ISSN- O: 2458-868X, ISSN –P : 2458 – 8687

International Journal of Medical Science and Innovative Research (IJMSIR)

IJMSIR : A Medical Publication Hub

Available Online at: www.ijmsir.com

Volume – 1, Issue – 2, July - August 2016, Page No. : 19 - 41

Molecular Docking an Imperative Methodology for Drug Discovery- A Systemic & Comprehensive Review Study

Virender Kumar[1], Kunal[1], Varun Arora[2], Neha Minocha[3], Davinder Kumar*[1]

. [1]College of Pharmacy, Pt. B.D. Sharma University of Health Sciences, Rohtak, Haryana 124001
[2]R.K.S.D College of Pharmacy Kaithal –Haryana-136027
[3]Department of Pharmaceutical Sciences & Research, Baba Mast Nath University- Rohtak-124001

Corresponding Author: Davinder Kumar, Assistant professor, College of Pharmacy, Pt. B. D. Sharma University of Health Sciences, Haryana, Rohtak, 124001

Abstract: Molecular docking and Quantitative structure activity relationship (QSAR), is now established as an important approaches in drug discovery and most important areas in chemistry, gives information that is useful for drug design. Two major bottleneck of molecular docking are availability of an efficient docking algorithm & availability of a selective and efficient scoring function. Comparison suggests that the best algorithm for docking is probably a hybrid of various types of algorithm encompassing novel search and scoring strategies. These are mathematical equations relating chemical structure to a wide variety of physical, chemical and biological properties. The derived relationship between molecular descriptors and activity are used to estimate the property of other molecules and/or to find the parameters affecting the biological activity.

Key words: Molecular docking, CADD, QSAR, Docking Software.

1. Introduction

The drug design and development process involves use of variety of computational techniques, such as structure–activity relationships (SAR), quantitative structure–activity relationships (QSAR), molecular mechanics, quantum mechanics, molecular dynamics, and drug-protein docking [1,2]. Quantitative structure–activity relationship (QSAR) studies are based on the premise that biological response is a function of chemical structure [3]. The QSAR establish a statistical relationship between biological activity or environmental behavior of the chemicals of interest and their structural properties [4,5]. QSARs predict chemical behavior directly from chemical structure and simulate adverse effects in cells, tissues and lab animals, minimizing the need to use animal tests to comply with regulatory requirements for human health and eco-toxicology endpoints. [6]. The fundamental hypothesis in QSAR is that similar chemicals have similar properties, and small structural changes result in small changes in property values [7]. SAR represent classification models that are used when an empirical property is characterized in a (+1/-1) manner, such as soluble/insoluble, active/inactive, inhibitor/non-inhibitor, ligand/non-ligand, substrate/non-substrate, toxic/non-toxic, mutagen/non-mutagen, or carcinogen/non-carcinogen [8]. In silico screening is typically a low cost high-throughput process, which can provide a fast indication of potential hazards for use in lead prioritization [9].

Machine learning (ML) is an important field of artificial intelligence in which models are generated by extracting rules and functions from large datasets. ML includes a diversity of methods and algorithms such as decision trees, lazy learning, k-nearest neighbors, Bayesian

methods, Gaussian processes, artificial neural networks (ANN), artificial immune systems, support vector machines and kernel algorithms. Machine learning algorithms extract information from experimental data by computational and statistical methods and generate a set of rules, functions or procedures that allow them to predict the properties of novel objects that are not included in the learning set. (Q)SAR models based on machine learning algorithms are applied during the drug development cycles to optimize the biological activity, target selectivity, and other physico-chemical and biological properties of selected chemical compounds [10]. The advantage of AI approaches is that they can be applied to learn from examples and develop predictive models even when our understanding of the underlying molecular processes is limited, or when computational simulations based on fundamental physical models are too expensive to carry out [11,12]. The number of proteins with a known three dimensional structure is increasing rapidly and structure produced by structural genomics initiatives are beginning to publicly available. The increase in number of structural targets is in part due to improvements in technique for structure determination such as high throughput X-ray crystallography. The action of drug molecules and the function of proteins targets are governed by principles of molecular recognition [13]. Binding events between ligands and their receptors in biological systems form the basis of physiological activity and pharmacological effects of chemical compounds. Accordingly, the rational development of new drugs requires an understanding of molecular recognition in terms of both structure and energetic. Docking and virtual screening are computational tools to investigate the binding between macromolecular targets and potential ligands. They constitute an essential part of structure-based drug design, the area of medicinal chemistry that harnesses structural information for the purpose of drug discovery.

Docking of small molecules to protein binding sites was pioneered during early 1980 and remains a highly active area of drug research. When only the structure of a target and its active site is available, high throughput docking is primarily used as a hit identification tool. Furthermore, docking can also be contributed to the analysis of drug metabolism using structure such as cytochrome P450 Isoforms [14].

Molecular docking can be defined as follows:

It is a term used for computational schemes that attempt to find the 'best' matching between two molecules: a receptor and a ligand [13]. (Figure 1)

The subject of docking is the formation of non-covalent complexes.

Given two molecules molecular docking determines:

Whether two molecules interact. If they interact then what is orientation that maximizes the 'interaction' while minimizing the energy of the complex [14].

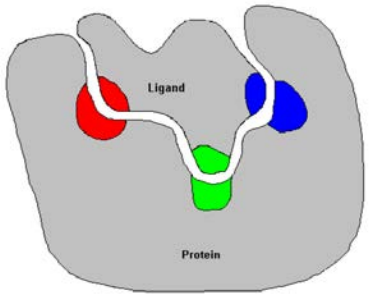

Figure 1: A best match between a protein and a ligand molecule

"Docking is actually an 'energy optimization problem' concerned with the search of lowest free energy binding mode of a ligand with a protein binding site." Example of Docking: HIV-Protease [13].

Figure 2(a): Active site of HIV-Protease.

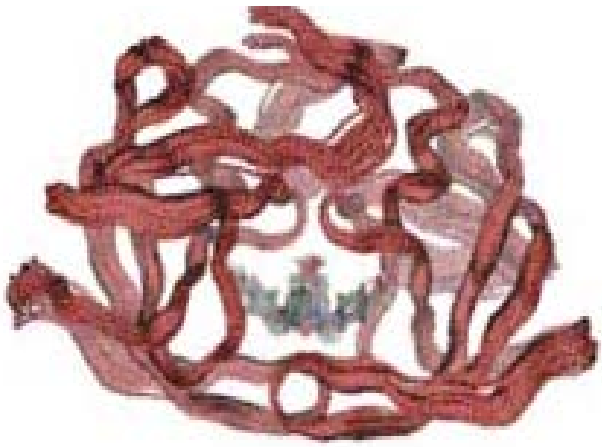

Figure 2 (b): Inhibitor bound with active site.

2. Types of docking

***1. Blind Docking*:** In this type active site of the protein is not known and search for both the binding site and subsequently the binding mode of ligand is required. It is important for investigating protein-protein interactions. A special sub case of blind docking is the docking to homology models of a target where the position of active site is assumed to be similar to the one in a template protein. Docking to models of Trans-membranes proteins, such as G-proteins coupled receptors, falls into this category.

2. *Direct Docking:* If the active site of the binding is known from X-ray diffraction or from NMR studies, docking into the known active is called a 'direct docking'. During direct docking, certain factors such as presence of cofactors, discrete crystal molecules of water, and catalytic metal ions in the active site of protein or ionized states of the compounds as well as the effects of the pH, induced fit and conformational changes of proteins must be taken into account if they are participating in protein-ligand interaction[15].

3. Requirements Of Molecular Docking

The set up for a ligand docking approach require the following components; a three dimensional structure of target protein with or without 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 docking procedure. Two main components of computational framework in the drug designing are an efficient searching procedure and good scoring function [16].

***Searching procedure*:** is used to explore the configuration space accessible for the interaction between the two molecules. The goal of this exploration is to find the orientation and conformation of interacting molecules corresponding to global minimum of the free energy of binding. Two critical elements in searching procedure are speed and effectiveness in covering the relevant conformational space.

Scoring: Scoring functions in docking procedure is used to evaluate and rank the configurations generated by the search process. It should be fast enough to so that it can be applied to number of potential solutions. Scoring function should include and appropriately weigh all the energetic ingredients. To solve the docking problem, ideally the best

matching algorithms and scoring schemes should be com bined [17].

