop and HypoGen provide two approaches for automatic pharmacophore generation. HipHop identifies pharmacophores by aligning the chemical features of active molecules[2]. Each conformation of each molecule is used as a reference for alignment and every configuration is scored. HypoGen is designed to correlate structure and activity for automatic pharmacophore generation[2], working in three steps. In the first step, common features are detected between the two most active compounds. In the second step, those features that are common between active and inactive compounds are removed from the pharmacophore. The last step is an optimization phase where simulated annealing is used to improve the predictive power of the pharmacophore. An exclusion volume can be added to Hypo Genpharmacophores to filter out too large molecules from the search[2].

3D-Quantitative Structure Activity Relationships (QSAR) methods can be considered as field-based automatic pharmacophore generation methods[2]. The most frequently used 3DQSARmethod is Comparative Molecular Field Analysis (CoMFA) devised by Cramer andco-workers. Other widely used 3D-QSAR methods are CoMSIA and GRID/GOLPE .An outline of the method is presented in Figure 10.

The molecular field is presented as alattice. Compounds are superimposed and their activity values, steric and electrostatic potentials are recorded in the QSAR table. From this table, an equation is derived with Partial Least Squares (PLS) data analysis method[2].

This equation can then be used in the prediction of activity for compounds outside the model. Although the basic idea is ratherstraightforward, the correct use of the method is difficult, as the results are critically dependent on conformation and superimposition of the compounds. Furthermore, the chemical parameters used to generate fields and the statistical evaluation methods have alarge influence on the models.

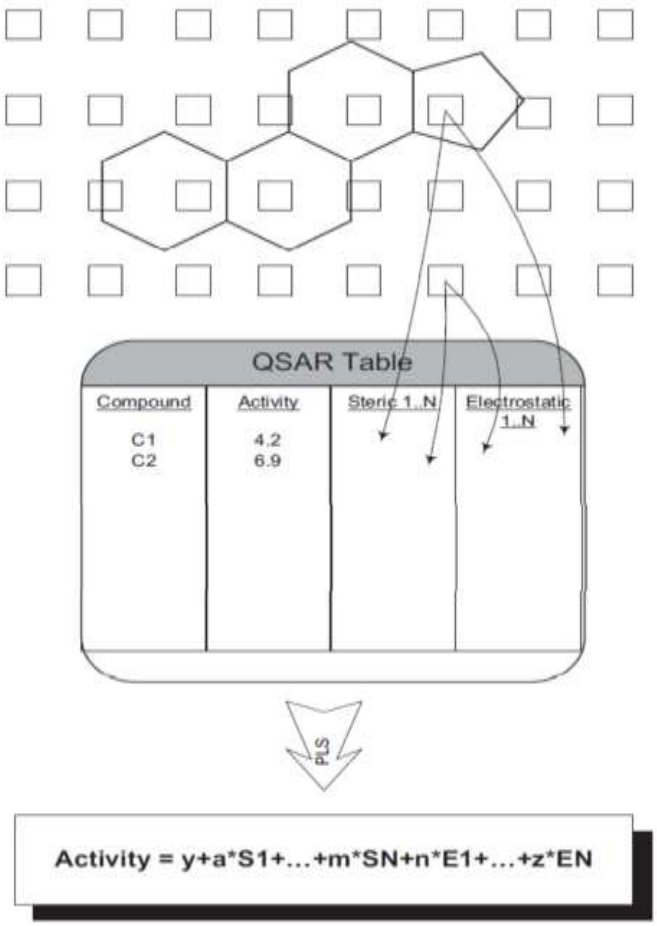



**Fig 9: Comparative Molecular Field Analysis (CoMFA) (adapted from Cramer et al. 1988).**

Even though there are hundreds of CoMFA studies published (PubMed lists over 900 citations with keyword "CoMFA"), most of these studies are mostly retrospective analyses and have very little predictive value that could be used in prospective virtual screening of new biologically active molecules. Also, the superimposing step is a major limitation forvirtual screening applications, as the compounds to be screened need to have a common scaffold to permit automatic alignment. It could be therefore concluded that CoMFA is more a tool for lead optimization rather than a virtual screening method for large databases. Thereis also Topomer-CoMFA available, which is easier to use than the traditional CoMFA[1].

In pharmacophore methods, a set of compounds is compared in order to find common features, which are then matched to a set of compounds in a database. One can also try to match the whole query molecule to database molecules by aligning them in a pair-wise manner. It is easier to find a reasonable alignment for a pair of molecules than for diverse set of molecules.The problem of molecular alignment is a complex issue due to the

degrees of freedom involved. From a practical point of view, there are several high-throughput molecular alignment methods publicly available (Table 2).