Steps Involved in Docking Procedure

Docking procedure involves the following steps:

4.1. Target selection for binding- mode assessment

Protein-ligand complexes are selected from the protein Data Bank according to the following criteria:

General Features

- Non-covalent binding between ligand and protein
- Crystallographic resolution around 3.0 Å or better Ligand features
- Molecular weight between 150 and 800 Da
- From 1 to 16 rotable bonds,
- Drug lead/nonlead like
- Structurally diverse
- Protein features
- Multiple structural motifs (wide spectrum of receptor families)
- Metal present in some of the binding pockets
- Range of the active site topologies and water accessibility
- Relevant for drug discovery

4.2. Receptor preparation for Binding-Mode Assessment

Generally, if a cofactor is present at the binding site, its bond order and protonation state are inspected and corrected if required. When relevant, metal ions at the binding site are preserved. All the crystallographic waters are deleted from the binding pockets expect of few tightly bound to the pocket. After removal of the ligand, solvent, and cofactor (when the latter two are not intrinsic parts of the binding site), additional domains not involved in ligand binding, stabilizing counter ions, and other extraneous small molecules far from the active site are also removed. Residues at the binding site of each receptor are then visually inspected, hydrogen's are added along with missing heavy atoms and partial charges, corrections are made to the orientations of hydroxyl groups and disulfide bonds, and the tautomeric states of histidine residues and the protonation states of basic and acidic residues are adjusted to be the dominant species at pH 7.0 [18].

Representation of system: There are three basic representation of the receptor; atomic, surface and grid. Among this atomic representation is generally used in conjunction with a potential energy function and often during final ranking. Surface based docking programs are typically used in protein-protein docking. These methods allow aligning points on surfaces by minimizing the angle between the surfaces of opposing molecules.

Grid: The basic idea in grid representation is to store information about the receptor's energetic contributions on grid points so that it only needs to be read during ligand scoring. Grid points stores two types of potentials; electrostatic and vander Waals.

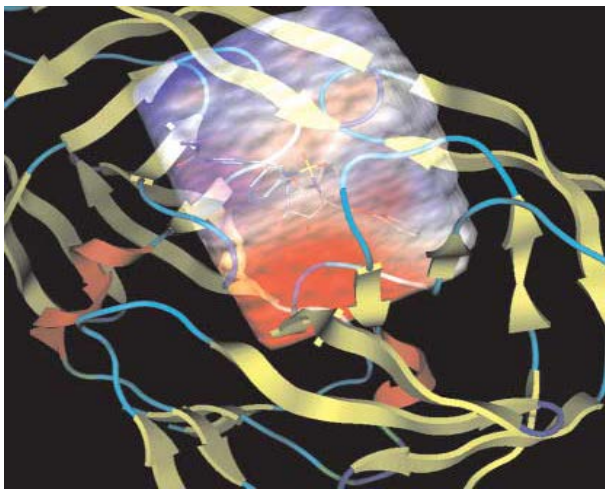

Figure 2 (a)

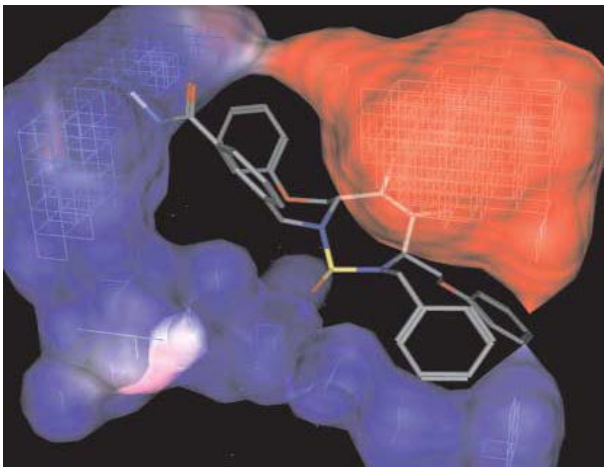

Figure 2 (b)

Figure2: Grid representation (a) shows surface plot of grid capturing the electrostatic potential of HIV-protease around its active site (b) shows electrostatic potential grid of the enzyme around the bound inhibitor [13].

4.3. Ligand Preparation for Binding-Mode Assessment

The X-ray coordinates of the ligands are extracted from each of the protein receptors. Each ligand is examined for bond order and protonation state, and written out as a three dimensional "reference ligand" .The ionization states of the ligands which we are attempting to dock are also of particular concern .A subset of commercially available chemicals is prepared by randomly selecting from larger data set compound structures. These compounds are also selected to satisfy the following criteria:

General Features

- Molecular weight between 150 and 750 Da
- Number of rotatable bonds less than 7
- At least one polar atom (N, O, S, or P)

Actives, decoys (compounds that are similar to the active compounds in every respect except for activity), and random selections (compounds that bear little resemblance to the active ligands) are selected with a similar distribution of molecular weight, to minimize the well-known tendency of a scoring function to favor larger molecules [18].

5. Scoring Functions

The purpose of scoring function is the identification of the correct binding pose by its lowest energy value and the ranking of protein ligand complexes according to their binding affinities. The aim of scoring is to compare the free energies of hundred or thousands of protein ligand complexes as generated by virtual screening.

Scoring is usually composed of three different aspects relevant to docking and design:

I .Ranking of configurations generated by the docking search for one ligand interacting with a given protein; this aspect is essential to detect the binding mode.

II. Ranking different ligand with respect to the binding to one protein, that is prioritizing the ligands according to their affinity; this aspect is essential in virtual screening.

III. Ranking one or different ligands with respect to their binding affinities to different proteins; this aspect is essential for consideration of selectivity and specificity.

If one were able to measure free energy of binding, all three aspects would be satisfied simultaneously [14]. There are three types of scoring functions

5.1. Empirical Scoring Functions

- LUDI
- CHEM SCORE
- F-Score
- X-Score

5.2. Force-Field Based Scoring Functions

- D-Score
- G-Score

- GOLD
- DOCK

5.3. Knowledge Based Functions

- PMF
- Drug Score
- SMOG[13]

5.1. Empirical Scoring Functions

These scoring functions are first proposed by 'Bohm'. The underlying idea for this scoring function is that binding free energy can be interpreted as a weighted sum of localized interaction terms. The interaction terms typically represent hydrogen bonding terms, ionic interactions, hydrophobic interactions, entropy change associated with binding. The interaction terms are usually calculated using experimental 3D structures of receptor ligand complexes. A typical empirical function:

$\Delta G = \Delta G_0 + \Delta G_{rot} * N_{rot}$ loss of entropy during binding

$+\Delta G_{hb} \sum f(\Delta R, \Delta\alpha)$ hydrogen bonding

$+\Delta G_{io} \sum f(\Delta R, \Delta\alpha)$ ionic interactions

$+\Delta G_{aro} \sum f(\Delta R, \Delta\alpha)$ aromatic interactions

$+\Delta G_{lipo} f*(\Delta R)$ hydrophobic interactions

The ΔG coefficients are unknown and are determined by multilinear regression in order to fit the experimentally measured binding affinities. The first terms are constant term taking into account the loss of entropy during ligand binding (ΔG_{rot}: energy loss per rotable bond, Nrot: number of rotable bonds). ΔG_{hb}, ΔG_{io}, ΔG_{aro}, ΔG_{lipo} give the binding energy for each hydrogen bond, ionic interaction, aromatic interaction and for liphophilic interaction respectively. $f(\Delta R,\Delta\alpha)$ is a scaling function penalizing deviations from the ideal interaction geometery in terms of distance (ΔR) and angle ($\Delta\alpha$). The function f *R for contacts with a more or less ideal distance and penalizes forbiddingly close contacts.

Examples of empirical scoring functions showing some promise include Chem Score, X-Score and PLP [15].

PLP: This empirical scoring function can be expressed as

$$E_{total} = E_{H\text{-}bond} + E_{repulsion} + E_{contact}$$

LUDI: Differentiate between neutral & ionic hydrogen bond and calculates hydrophobic contribution on the basis of representation of molecular surface area.

***Chem Score*:** Includes terms for hydrogen bonding, metal-ligand interaction, lipophilic contact & rotational entropy. It does not differentiate between ionic and neutral hydrogen bonding. It evaluates contact between hydrophobic atom pairs.

F-score: An empirical scoring function implemented in docking program flexX and is twist of LUDI scoring function

***X-score*:** It includes vander waals interaction term, a hydrogen bonding term, a hydrophobic effect term, a torsional entropy, and a regression constant [19].

Advantage: Empirical scoring functions are faster than force field based methods.

Disadvantage: One major disadvantage of empirical scoring functions is that it requires a training set to derive the weight factors of the individual energy terms [15].

5.2. Force Field Functions

These methods use non bonding energies of molecular mechanics force fields (ex. AMBER, CHARMM) to estimate the binding affinity. The non bonded interaction energy takes the following form:

$$E = \sum_{i=1}^{lig} \sum_{j=1}^{rec} \left[\frac{A_{ij}}{r^{12}} - \frac{B_{ij}}{r^{6}} + \frac{q_i q_j}{D r_{ij}} \right] + 332$$

Where A_{ij} and B_{ij} are vander waal's repulsion and attraction parameters between two atoms i and j at a distance r_{ij}, q_i and q_j are the point charges on i and j. D is dielectric function and 332 is a factor that converts the electrostatic energy into kilocalories per mole.

Auto-Dock program uses the force field based scoring functions which utilizes the parameters from the AMBER force field. In Auto Dock, the overall docking energy of a given molecule is expressed as a sum of intermolecular interaction between the complex & the internal steric energy of ligand.

G-Score: It is the sum of protein-ligand complexation term, hydrogen bonding term, and an internal energy term.

D-Score: It is a classical force field energy function, which sums vander Waals & electrostatic interactions between the complexes.

Disadvantages: The main drawback of force field calculations is the omission of the entropic component of the free energy of binding. They are time consuming and sensitive to errors in the protein structure models [15].

5.3. Knowledge Based Functions

Knowledge based scoring functions are based on the inverse formulation of the Boltzmann law. The frequency of occurrence of individual contacts is used as a measure of their energetic contribution to binding. A high frequency of occurrence of individual contacts an attractive attraction, while a low frequency indicates a repulsive interaction. These functions include the potential of mean force (PMF) and Drug Score.

PMF: A potential of mean force converts structural information gathered from protein-ligand X-ray coordinates into Helmholtz free interaction energies of protein-ligand atom pairs.

A new version of PMF scoring (PMF04) has been generated using ~10-fold more protein- ligand complexes from the PDB as knowledge base, compared to PMF99. The PMF04 have allowed for the introduction of a metal ion protein atom type and more halogen-containing atom pair potentials. PMF04 and PMF99 have been compared using a series of test sets that were previously used for the validation of PMF99. In most of the reported cases PMF04 performs either slightly or significantly better than PMF99 [20].

Drug Score: *In* the Drug Score equation solvent-accessible surface dependent singlet potentials for protein and ligand atoms are included. A protein-ligand interaction free energy A (r) is then assigned to each interaction type between protein atoms of type i and a ligand atom type j in a distance r_{ij}, depending on its frequency.

$$A(r) = - k_B T \ln g_{ij} (r)$$

Where k_B is the Boltzmann constant, T is the absolute temperature and g_{ij} (r) the atom pair distribution function for a protein -ligand atom pair ij. The distribution function is calculated from the number of occurrences of that pair ij at a certain distance r in a database of protein -ligand complexes (usually the PDB). The scored is defined as the sum over all interactions of the protein-ligand complexes.

SMOG: It is another scoring function that utilizes pair-wise atom potentials to evaluate protein-ligand interaction.

***M-score*:** Knowledge based potential scoring function, which considers the mobility of protein atoms.

Advantages: The main advantage of this approach that there is no need for training set and entropic terms are implicitly included.

Disadvantage: A main disadvantage of these functions is that their derivation is essentially based on the information

implicitly encoded in limited sets of protein-ligand complex structures [15].

6. Docking Methods

Docking algorithms are used to treat ligand flexibility, and to some extent protein flexibility. Treatment of ligand flexibility can be divided into three basic categories: systematic methods, random or stochastic methods, and simulation methods.

***Systematic search*:** These algorithms try to explore all the degree of freedom in a molecule. It is performed by varying systematically each of the torsion angle of the molecule in order to generate all possible conformations. Most popular systematic approach is incremental construction method.

***Random search*:** These algorithms operate by making random changes to either a single ligand or a population of ligands. A newly obtained ligand is evaluated on the basis of pre-defined probability function. Two popular random approaches are Monte Carlo and genetic algorithm.

Simulation Methods: Molecular dynamics is currently the most popular simulation approach. However, the molecular dynamics is unable to cross high energy barriers within the feasible simulation time periods, therefore might only accommodate ligands in local minima in energy surface [13].

6.1. Molecular Dynamics

Molecular dynamics is a simulation technique that solves Newton's equation of motion. For an atomic system: $F_i=m_ia_i$ in which 'F' is force, m is mass and 'a' is acceleration. The force on each atom is calculated from a change in potential energy (usually based on molecular mechanics terms) between current and new positions: $F_i= – (dE/r_i)$, in which r is distance. Atomic forces and masses are then used to determine atomic positions over series of very small time steps:

$F_i= m_i (d^2 r_i/dt^2)$, in which t is time.

This provides a trajectory of changes in atomic positions over time. It is easier to determine time-dependent atomic positions by first calculating accelerations a from forces and masses, then velocities v_i from $a_i=dv_i/dt$ and, ultimately, positions from velocities $v_i= dr_i/dt$[1].

6.2. Monte Carlo Method

Monte carlo simulation method occupies a special place in the history of molecular modeling, as it was the technique used to perform the first computer simulation of a molecular system. This method generates random moves to the system and then accepts or rejects the move based on Boltzmann probability.

Monte Carlo algorithm in its basic form:

- Generate an initial configuration of a ligand in an active site consisting of a random conformation, translation and rotation to minimize the intermolecular overlap.
- Score the initial configuration.
- Generate a new configuration and score it.
- Use a Metropolis criterion (explained below) to determine whether the new configuration is retained.
- Repeat previous steps until the desired number of configurations are obtained [13].

Metropolis criterion

If a new solution scores better than the previous one, it is immediately accepted. If the configuration is not a new minimum, a Boltzmann-based probability function is applied. If the solution passes the probability function test, it is accepted; if not, the configurations rejected. The probability of acceptance P is given as:

$$P = e^{(-\Delta E/Kt)}$$

Where ΔE is the difference in energy from previous step, T is absolute temperature in Kelvin, and k is a Boltzmann

constant. This means higher the temperature of the cycle higher is the probability that the new state is accepted [15]. Programs using MC methods include ProDock, ICM, MCDOCK, Dock vision [21].

6.3. Simulated Annealing

Simulated annealing is a special molecular dynamics simulation in which the system is cooled down at regular time intervals by decreasing the simulation temperature. The system is thus trapped in the nearest local minimum conformation. The disadvantage of simulated annealing is that the result depends upon the initial placement of ligand and the algorithm does not explore the solution space exhaustively. AutoDock2.4 uses Monte Carlo simulated annealing [16].

6.4. Point Complementary Method

These methods are based on evaluating the shape and chemical complementarity between interacting molecules. The interacting molecules are usually modeled in an easy way, for example using spheres or cubes as atoms. The ligand description is then rotated and translated to obtain the maximum number of matching between ligand and protein surfaces minus the number of volume cube (cube inside the molecule) overlaps. The docking solutions are then clustered based on translation vectors and rotation angles. The average value for each cluster is then scored using a geometric sum of atom descriptors, that are based on charges hydrogen bond donors/acceptors and hydrophobicity. Examples of programs using point complementary methods are FTDOCK, SANDOCK [21].

6.5. Genetic Algorithm

The essential idea of genetic algorithm is evolution of population of possible solutions via genetic operators (mutation, crossovers) to a final solution, optimizing a predefined fitness function. Main features of Genetic algorithm;

- Ligand translation, rotation, configuration variables constitute the genes.
- Crossover mixes ligand variables from parent configurations.
- Mutation randomly changes variables.
- Natural selection of current generation based on fitness.
- Energy scoring function determines fitness.
- Some programes using GAs are GOLD, AutoDock, DIVALI, and DARWIN.

Disadvantage: Genetic algorithms require the longest time for single energy calculation and hence are the least efficient [13].

6.6. Tabu Searches

These methods are based on stochastic process, in which new states are randomly generated from an initial state (referred to as current solution). These new solution are then scored and ranked in ascending order. The best new solution is then chosen as the new current solution and same process is then repeated again. Tabu search maintains a tabu list that stores a number of previously visited solutions .Thus by preventing the search from revisiting these regions, the exploration of new search is encouraged. Only one current solution is maintained during the course of a search. The highest ranked move is always accepted as new “current solution” if its energy is lower than lowest energy obtained so far and replaces at the same time the previous “best solution”. An example of docking algorithms using tabu search is PROLEADS [15].