**Table 2: High-throughput small molecule alignment-based similarity methods suitable forvirtual screening. License abbreviations: O=Open Source, F=Free, FA=Free for Academicuse and C=Commercial**.

| Program | Reference | Lic. | Website |
|---|---|---|---|
| ROCS | Grant et al. 1995 | FA | www.eyesopen.com |
| EON | Nicholis et al. 2004 | FA | www.eyesopen.com |
| PAPER | Haque and Pande 2009 | O | Simtk.org/home/paper |
| BRUTUS | Ronkko et al. 2006 | C | www.visipoint.fi |
| ShaEP | Vainio et al.2009 | F | User.abo.fi/mivainio/shaep |
| FlexS | Lemmen et al.1998 | C | www.biosolveit.de |

The most widely used molecular alignment method for virtual screening is Rapid Overlay of Chemical Structures (ROCS) from OpenEye Scientific Software[2]. In this method, molecules are superimposed with a smooth Gaussian function representing the molecular volume. ROC optimizes this function by rigidly translating and rotating the molecule with respect to the query molecule. The similarity *S* between two molecules A and B is calculated from the volumes of the molecules (ShapeTanimoto score):

$$S = \frac{O_{AB}}{O_{AA} + O_{BB} - O_{AB}}$$

where*OAA* is the volume of molecule A, *OBB* is the volume of molecule B and *OAB* is the overlapping volume between these molecules.

## STRUCTURE-BASED VIRTUAL SCREENING (SBVS)[2]

### Overview

Structure-Based Virtual Screening (SBVS) is usually based on molecular docking. In molecular docking, a small molecule is fitted into the protein model's active site. The aim of docking is to predict the structure of the complex [P+L] = [PL] under equilibrium conditions in water and to estimate the Gibbs energy of binding $\Delta G$. $\Delta G$ can be described by the equation $\Delta G=\Delta H-T\Delta$. Enthalphic factors ($\Delta H$) include steric and electrostatic complementary, hydrogen-bonding, protein strain and also ligand strain, if the ligand is flexible. Desolvation, rotational and translational entropy are important factors in entropy ($\Delta S$)

There are two major components in a docking program: a search algorithm that produces relevant binding modes (poses), and a scoring function, which should be able to predict the affinity of the docked compound to the protein i.e. estimate $\Delta G$. The searching problem has been basically solved, but the scoring problem persists. Due to the number of atoms involved in the protein-ligand interaction, the problem is extremely complex. A typical approximation in order to speed up the calculations is to use a rigid protein and torsionally flexible ligand instead of a fully flexible protein and ligand. Even with these simplifications, molecular docking is still a time consuming process compared to the ligand-based virtual screening methods. There are over 60 docking programs and more than 30 scoring functions described in the literature. However, only a fraction of the proposed methods are readily available for virtual screening studies. Most of the docking software is commercial, so licensing might represent a rate-limiting step in a virtual screening study even though supercomputing capability is available. Commonly used docking methods include AutoDock, DOCK, LigandFit, FlexX, FRED, GLIDE and GOLD.

**List of docking softwares**

- AutoDock
- DOCK
- LigandFit
- FRED
- GLIDE
- GOLD

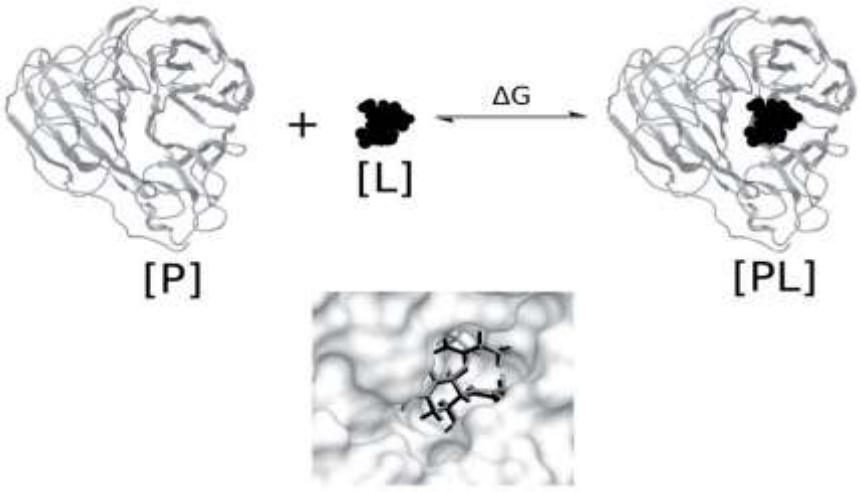



**Fig 10: The concept of molecular docking. Ligand (L) is docked to the protein (P) to form a protein-ligand complex (PL). PDB-Complex 2HU4 (Russell et al. 2006) formed by oseltamivir and N1 neuraminidase is closely reproduced by a docking program. The**

**docking programs best scored solution is shown in black and that experimentally observed in gray. Images created with GLIDE and Maestro (Schrodinger Inc).**

### Docking

Docking procedures aim to identify correct poses of ligands in the binding pocket of a protein and to predict the affinity between the ligand and the protein. In other words, docking describes a process by which two molecules fit together in three-dimensional space. One main motivation in drug discovery is the identification of innovative small molecular scaffolds exhibiting high binding affinity and selectivity for the target together with a reasonable ADME (absorption, distribution, metabolism, excretion) profile, lead and/or drug likeness. Such chemical entities are likely to be able to enter higher phases in the further drug development process. Molecular docking, compared to the fast and successful method of three-dimensional pharmacophore modeling [1,2] is a rather complex and computer-intensive approach to find new compounds by virtual screening.