6.7. Incremental Constrution Method

In the incremental construction method ligand is not docked completely at once, but is divided into single fragments, docking the fragments and incrementally reconstructed inside the active site. These methods require subjective decisions in the importance of the various functional groups in the ligand, because a good choice of

base fragment is essential for these methods. A poor choice can significantly affect the quality of the results. The base fragment must contain the predominant interactions with the receptor. Some well known programs using this method are FlexX and DOCK, Hammerhead.

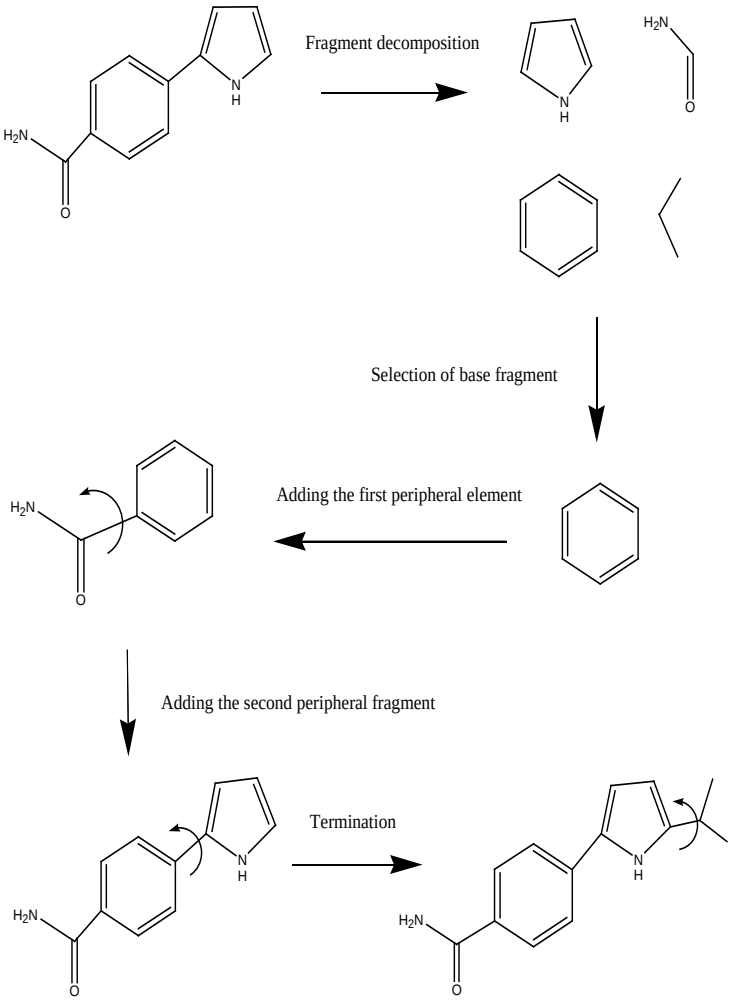

Figure: 3. Incremental construction method [26-27]

7. Assessment of Docking Method

Docking method are usually assessed by their ability to reproduce the binding mode of experimentally resolved protein ligand complexes: the ligand is removed from the complex, a search area is defined around the actual binding site, the ligand is redocked into the protein, and the achieved binding mode is compared with the experimental position, usually in terms of a root-mean-square deviation (rmsd). If the rmsd is below 2Å, it is generally considered as a successful prediction. Virtually any introduction of a new docking method has been accompanied by such a test [14].

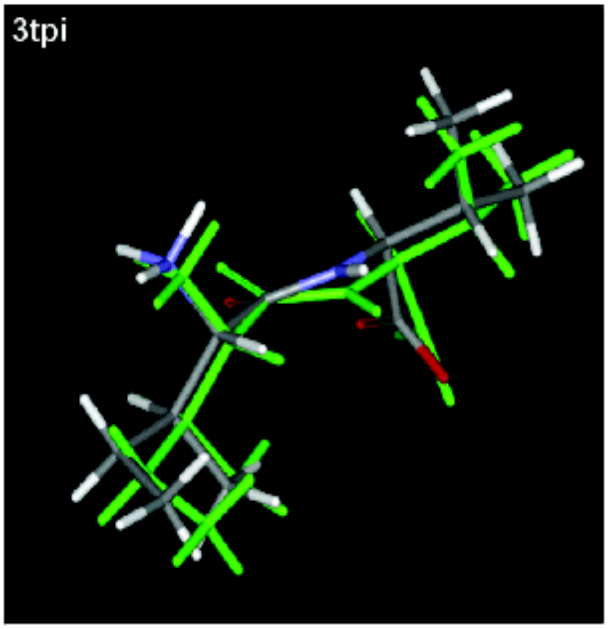

Figure 4(a)

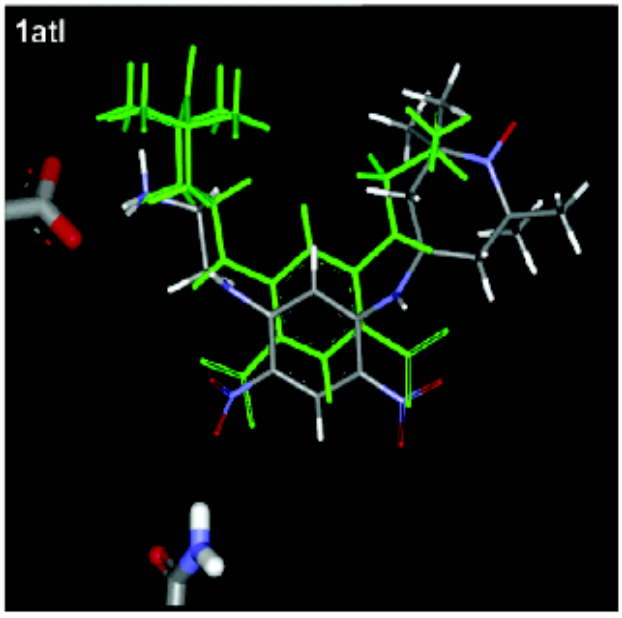

Figure 4(b)

Figure 4(a): shows a good accuracy of docking method and (b) shows a poorly docked cases

Enrichment factor: When the percentage of active compounds in the screening set can be reliably estimated then success is quantified by enrichment factor. It is

defined as the ratio between the percentage of active compounds in the selected subset & the percentage in the entire subset [18].

8. Docking Software

A represented table (1) which shows some of the available docking program.

Docking program	Docking method	Developer	Year published
Dock Dock 4.0	Point complementary method Incremental method	Kuntz et al Ewing & kuntz et al	1982 2001
Auto Dock	Lemarckian genetic algorithm & monte carlo method	Morris et al	1998
GOLD	Genetic algorithm	Jones et al	1997
FlexX	Incremental construction method	Rarey et al	1996
Hammerhead	Incremental construction method	Welch et al	1996
ICM	Monte carlo method	Abagyan et al	1994
MC Dock	Monte carlo method	Liu & Wang	1994
SLIDE	Incremental method	Schnecke et al	2002
Glide	Simulation method	Frienser et al	2004

Among these DOCK, AutoDock, FlexX, GOLD, Hammerhead are the old programs used for molecular docking.

8.1. DOCK

DOCK is oldest and best known ligand-protein docking programs. DOCK uses fragment based method using shape and chemical complementary methods for creating possible orientations for ligand. The initial version used rigid ligands: flexibility was later incorporated via incremental construction of ligand in the binding pocket in DOCK 4.0. DOCK works in 5 steps:

1. Start with crystal coordinates of the target receptor.
2. Generate molecular surface for receptor.
3. Generate negative image of the binding site from the molecular surface of the receptor.\ the negative image consists of sets of overlapping spheres of varying radii. (Figure 5 a)
4. Matching: Ligand atoms are then matched to the sphere centers to find the matching sets in which all the distances between the ligand atoms are equal to the corresponding sphere center-sphere center distances & possible orientations of ligand are determined. (Figure 5 b)
5. Scoring: Find the top scoring of orientation.

DOCK seems to handle well polar binding sites and is useful for fast docking, but it is not the most software.