## BASIC REQUIREMENTS FOR MOLECULAR DOCKING

The setup for a ligand docking approach requires the following components: A target protein structure with or without a bound ligand, the molecules of interest or a database containing existing or virtual compounds for the docking process, and a computational framework that allows the implementation of the desired docking and scoring procedures. The three dimensional structure of the protein-ligand complex has to be detailed at atomic resolution. In many cases only the unbound (ligand-free, apo) form of the protein is determined, without the bioactive conformation of the ligand. Most docking algorithms assume the protein to be rigid, according to the high computational cost that the demand of flexibility implicates. The ligand is mostly regarded as flexible. Beside the conformational degrees of freedom the binding pose in the protein's binding pocket must be taken into consideration.

### Active Site Prediction

Proteins often exhibit an active site: a particular area of the molecule where ligands (or some usual ligand) normally bind. This is sometimes a crevice or other prominent geometric feature, but may also be an area with smoother shape but favourable electrostatic forces. Identifying the active site (also known as the binding site) of a protein is likely to help in the assessment of whether and how a ligand might be docked. For those proteins whose native and complexed structures are documented, it is sometimes possible to determine the active site quite simply; using knowledge derived from known structures, predictive methods may

be developed. Active site proposals can be generated as a side-effect of docking algorithms by assessing a scoring function with a probe ligand.

**Docking protocols** can be described as a combination of two components; a search strategy and a scoring function. The search algorithm should generate an optimum number of configurations that include the experimentally determined binding mode.

**Searching procedures applied for docking**

a. Molecular Dynamics(MD).
b. Random/Monte Carlo search(MC).
c. Genetic Algorithms(GA).
d. Lamarckian Genetic Algorithm (LGA).
e. Tabu searches.
f. Incremental/Fragment Based.

**Scoring functions procedures applied for docking**

a. Force fields
b. Empirical scoring
c. Geometrical scoring
d. Knowledge based parameters.
e. Consensus scoring

**Types of docking**

The accurate prediction of the binding modes between the ligand and protein, (the docking problem) is of fundamental importance in modern structure-based drug design. The task of molecular docking can be divided according to the molecules being involved:

1. Protein-Ligand docking
2. Protein-Protein docking

Specific docking algorithms are usually designed to deal with one of these problems but not with both (different contact area, flexibility, level of representation). Assuming the receptor structure is available, a primary challenge in lead discovery and optimization is to predict both ligand orientation and binding affinity. The major techniques currently available are: molecular dynamics, Monte Carlo methods, genetic algorithms, fragment-based methods, point complementarity methods, distance geometry methods, tabu searches and systematic searches.

**Protein-protein docking methods**

The approaches to protein-protein docking have a lot in common with small molecule docking. The methods are still based on the combination of search algorithm coupled to scoring function. The scoring functions are essentially the same (since we are still dealing with atomic interactions), however the major problem is that the conformational space we now need to search is massive.

**Local and global docking**

The general problem includes a search for the location of the binding site and a search to figure out the exact orientation of the ligand in the binding site. A program that does both makes a **Global docking**

**Rigid vs. flexible docking**

Docking procedures that perform rigid body search are termed **rigid docking.** Docking procedures that consider possible conformational changes are termed **flexible docking**.

**Bound and unbound docking**

In **bound docking** the goal is to reproduce a known complex where the starting coordinates of the individual molecules are taken from the crystal of the complex. In the **unbound docking**, which is a significantly more difficult problem, the starting coordinates are taken from the unbound molecules.

**Clustering and results of docking**

No docking algorithm can produce a single, trustworthy structure for the bound complex, but instead they produce an ensemble of predictions. Each predicted structure has an associated energy (or enthalpy as the case may be) as well as the relative population. By clustering our data based on some "distance" criteria, we can gain some sense of similarity between predictions. The distance can be any of a number of similarity measures, but for 3D structures, RMSD is the standard choice.

**Search algorithms**

**Overview**

Docking protocols can be described as a combination of two components; a search strategy and a scoring function. The search algorithm should generate an optimum number of configurations that include the experimentally determined binding mode. A rigorous search

algorithm would exhaustively check all possible binding modes between the ligand and receptor. All six degrees of translational and rotational freedom of the ligand would be explored along with the internal conformational degrees of freedom of both the ligand and protein. However, this is impractical due to the size of the search space. The practical application of such an extensive search involves the sampling of many high energy unfavorable states which can restrict the success of an optimization algorithm.

A common approach in modeling molecular flexibility is to consider only the conformational space of the ligand, assuming a rigid receptor throughout the docking protocol. However, the searching algorithm is only half the docking problem; the other factor to be incorporated into a docking protocol is the scoring function.

**Searching algorithms**

Different approaches for the docking pose generation have been applied. The methods can be roughly divided into three main types: rigid-body, incremental construction and stochastic algorithms.

**Rigid-body docking**

Rigid-body docking algorithms use either single or multi-conformation databases to account for ligand flexibility. The molecules are fitted into the binding sites of proteins by shape complementary or by interaction matching algorithms. These are the fastest structure-based virtual screening methods, but their accuracy may be limited due to the fact that ligand conformation is not refined at the binding site.

An example of rigid-body docking software is FRED from OpenEye Scientific Software. It uses pregenerated multi-conformation database as its input. First, all possible poses of the ligand around the active site are enumerated for each of the conformations. These poses are then filtered, based on the volume of the active site. The remaining poses are then scored with a scoring function. FRED is one the fastest docking program currently available, as it requires just a few seconds per ligand.

**Incremental construction docking**

There are also docking programs based on incremental construction algorithms. These programs build up the ligand in the active site. First, the ligand is fragmented and one fragment is selected as the anchor fragment. The anchor fragment is then rigidly docked into

the active site and the other fragments are connected with the knowledge of preferred conformations. FlexX is an example of a program that is based on incremental construction .It uses a pose-clustering technique similar to those used in pattern recognition.

**Stochastic docking**

Both multi-conformation and incremental construction docking algorithms are deterministic. There are also stochastic docking algorithms available that have a random element in them. Therefore, they do not usually produce exactly the same results in every run. The two most widely used stochastic approaches are Monte Carlo methods and genetic algorithms.

**Monte Carlo methods**

Monte Carlo (MC) methods are stochastic techniques; this means they are based on the use of random numbers and probability statistics to investigate problems. The MC methods are used in many areas from economics to nuclear physics to regulating the flow of traffic.

MC methods are based on repeated random sampling. The ligand to be docked is randomly rotated and translated one parameter at the time. The modified conformation is then evaluated by a scoring function. If the new conformation has a lower energy than the previous one, it is kept. The process is repeated until a satisfactory pose has been generated. A typical example of Monte Carlo docking method is ICM. GLIDE has also a Monte Carlo element, as final poses from hierarchical filtering are generated by the Monte Carlo method.

**Genetic algorithms** are based on Darwin's theory of evolution. A docking pose is stored in a data structure called a "chromosome", which is made up of numbers called "genes" that store each translational angle, rotation and translation of the ligand. Chromosomes then evolve through a process of reproduction and are altered by genetic operators like mutation and crossover. The next generation is then selected by the survival of the fittest, where the two lowest energy chromosomes are kept.

**Table 3. The Genetic Algorithims take inspiration from Nature**

| GA | Nature | Description |
|---|---|---|
| Population | Population | The set of individuals in a given time point |
| Chromosome/ solution. Information is hold in strings | Individual ( information is in the DNA, one or more chromosomes) | Individual of the population and the information it carries |
| Parameter | Gene | Basic information unit |

| | | |
|---|---|---|
| Value assign to the parameter | Allele | The content of information in the basic unit |
| All parameter values | Genotype | Entire coded information |
| Fitting value | Phenotype | How information is expressed |
| Scoring function | environment | Selector |

**Genetic algorithms** are based on the Darwinian principles of natural selection and evolution. They manipulate a population of potential solutions to an optimization (or search) problem. The genetic algorithm cycle *is repeated until a satisfactory solution* to the problem is found or some other termination criteria are met.

**Lamarckian Genetic Algorithm (LGA)**[2]is a modification of the genetic algorithm that is used in AutoDock.LGA is hybrid method which contains an adaptive global optimizer with a local search. The local search method uses a random search optimization, which is allowed to change the chromosome of the global optimizer. The use of LGA instead of the regular genetic algorithm increases the performance of AutoDock.

**Tabu Searches**

– Start with random conformation.

– Generate about 100 new ones from current

– Pick new current (best)

– Generate 100 repeat.

– If new 100 do not contain a better one than Current, pick from old currents

**Molecular Dynamics**

There are many programs to perform molecular dynamics (MD) simulations such as AMBER [32] and CHARMM [25]. MD involves the calculation of solutions to Newton's equations of motions. Using standard MD to find the global minimum energy of a docked complex is difficult since traversing the rugged hyper surface of a biological system is problematic. Often an MD trajectory will become trapped in a local minimum and will not be able to step over high energy Conformational barriers. Thus, the qualities of the results from a standard MD simulation are extremely dependent on the starting conformation of the system.

**Energy minimization**

Energy minimization methods, such as direct searches, gradient methods, conjugate-gradient methods, second-derivative methods, and least-squares methods have been used inconjunction with other docking algorithms in order to improve the initial structures of theligand to be docked.

Examples where energy minimization is used for refinement include DOC and ICM. Energy minimization should not be used alone as a docking algorithm, as it generally reaches localminima only. In general, the main consideration in selecting a docking program is the balance between speed and performance in covering the relevant conformational space[1].