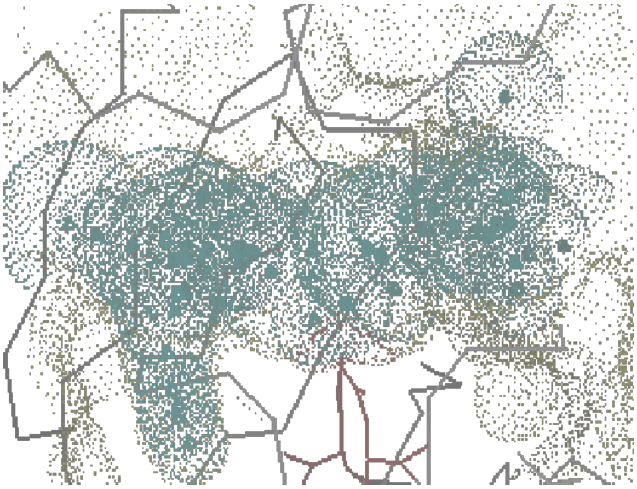

Figure 5 (a)

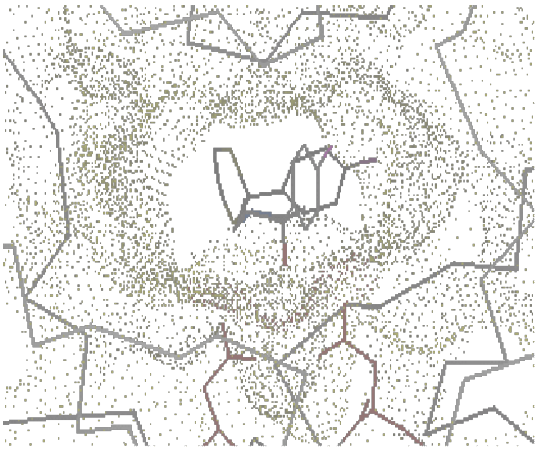

Figure 5 (b)

8.2. AutoDock

Auto dock uses Monte Carlo simulated annealing and Lamarckian genetic algorithm to create a set of possible conformations. LGA is used as a global optimizer &energy minimization as a local search method. Possible orientations are evaluated with AMBER force field model in conjunction with free energy scoring functions. In this implementation ligand is flexible and the receptor is rigid &represented as a grid. The genetic algorithm uses two point crossover & mutation operators. The fitness function comprises five terms: a lennard-jones 12-6 repulsion term; a directional 12-10 hydrogen bond term; a coulombic electrostatic potential; a term proportional to number of sp^3 bonds in the ligand to represent unfavourable entropy of ligand binding due to restriction of conformational degree of freedom & a desolvation term. The algorithm was originally tested on seven complexes and for this test samples all lowest energy structures were within 1.14 $\overset{o}{A}$ RMSD of the crystal structure [21].

8.3. GOLD

GOLD uses genetic algorithm to provide docking of flexible ligand and a protein with flexible hydroxyl groups. Otherwise protein is considered as rigid. This makes a good choice when the binding pocket contains amino acid that forms hydrogen bonds with the ligand. The Gold validation test is one of the most comprehensive (comprising 100 different protein complexes) of all the docking methods reviewed, and achieved a 71% success rate based primarily on visual inspection of the docked structures. 66 of the complexes had an RMSD of $2\overset{0}{A}$ or less, while 71 had an RMSD of $3\overset{o}{A}$ or less.

Disadvantage: The disadvantage is that it does not include hydrophobic interactions and a solvent model which leads to some of docking failures when ligands are hydrophobic & complexes containing poorly resolved active site [21]

8.4. FlexX

FlexX is a fast, flexible docking method that uses an incremental construction algorithm to place ligands into active site. It offers:

- A fast method for docking conformationally ligands.
- A full specification of the active site, including oxidation states, metals ions, side chain protonation states
- Automated ligand positioning.

It divides the ligands into rigid fragment along its rotational bonds, docks first a base fragment into the active site & reattaches the remaining fragments. The remaining ligand components are then incrementally attaches to the core. At each growing step, a list of preferred torsional angle values is read & the best conformation in terms of protein-ligand interactions is maintained for further growing of ligand. Finally, the conformations of the complete ligand with the lowest score are selected. It uses 'Bohm' as a scoring function (i.e. empirical scoring function). It differs significantly from DOCK in the method used for determining the placement of base fragment. Rather than defining points where ligand atoms may be located, FlexX defines interaction sites for each possible interacting group of active site & ligand. It has a lower hit rate than DOCK but provides better estimates of root mean square distance for compounds with correctly predicted binding mode. There is an extension of flexX called flexE which consider receptor as flexible which has shown to produce better results with significantly lower running times [15].

9. Recent Strategies

9.1 MCDOCK: A Monte Carlo simulation approach to the Molecular docking

MCDOCK is a new docking method in which a non-conventional Monte Carlo simulation technique is applied. A computer program MCDOCK, developed to carry out molecular docking operation automatically. The current version of the MCDOCK program allows for full flexibility of ligands in the docking calculations. The scoring function used in this method is the sum of interaction energy between ligand & its receptor, and the conformational energy of the ligand. To validate this method, 19 small ligands, the binding mode of which had been determined experimentally using X-ray diffraction, are docked into their receptor binding sites. Result showed that scoring function used in MCDOCK program is fairly adequate for accurate prediction of the ligand binding mode. Using the binding mode with lowest potential energy as the predicted binding mode, the rms value for these 19 ligands is between 0.25$\overset{0}{A}$ and 1.84$\overset{0}{A}$. The CPU time for each MCDOCK run is from 1 to 15 min for a ligand, depending upon the size & flexibility of the ligand.

Limitation: The scoring function used in this method does not include solvation effect & a better description of intra-molecular interactions for ligands [22].

9.2. Mining Minima Optimizer

Mining minima method is a novel approach in docking studies. This method computes molecular free energy by rapidly identifying the most stable conformations of a molecule. The optimization algorithm draws an idea from the Global Under estimator method, genetic algorithm & tabu search. This method has been adapted in the protein-ligand docking due to two reasons: First, because no free energy calculation is done, the time consuming integration of the boltzmann factor with in each energy well is removed. Second, an exclusion zone of uniform dimensions is placed around each energy minimum as it is discovered, in order to avoid rediscovering it in future docking iterations. Global under estimator method includes the following steps: a collection of local energy minima is generated; the coordinates & energies of these minima is used to construct a concave-up parabola of energy versus conformation that lies at or below each of minima; and a new set of local minima is then generated in the vicinity of the global minimum of the parabolic function, with the idea that the global energy minimum of the actual energy function probably lies near the minimum of the parabolic 'under estimator' function. This new method is competitive in terms of both speed & accuracy, with another energy based methods. In most test, a configuration with RMSD less than 1.5$\overset{0}{A}$ was found with in 25 dockings. One of the most rigid ligand, thiazoline is not docked well, while reasonably good results were obtained for some of the flexible ligands, such as hexadecanesulfonic acid [23].

9.3. DOCK 4.0

The search strategies incorporated into the widely distributed DOCK software include incremental construction method & random conformation search and utilizing the existing columbic & lennard-jones grid based scoring function. The incremental construction strategy is used with a panel of 15 crystallographic test cases. For 7 of the 15 test cases, the top scoring position is also with in 2$\overset{0}{A}$of the crystallographic position.

An important application of DOCK is the screening of a molecule database. Steptavidin & dihydrofolate reductase are used as test sites to which a set of 49 randomly selected molecules from the Current Medical Chemicals molecular database are screened. The test database is seeded with biotin and methotrexate so that at least one tight binding molecule is included. The algorithm is fast enough to successfully dock a few test cases within seconds [24].

9.4. Surflex

Surflex is a fully automatic flexible molecular docking algorithm that combines the Hammerhead empirical scoring function with a search engine that relies on a surface based molecular similarity method. Hammerhead is fragment based docking program. In this program, the head fragments are generated by dividing the ligands into sections. The highest scoring fragments are considered head fragments. To each head fragment tails (remaining fragments) are added one at a time. The best scoring orientations are then retained for addition of next fragment. Scoring function used in this algorithm is the sum of hydrophobic interactions, polar complementarity, entropic terms and salvation terms. Surflex's utility is used as a screening tool on two protein targets (thymidine kinase and estrogen receptor) using data sets on which competing method are also run. Result shows that Surflex is more accurate in terms of rmsd of docking ligands as compared to other methods. Surflex is fast in terms of docking speed and significantly more accurate in terms of scoring to the extent that false positive rates are 5 to 10-fold lower for equivalent true positive rates compared to other methods.

Limitation: scoring function used does not include non-bonded self-interactions with in the ligands and does not account for protein flexibility [25].