**Table 4. Main Types of Flexible-Ligand and Flexible-Receptor Search Algorithims**

| Flexible-ligand docking |
|---|
| *Systematic*<br>Conformational<br>Fragmentation<br>Database<br>*Random / stochastic*<br>Monte carlo (MC)<br>Genetic algorithm (GA)<br>Tabu search<br>*Simulation methods*<br>Molecular dynamics (MD)<br>Energy minimization |
| Flexible- protein docking |
| Molecular dynamics (MD)<br>Monte carlo (MC)<br>Rotamer libraries<br>Protein- ensemble grids<br>Soft- receptor modeling |

Table search algorithm[7]

**Table 5: Currently available docking programs (adapted and modified from Moitessier et al.2008). The most commonly used programs are shown in italics (McInnes 2007). Licenseabbreviations: O=Open Source, F=Free, FA=Free for Academic use and C=Commercial.**

| Program | References | Search algorithm | Lic. | Website |
|---|---|---|---|---|
| Autodock | Morris *et al.*2009 | LGA | O | Autodock scipps.edu |
| Autodock | Trot and Oison 2010 | Gradient optimization based | FA | Vina scripps.edu |

| | | | | |
|---|---|---|---|---|
| CDOCKER of Discovery Studio | Wu *et al.* 2003 | MD/ simulated annealing | C | www.accelrys.com |
| CHARMM(Galgor) | Veith*et al.* 1998 | GA/MC | C | www.charm.org |
| DOCK | Lang et al. 2009 | sphere matching | FA | Dock.combio.ucsf.edu |
| Dockth | | sphere mathcing | C | www.simbiosys.com |
| eHiTS | Zsoldos*et al.* 2006 | rigid docking of fragments | C | www.simbiosys.com |
| DAIM-SEED-FELD | Kolb and Caflish 2006; Majeux*et al.* 2001; Budin Et Al. 2001 | docking of fragments | F | www.biochemcaflish.uzh.ch |
| FITTED | Corbeil*et al.* 2007 | GA | FA | www.fitted.ca |
| LibDock of discovery studio | Diller *et al.* 2001 | pregenerated ligand conformations with gradient based optimization | C | www.accelrys.com |
| ligand fit of discovery- studio | VenkataChalame*t al.* 2003 | shape-based method with mc | C | www.accelrys.com |
| flex | Rarey*et al.* 1996 | incremental construction | C | www.accelrys.com |
| FlipDock | Zhao and Sanner 2007 | ga | FA | Flipdockscripps.edu |
| FRED | Mcgann*et al.* 2003 | gaussian docking function | FA | www.eyesopen.com |
| FTDock | Gabb*et al.* 1997 | fourier correlation algorithm | O | Bmm.cacerresearchuk.org |
| GEMDOCK | Yang and Chen 2004 | GA | F | Gemdocklife.nctu.edu.tw |
| GlamDOCK | Tietze and Apostolakis 2007 | MC | FA | www.chil2.de |
| GLIDE | Verdonk Et Al. 2003 | GA | C | www.ccdc.cam.ac.uk |
| HADDOCK | Devries *et al.* 2007 | mainly for protein-protein docking | FA | www.nmr.chem.uu.nl |
| MolDock | Thomson and Christensen 2006 | heuristic search | C | www.molegro.com |
| patchDOCK | Schneidman-Duhovny *et al.* 2005 | shape complementary | FA | Bioinfo3d.cs.tau.ac.il |
| PLANTS | Korbe*et al* . 2009 | ant colony | FA | www.tcd.unikonstanz.de |

| | | | | |
|---|---|---|---|---|
| | | optimisation | | |
| ICM | Abagyan*et al.* 1994 | MC | C | Moisoft.com |
| rDOCK | Morley and Aeshar 2004 | MC | F | www.ysbl.york.ac.uk/edoc |
| ROSETTA-LIGAND | Meiler and Baker 2006 | MC | FA | www.roettacommnds.org |
| SLIDE | Schnecke and Kuhn 2000 | incremental construction | FA | www.bch.msu.edu/-kuhn |
| SODOCK | Chen *et al.* 2007 | particle swam optimisation for autodock 3.05 | F | Icaiab.life.nctu.edu.tw/sodoc |
| SurfleX-Dock | Jan 2007 | incremental construction | FA | www.biopharmics.com |
| MOE-Dock | - | MC | C | www.chemcomp.com |

**Scoring functions**

**Overview**

A large number of current scoring functions are based on force fields that were initially designed to simulate the function of proteins. A force field is an empirical fit to the potential energy surface in which the protein exists and is obtained by establishing a model with a combination of bonded terms (bond distances, bond angles, torsional angles) and non-bonded terms (van der Waals and electrostatic). Some scoring functions used in molecular docking have been adapted to include terms such as solvation and entropy. A separate approach is to use pure empirical scoring functions that are derived using multivariate regression methods of experimental data.

**Scoring functions**

The scoring functions can be roughly divided into force field, empirical and knowledge-based .Scoring functions can be also hybrids of molecular mechanics and empirical terms .