9.5 GLIDE 3.5

The GLIDE algorithm approximates a systematic search of positions, orientations, and conformations of the ligand in the protein-binding pocket via a series of hierarchical filters. The shape and properties of the receptor are represented on a grid by several different sets of fields that provide a progressively more accurate scoring of the ligand pose. The fields are computed prior to docking. The binding site is defined by a rectangular box confining the translations of the center of mass of the ligand. A set of initial ligand conformations is generated through an exhaustive search of the torsional minima, and the conformers are clustered in a combinatorial fashion. The search begins with a rough positioning and scoring phase that significantly narrows the search space and reduces the number of poses to be further considered to a few hundred. In the following stage, the selected poses are minimized on precomputed OPLS-AA vander Waals and electrostatic grids for the receptor. In the final stage, the 5-10 lowest-energy poses obtained in this fashion are subjected to a Monte Carlo procedure in which nearby torsional minima are examined, and the orientation of peripheral groups of the ligand are refined. The minimized poses are then rescored using the GLIDE Score function, which is a more advanced version of Chem Score 31 with force-field-based components and include additional terms accounting for solvation and repulsive interactions.

Limitation*:* Glide is the slowest program & therefore it is not advisable for usage in docking using large data bases without prior filtering [18].

10. Induced Fit Method

Induced fit is a novel protein-docking (IFD) method that accurately accounts for both ligand and receptor flexibility by iteratively combining rigid receptor docking with protein structure prediction (prime) technique. In order to tacle the full protein/ligand structure prediction problem in a robust &accurate manner, it is essential to allow both the structure of the protein and ligand to reorganize. Application of this methodology to 21 pharmaceutically complexes reported that the average ligand RMSD for docking to a flexible receptor for the 21 pairs is 1.4$\overset{0}{A}$; the RMSD is $\leq$1.8 $\overset{0}{A}$ for 18 of the cases.

Induced fit methodology:

The overall procedure has four steps (outlined in figure 6):

- Initial softened-potential docking into a rigid receptor to generate an ensemble of poses.
- Sampling of the protein for each ligand pose generated in first step.
- Redocking of the ligand into low energy induced-fit structures from the previous step.
- Scoring by accounting for the docking energy (G-score), and receptor strain & solvation terms (prime energy).

Initial ligand sampling: The key challenge in the initial ligand docking step is minimizing the protein-ligand steric clashes that are manifested when docking into the unmodified binding site, while retaining the sufficient structure in the modified binding site to avoid generating a large number of infeasible poses. This is done by scaling the vander waals radii of ligand & receptor atoms by 50% & by temporarily replaces the residues predicted to highly flexible with alanine.

***Receptor sampling*:** Any residue that was replaced with alanine in the first step is restored to their original residue type, and then side-chain prediction &minimization are performed for all 20 ligand/protein complexes. Only residues having at least one atom within $5\overset{0}{A}$ of any of the 20 ligand poses are sampled.

***Ligand resampling*:** In this stage, the ligand is redocked into the induced-fit structures from the previous stage that are within 30kcal/mol of the lowest energy structure.

***Final scoring*:** final scoring is achieved by combining the prime energy and Glide Score in suitable proportions. G-score uses force field based scoring function. In this ligand affinity is primarily driven by hydrophobic effect. In prime energy binding affinity is based on ligand strain & solvation term.

Advantage: Induced fit method is a robust across a wide range of targets, can be applied in an automated fashion, and completes uses an acceptable amount of computation time. The modeled receptor/ligand complexes generated by this methodology can be visually inspected by modelers & medicinal chemists to obtain qualitative ideas about how to modify lead compounds.

Limitation: This methodology only consider the primary changes in receptor side chains but there are certain receptor which exhibit changes in loop conformation upon ligand binding, for example kinases. The induced fit methodology can be applied to such problems by introducing loop prediction into the protocol [26].

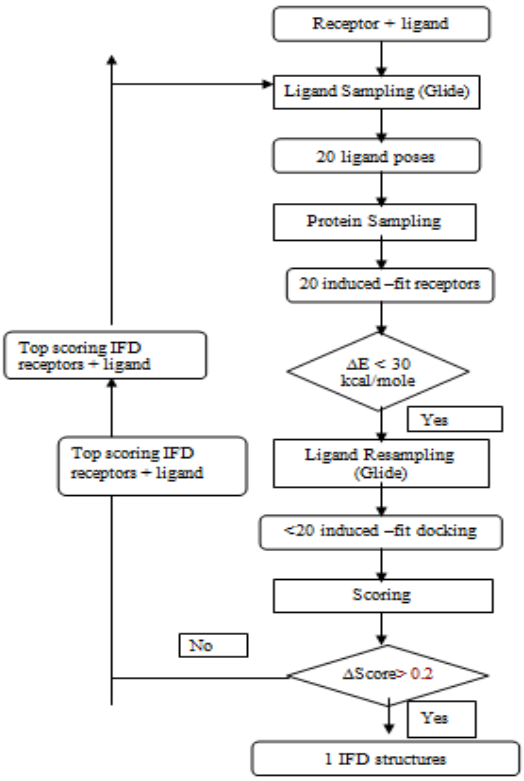

Figure 6: IFD flowchart. ΔE is the energy gap from the lowest energy structure [26].

11. PAS DOCK

Protein alpha shape dock is a new Gaussian based scoring function suitable for virtual library screening using homology modeled protein structures. Here, the scoring function is used in combination with the geometry search

method Tabu search. A description of the protein binding site is generated using Gaussian property fields like in protein alpha shape similarity analysis (PASSA).Gaussian property fields are also used to describe the ligand properties. The overlap between the receptor & ligand hydrophilicity and lipophilicity fields is maximized, while minimizing steric clashes. Gaussian function makes the smoothing of the property fields. This makes the scoring function robust against small structure variations, and suitable for use with homology models since, Gaussian functions give a less detailed representation than force field based models. Two different score functions used in PAS-Dock, a rough estimation of the match between the protein and ligand structures for the geometry search, and a more accurate scoring for the final estimation of the binding affinities. In this way a good prediction of the free energy of binding is obtained.

The performance of PASSA is compared with other docking methods, Auto-Dock &MOE-Dock and PAS-Dock is found to be more computationally efficient than Auto-Dock & MOE-Dock, and gives a better prediction of the free energies of binding.

Limitation: The major limitation is that calculations used in this scoring function are independent of the placement of hydrogen atoms, the fact that hydrogen atom can form hydrogen bonds is not accounted [27].

Consensus scoring

Consensus is the recent trend in the field of scoring function.. Consensus scoring combines balance errors in single scores and improve the probability of identifying 'true' ligands. Consensus score also known as C-score integrates a number of popular scoring function for ranking the affinity of ligands bound to active site of receptor. The strengths of individual scoring functions combine to produce a consensus that is more robust and accurate than any single function for evaluating the ligand-receptor interactions.

C score combines several functions like:

- G-score
- D-score
- PMF-score
- Chem score

Advantage: Consensus lists generated from two or three different scoring functions contain significantly lower number of false positive than any hitlist obtained by single scoring function and conclude that an optimal combination of scoring function significantly enhanced hit rates. False positive occurs because uncertainties in the crystal structure with respect to protein side chain distorts scores in favour of a secondary or trasitional binding mode that in fact is slightly higher in energy.

Disadvantage: Consensus score always present an average value & cannot perform as well as any specific function in a specific instance [13].

12. MOLDOCK

Mol-Dock is a new technique for high-accuracy molecular docking. Mol Dock is based on a new hybrid search algorithm, called guided differential evolution. The guided evolution algorithm combines the differential evolution optimization technique with a cavity prediction algorithm. Differential evolution was introduced by Storn and Price in 1995. The scoring function used is derived from the PLP (piecewise linear potential).The scoring function used by MolDock improves these scoring functions with a new hydrogen bonding term &new charge schemes. The docking scoring function, Escore is defined by the following terms

$$E_{score} = E_{inter} + E_{intra}$$

Guided differential evolution (DE) compared to more evolutionary programming, DE uses a different approach

to select & modify candidate solutions. The main innovative idea in DE is to create offspring from a weighted difference of parent solutions. Parent solutions are randomly selected from the population. Afterward, the offspring replaces the parent, if and only if it is fitter. Mol-Dock automatically identifies potential binding sites using the cavity detection algorithm. The cavities found by the cavity detection algorithm are actively used by search algorithm to focus the search during the docking simulation. The docking accuracy of Mol-Dock has been evaluated by docking flexible ligands to 77 protein targets. Mol-Dock was able to find the correct binding mode of 87% of the complexes. In comparison, the accuracy of Glide and Surflex was 82% and 75% respectively. FlexX obtained 58% and Gold 78% on subsets containing 76 & 55 cases, respectively. The primary reason for success of MolDock is its search algorithm and the re-ranking scoring function [28].