**Force field-based scoring functions**

Molecular mechanics force fields are used in scoring functions to calculate the protein-ligand interaction energy and the internal ligand energy. The two factors contributing to the energy are van der Waals and electrostatic terms. Van der Waals energy is most often described by a Lennard-Jones potential (also known as the 12-6 potential)[2]:

$$E_{vdw}(r) = \sum_{j=1}^{N_A} \sum_{i=1}^{N_B} 4\varepsilon \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right]$$

where*NA* and *NB* are the number of atoms in molecules A and B, *r* is the distance between atoms *i*and *j*, *σ* is the collision diameter between atoms *i*and *j*, and *ε* is the well depth of the potential. Different modifications of Lennard-Jones potential have been formulated. For example, the 12-10 potential is used in AutoDock to model hydrogen. The electrostatic potential energy is calculated from a Coulombic equation:

$$E_{cou}(r) = \sum_{i=1}^{N_A} \sum_{j=1}^{N_B} \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}$$

Where $N_x$are the number of atoms in molecule x, $\varepsilon_0$the electric constant and $q_y$is the charge of each atom y. These kinds of descriptions suffer from obvious serious limitations such as modeling protein-ligand binding in water, as they were originally formulated to model gas-phase interactions and do not take solvation or entropy into account. Furthermore, arbitrary cut-off values are required for modeling of non-bonded interactions, which complicates the estimation of long distance interactions. Given these limitations, additional terms besides van der Waals and Coulombic energy have been added to the scoring functions. The AutoDock scoring function includes a desolvation potential Esol based on the general approach by Wesson and Eisenberg. It has an atomic solvation parameter $S_i$ and volume $V_i$ of the atoms surrounding given atom i:

$$S_i = A_i + Q * |q_i|$$

$$E_{sol} = \sum_{i,j} (S_i V_j + S_j V_i) e^{\left(\frac{-r_{ij}^2}{2\sigma^2}\right)}$$

Where σ is a distance weighting factor, $A_i$ and Q are atomic solvation parameters calibrated using six atom types[2].

**Empirical scoring functions**

Scoring functions can also take advantage of existing experimental data. Empirical scoring functions are derived with regression analysis from determined binding energies and/or crystallography data. The concept was originally implemented in de novo design program LUDI and since then, several empirical scoring functions have been proposed. Empirical scoring function estimate the binding affinity of a complex on the basis of a set of weighted energy terms

$$\Delta G = \sum_i W_i . \Delta G_i$$

Where $\Delta G_i$represents different energy terms such as VDW energy, electrostatics, hydrogen bond, desolvation, entropy, hydrophobicity, etc. The corresponding coefficients $W_i$are determined by fitting the binding affinity data of a training set of protein–ligand complexes with known three-dimensional structures[8].

**Knowledge-based scoring functions**

Knowledge-based scoring functions are also very quick to calculate .They are designed to reproduce experimentally observer structures instead of devising predictions of affinity like empirical scoring functions. As the name implies, knowledge-based scoring functions use data about protein-ligand interactions. Pre-defined atom-pair interactions are used to evaluate the docking pose. The principle behind knowledge-based scoring functions is simple: Pair wise potentials are directly obtained from the occurrence frequency of atom pairs in a database using the inverse Boltzmann relation. For protein–ligand studies, the potentials are calculated by:[2]

$$W(r) = -k_B T \ln[g(r)], \quad g(r) = \rho(r)/\rho^*(r)$$

where $k_B$ is the Boltzmann constant, T is the absolute temperature of the system, ρ(r) is the number density of the protein–ligand atom pair at distance r, and ρ*(r)is the pair density in a reference state where the interatomic interactions are zero.

**Consensus scoring**

Despite a good number of scoring functions that have been developed, none of them is perfect in terms of accuracy and general applicability. Every scoring function has its advantages and limitations. To take the advantages and balance the deficiencies of different scoring functions, the consensus scoring technique has been introduced to improve the probability of finding correct solutions by combining the scores from multiple scoring functions. The critical step in consensus scoring is the design of an appropriate consensus scoring strategy of individual scores so that the true modes/binders can be discriminated from others accordingly. Commonly used consensus scoring strategies include vote-by number, number-by-number, rank-by number, average rank, linear combination, etc.111 Examples of consensus scoring are MultiScore, X-Cscore, GFscore, SCS, and SeleX-CS

**Table 6: Currently available scoring functions (adapted and modified from Moitessier*et al.*2008).**

| Scoring function | References | Types | Software or Website |
|---|---|---|---|
| ChemScore | Eldrige*et al.* 1997 | Empirical | GOLD,FRED, CSore |
| shapeGauss | meGann*et al.* 2003 | Empirical | FRED |
| chemGauss | - | Empirical | FRED |
| eHiTs | Zsoldos*et al.* 2006 | Empirical | EHiTs |
| Glidescore | Friesner*et al.* 2004 | Empirical | Glide |
| FlexX | Rarey*et al.*1996 | Empirical | FlexX |
| Hammerhead | Pham and Jain 2006 | Empirical | Surflex- Dock, Discovery Studio |
| LigScore | Lrammer*et al.* 205 | Empirical | Discovery Studio |
| PLP | Verkivker*et al.* 2000 | Empirical | Discovery Studio, FRED, Dockit |
| rankScore | Moitesier *et al.* 2006 | Empirical/FF | FITTED |
| ScreenScore | Stahl and rarey 2001 | Empirical/ consensus | FRED |
| SLIDESCORE | Schnecke and Kuhn 2000 | Empirical | SLIDE |
| X-Score | Wang *et al.* 2003 | Empirical/ consensus | Sw16.im.med.umich.edu/software/xtool |
| AutoDock4SF | Huey *et al.* 2007 | FF/Empirical | AutoDock, SODOCK |
| DockScore | Meng*et al.* 1992 | FF | DOCK,Csore |
| Zou GB/SA Score | Liu *et a*l.2004 | GB/SA | DOCK |
| GoldScore | Jones *et al.*1997 | FF | ICM |
| HADDOCK | Van Dijk*et al.* 2006 | FF | HADDOCK |
| ICM | Abagyan*et al.* 1994 | FF | ICM |
| DrugScoreCSD | Velee*et al.* 2005 | Knowledge based | Pc1664.pharmasize.unmarburg.de/drugscore |
| DrugScorePDB | Gohlke*et al.* 2000 | Knowledge based | Pc1664.pharmasize.unmarburg.de/drugscore |
| M-Score | Yang *et al.* 2006 | Knowledge based | Sw16.m.med.umich.edu/lab/members/chaoyie |
| PMF | Muegge 2006 | Knowledge based | Discovery Studio, dockIt, Csore |
| Zapbind | Grant *et al.* 2001 | Empirical/ PBSA | FRED, DOCK |

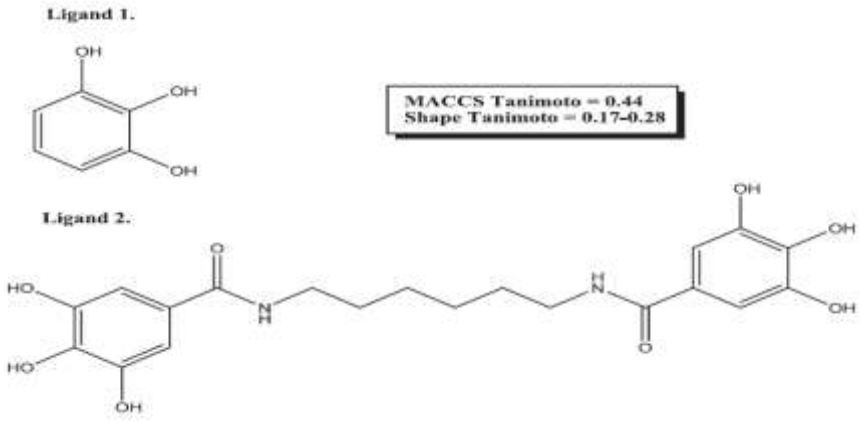
| | | | |
|---|---|---|---|
| Astex Scoring potential | Mooji and Verdonk 2005 | Knowledge based | GOLD |
| Cscore | Clark *et al*. 2002 | Consensus | SYBYL |
| LUDI | Bohm*et al*. 1998 | Empirical | Discovery Studio |
| ASE | - | Gaussian | MOE |
| London dG | - | Empirical/FF | MOE |

## THE LIMITATIONS OF VIRTUAL SCREENING

Even though virtual screening has been successful in drug discovery projects, there are some fundamental limitations in both LBVS and SBVS that are good to keep in mind when designing new experiments.

### Limitations of LBVS

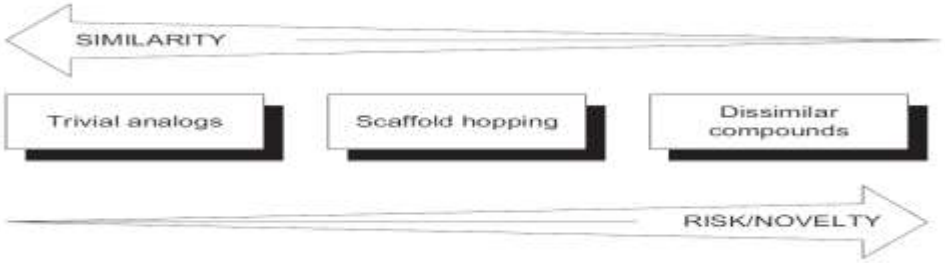
The first limitation of LBVS is the classical chicken and an egg problem: at least one biologically relevant molecule must be identified before database can be screened. This is a major limitation as there are many potential targets for which there are known ligands available. It is unreasonable to expect something completely different from a methodology that is based on searching for similar molecules. The issue is illustrated on Fig. 11, which shows two inhibitors for the catecholamine-O-methyltransferase (COMT) enzyme. They have both low 2D- and 3D-similarities even though they have similar biological activities. Total similarity based on a single molecule is therefore a relatively limited technique. This problem is alleviated by the fact that often several active molecules are known.