13. FLEXNOVO

FLEXNOVO is a new molecular design program for structure-based searching within large fragment spaces following a sequential growth strategy. The fragment spaces consist of several thousands of chemical fragments and a corresponding set of rules that specify how the fragments can be connected. FlexNovo is based on the FlexX molecular docking software and makes use of incremental construction algorithm. Interaction energies are calculated by using standard scoring functions.. FlexNovo has been used to design potential inhibitors for four targets of pharmaceutical interest (dihydrofolate reductase, cyclin-dependant kinase 2, cyclooxygenase-2, and the estrogen receptor). Calculations using different diversity parameters for each of these targets and generated solution sets containing up to 50 molecules. The compounds obtained show that Flex-Novo is able to generate a diverse set of reasonable molecules with drug-like properties. A FlexNovo calculation consists of a preprocessing phase and a "build-up" phase. The preprocessing phase consists of two different docking calculations for all fragments. The first is performed without constraints and serves to estimate the highest possible score for each fragment according to the scoring function used for a particular receptor (This information is stored on disk and used later in the build-up process.) The second calculation is done by using pharmacophore-type constraints for generating docking solutions for all fragments that are able to fulfill these constraints. The "placements" are used as the starting positions in the build-up process. In the build-up phase, a fixed number of "extension cycles" is carried out. In each extension cycle, the fragments with the best scores from the previous cycle are identified.

Unlike FlexX, which deals with one molecule at a time, Flex-Novo deals with fragment spaces.

Filters used in Flex novo:

***Property filters*:** In these the user can specify property ranges for molecular weight, the number of hydrogen-bond donors & acceptors, rings, non-terminal single bonds, molecular logP, refractivity values.

***Diversity filters*:** In this user can specify the maximum number of common fragments or minimum number of different fragments, for each pair of molecules in the final solution list.

Pose-geometry filters: These include polarity filter, repulsion filter, and saturation filter.

Results for cyclin-dependent kinase:

Cyclin-dependent kinase 2 plays a dominant role (among other kinases) in the modulation of diseases like cancer and is a well-established target in pharmaceutical research .ATP-competitive kinase inhibitors all form one or more hydrogen bonds to the "hinge region" of the active site. Therefore, an essential anchor-pharmacophore constraint

for the generation of the start placements was the formation of a hydrogen bond to the central backbone NH group of Leu83. Most CDK2 inhibitors form additional hydrogen bonds to the nearby carbonyl groups of Glu81 and Leu83.These were also used as a constraint such that at least one interaction with these residues has to be formed. The initial docking calculation of dimethoxyquinazoline derivative yielded solutions for approximately 35% of the fragments that satisfy the specified anchor pharmacophore constraints. The predicted scores were between small positive values and about -30 (kJmol-1).

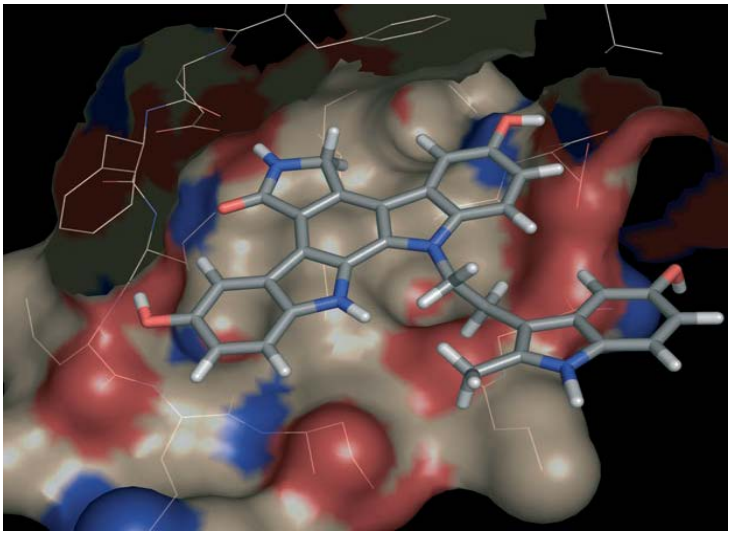

Figure7. Predicted binding mode of a molecule obtained from an earlier version of Flex-Novo in which no filter criteria is used. The active site of CDK2 is shown with its conelly surface. The results obtained for type of targets (shown in figure 8) demonstrate that Flex-Novo is able to generate diverse sets of molecules that are highly complementary to different target proteins. These molecules exhibit drug-like properties & have reasonable predicted binding orientations. Flex-Novo handles very large fragment spaces with up to several thousand fragments [29].

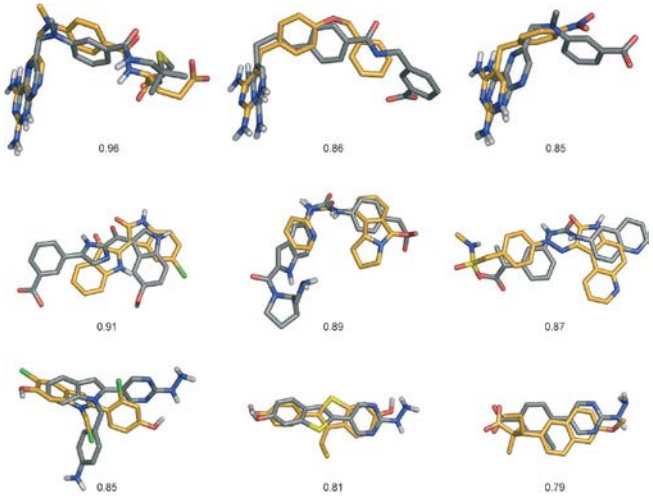

Figure 8 : Superimposed pairs of Flex Novo solution list molecules (gray) for targets DHFR, CDK2,and ER (from top to bottom), and known inhibitors (orange) .

Limitations: Flex-Novo is not able to close rings during the structure-generation process .second halogen atoms are not included in the chemical model. Another principle limitation is that pharmacophore-type constraints can only be used for the initial fragment-placement [29].

14. Cross-Docking Methodology

Non-nucleoside reverse transcriptase inhibitors (NNRTIs) have, in addition to the nucleoside reverse transcriptase inhibitors (NRTI) & protease inhibitors (PIs), a definitive role in the treatment of HIV-1 infections. The NNRTI interact with a specific site of HIV-1 RT (non-nucleoside binding site NNBS) that is close to, but distinct from the NRTI binding site. Mutations of some amino acids cause a variation of the NNBS pocket properties, thus decreasing affinities of most the inhibitors.

Application of several Auto dock program has been widely used with success in reproducing the bound conformation of different ligand/protein systems. A striking feature of the RT is its considerable conformational flexibility (responsible for several of its catalytic actions). Inclusion of such structural variability in a docking study becomes of fundamental importance.

The ideal situation would be a program able to dock a ligand into a protein structure from a different complex easily & with reasonable accuracy. This has been referred to as 'cross-docking'. Auto dock as well as many docking program is not able to consider protein flexibility. To overcome this limitation an extensive cross-docking study is used in order to check if different ligands can still bind different NNBSat a low energy level. Delarvidine which is a peculiar NNRTI bearing unique feature was the worst cross-docked compound. The cross-docking approach in which the different NNRTI binding pockets allows each ligand to adopt different poses .Cross-docking experiments were also conducted on the mutated RT set. Observing NNRTI-bound conformations in the wild-type or mutated RT forms, it seems that for ligands also able to inhibit the RTmutants a unique binding mode exist for both the enzymes. Thus, parallel cross-docking experiments on both wild-type and mutated would be a useful tool for structure based drug design studies to design new inhibitors able to tightly bind to both enzyme. Cross-docking experiments conducted by docking 41 NNRTIs into 41 different RTs proved the autodock program to be a useful tool for structure-based drug design in developing new anti-RT agents against wild-type and mutated forms of the enzyme. Cross-docking is recently applied in a unique way the binding mode of the L-737126 lead compound & design new and potent anti-HIV agents. Application of cross docking on using mutated RTs to design anti-HIV agents active resistant strain is underway & will be reported soon[30].