**Fig; 11: Global similarity metrics miss sub-structural similarity. The two ligands of the COMT-enzyme have both low 2D- and 3D-similarities.**

There is also a clear paradox in the whole fundamental idea of finding novel bioactive molecules from LBVS, since there is the similarity principle that states that structurally

related compounds display similar biological activities. This of course means that the more different compounds that there are, the less likely they are going to have similar activity. Even though there are various ways to measure the similarity between two molecules, there is always a tradeoff between scaffold hopping and the probability of finding an active compound (Figure 11). It depends on the project if one wishes to find rather similar compounds with a high probability of being active or simply a large number of diverse compounds.




**The tradeoff between chemical similarity and the probability of finding an active compound.**

LBVS methods that require molecular alignment of multiple compounds, such as pharmacophores, assume that all of the active molecules bind in a similar conformation. Aligning several active conformations simultaneously is far from trivial, as the crystallized structures of protein-ligand complexes have well demonstrated. Two commonly used inhibitors of phosphodiesterase 5(PDE5), sildenafil and tadalafil, both have the same binding pocket, but the alignment is not obvious from the molecular structures alone (Fig.12).

**Fig.12: The binding conformations of two inhibitors of PDE5-enzyme. Sildenafil is in gray (PDB 2H42) and Tadalafil in black (PDB 1XOZ) (Wang et al. 2006; Card et al. 2004).**

**Limitations of SBVS**

X-ray crystallography is a rather difficult and laborsome science and therefore, it is not surprising that the crystal structures of most drug targets are not available. The structures of only a few G-protein Coupled Receptors (GPCRs) have been successfully solved, even though this class accounts for approximately 30% of targets of all marketed drugs. Homology modeling-based structures have been used instead, but it is still unclear if such models are truly suitable for virtual screening. In a recent GPCR modeling and docking contest, most of the 29 homology models submitted were not accurate enough to permit virtual screening .A protein model based on X-ray crystallography is an interpretation of experimental data .Two crystallographers may reach different conclusions from the same diffraction data. For example, a functional group of the bound ligand might be confused with a water molecule. This subjective nature of X-ray crystallography is often ignored when utilizing structures from Protein Databank.

In addition to the issues related to X-ray crystallography, there are major problems with current docking methods. The assumption that there is a rigid protein over-simplifies the modeling of protein-ligand interaction. The inductive effects are rarely considered and therefore the binding pocket may be of the wrong shape. A greater problem is that a macromolecular complex is not a single structure, but an ensemble of structures .Changes in conformations of both ligand and protein during the binding have a significant impact on the binding energy. Scoring functions assume that binding free energy can be formulated by additive terms from various protein-ligand interactions. In reality, different molecular interactions are nonadditive and should be designated with different amounts of Gibbs energy in different contexts .Another serious deficiency in docking is that it does not take enthalpy-entropy compensation properly into account An increase in entropy can compensate for a loss in enthalphy.

A good example of this phenomenon is the study of Christof and co-workers on a pair of thrombin inhibitors. The cyclopentyl group of the first compound was switched to cyclohexyl group in the second molecule. Both compounds had identical binding affinity even though X-ray crystallography indicated that the cyclopentyl group was located inside the binding pocket, whereas the cyclohexyl group was not. This similar binding affinity with a different binding mode was caused by enthalphy-entropy compensation as revealed by isothermal titration calorimentry. It is highly doubtful that this phenomenon would have been detected

from molecular docking studies. One can indeed wonder how docking can work at all, given all of these problems .There are successful structure-based virtual screening studies where novel biologically active compounds have been identified, but rarely has the docking pose been experimentally validated by comparing it to the crystallized structure. It is therefore possible that at least some of the reported findings are either based on crude features like molecular shape or just sheer luck. Indeed, for more sophisticated tasks like lead optimization, molecular docking does not seem to be a reliable enough technique.

**REFERENCES**

1. Pedro Fong, Hui Lei, A Systematic Review On Computer-Aided Drug Design, Docking And Scoring, Journal Of Macao Polytechnic Institute 2010, pp.- 47-51
2. TumoKalliokoki, Accelerating Three Dimensional Virtual Screening, Dissertations In Health Sciences, 2010
3. E.M. Krovat, T. SteindlAnd T. Langer, Recent Advances In Docking And Scoring, Current Computer- Aided Drug Design, 2005,1, 93-102
4. Gwyn S. Skone, Stratagems for Effective Function Evaluation in Computational Chemistry, May 2010
5. Rajmund Kazamierkiewicz, Introduction to Molecular Modelling, Intercollegiate Faculty of Biotechnology UG-MUG
6. R.D. Taylor, P. J. Jewsberry, J. W. Essax, A Review Of Protein- Small Molecule Docking Methods, Journal Of Computer- Aided Molecular Design , 16: 151-166, 2002.
7. Beanhard Knapp, An Introduction Into “Docking” And “Molecular Dynamics Simulations”, 2012.
8. Sheng-You Huang, Sam Z. Grinterand Xiaoquinzou, Scoring Funtions And Their Evaluation Methods For Protein-Ligand Docking: Recent Advances And Future Directions, Physical Chemistry Chemical Physics, 2010, 12, 12899-12908.