MM-GBSA Scoring

MM-GBSA is a new scoring function has recently become of interest in drug discovery for predicting relative binding free energies of drug discovery project. In this approach, the binding free energy ΔG_{bind} is estimated as

$$\Delta G_{bind} = \Delta E_{MM} + \Delta G_{solv} + \Delta G_{SA}$$

Where ΔE_{MM} is the difference in energy between the complex structure & the sum of energies of the ligand and unliganded protein, ΔG_{solv} is the difference in the GBSA salvation energy of the complex & the sum of the solvation energies for the ligand and unliganded protein, and ΔG_{SA} is the difference in the surface area energy for the complex & sum of surface area energies for the ligand and uncomplexed protein. Corrections for entropic changes are not applied. The relative potencies of members of a series of kinase inhibitors are successfully predicted by using molecular docking program Glide and MM-PBSA as a post docking scoring protocol[31].

15. Applications of Molecular Docking

The most important application of docking is virtual screening. In virtual screening the most interesting and most promising molecules are selected from an existing database for further research. Docking provides a reliable and fast filter in HT virtual screening. Molecular docking is a key to rational drug design: the results of docking can be used to find drugs for specific target proteins and thus to design new drugs.

Here, I have mentioned two examples in which molecular docking was used to investigate the key interaction between ligand and enzyme and successfully developed potent lead molecules..

A series of substituted acyl(thio)urea and 2H-1,2,4-thiadiazolo [2,3-a] pyrimidine derivatives were prepared. Molecular docking has been performed to evaluate of a new series of substituted acyl (thio) urea and thiadiazolo [2, 3-a] pyrimidine derivatives as potent inhibitors of influenza virus neuraminidase. Influenza virus commonly known as flu is the contagious etiologic agent that causes an acute respiratory infection; hence it has always been a major threat to human health worldwide and cause for economic costs.

The only neuraminidase (NA) inhibitors which received FDA approval are zanamivir and oseltamivir. In general, the NA inhibitors known as NAI do not inhibit virus replication but do prevent the release and spread of the virus from infected cells, and effectively retard its propagation.

Molecular docking procedure

FlexX 1.11.1 within SYBYL package was employed to explore the interaction between the ligand and enzyme. The crystal structure of influenza virus neuraminidase complexed with zanamivir was retrieved from PDB with corresponding entry code 1a4g. The crystal structure of neuraminidase-zanamivir complex (Fig.9) demonstrated the pattern of protein–ligand interactions, which consist of strong charge-charge- and charge-partial charge-based hydrogen bonds. The protein was prepared by removing heteroatom's and water molecules and adding all hydrogen atoms. The active site of 1a4 g was defined as residues with at least one atom within a radius of 9 $\overset{0}{A}$ from any atom of zanamivir. Then all compounds were sketched using sybyl with all hydrogen atoms. Furthermore, their conformers with low energy were ensured by RANDOM searches available in SYBYL. Then the compounds were docked to neuraminidase from influenza virus by FlexX facilities. FlexX scoring function was employed to evaluate the docking pose of the compounds.

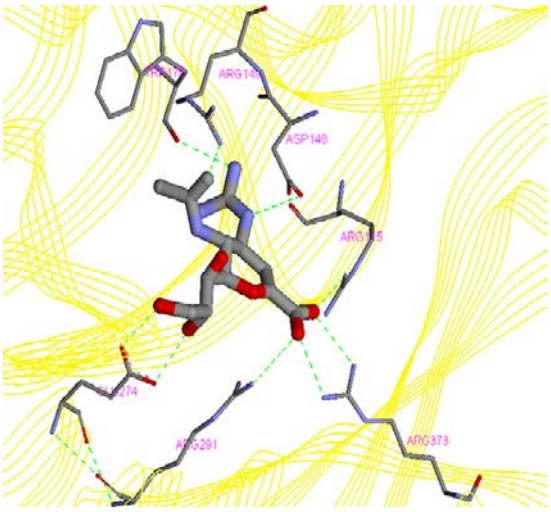

Figure. 9(a)

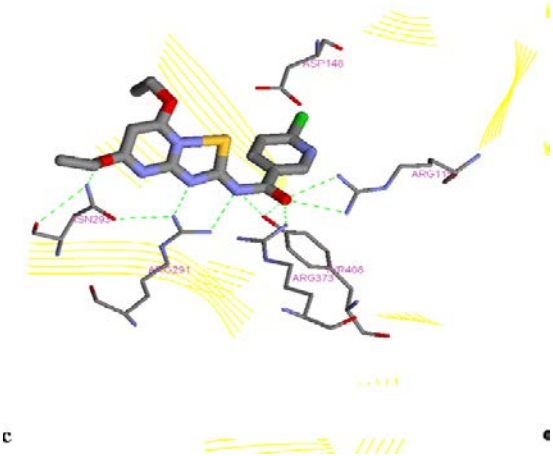

Figure. 9(b

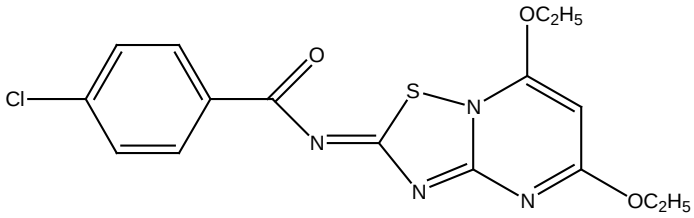

Figure. 10 .Advanced potential development canditate

Result: FlexX Score was found to be lowest for analogue 1. Analogue 1 (Fig.10) inhibited the influenza virus with an IC_{50} of 0.09 µM, and this novel inhibitor was investigated as candidate compounds with the most potential for future development.

Analogue 1 shared different structural properties with those of zanamivir, they seemed to bear a similar binding

mode (figure 7). The oxyethyl groups of 2 occupy a small pocket at the entrance of the binding cavity, which may prevent the substrate of NA from entering into the active site.

This research leads to a better understanding of SAR of influenza virus inhibitors and thereby provides some insight into the rational design of anti-flu agents and the discovery of new effective drug [32-34].

I. Novel antagonist for a nuclear hormone receptor has been identified by the program ICM, and the ICM with DOCK has been used to find inhibitors for the RNA transactivating response element (TAR) of HIV-1. The virtual screening protocol started with 153,000 compounds for the Available Chemicals Directory (ACD). In the HIV-1 TAR study, the ACD library was first rigidly docked into the binding site using DOCK program along with a simple contact scoring scheme. Then, 20% of best scoring compounds were subjected to flexible docking with ICM and empirical scoring function providing a selection of approximately 5000 compounds. This was followed by two additional steps involving longer sampling of conformational space to retrieve 350 most promising candidates. Of these, a very small fraction was tested experimentally & two compounds were successfully found to significantly reduce the binding of the Tat protein to HIV-1 TAR [28-30].

Docking techniques are currently applied to aid in structure-based absorption, distribution, metabolism and excretion evaluation. Cytochrome P450 isoforms are major drug-metabolizing enzymes and have become focal points in the study of rapid metabolism and drug-drug interactions. Several groups have therefore developed structure-based approaches for the prediction of compounds that would be metabolized by or inhibit P450 isoforms. Various homology models of human P450 isoform have been generated for these purposes as templates for docking to predict drug metabolism. These structural insights should help to further refine docking studies on human P450s and increase their predictive value [32-34].

Limitations of Docking

Docking does not reflect the actual physical process of binding and in some cases even prevents the correct identification of potential drug candidates. It can not differentiate between agonist & antagonist. Some of approaches have high computational cost and require a deep understanding of biological system makes automation difficult [33]. Docking is computationally difficult because there are many ways of putting two molecules together. Both molecules are flexible and alter each other when they interact. There are 100 to 1000 degree of freedom. The number of possibilities grows exponentially with the size of components. Combining all patches of the surface of one protein molecule with all patches of a second molecule takes an order of 10^7 trials. The computational problem is even more profound when we consider protein flexibility and the increasing demand to screen large databases (of protein structures & of potential drugs) [30-34].

16. Conclusion

Molecular docking is now established as an important approach in drug discovery. Two major bottleneck of molecular docking are availability of an efficient docking algorithm & availability of a selective and efficient scoring function. Many docking methods are available but all of them certain limitations. Comparison suggests that the best algorithm for docking is probably a hybrid of various types of algorithm encompassing novel search and scoring strategies. The issue of flexibility & induced-fit motions of the protein will gain in importance over the coming years in the design and discovery of novel lead

candidates[1]. The identification of an overall reliable and robust scoring function seems to be one of the main challenges to be addressed in the near future. However, the combination of scoring functions in a consensus score presents a well established scoring function. We conclude by summarizing our perspective on major challenges in the further development of docking procedures & scoring functions.

- The fact that protein-ligand interaction occurs in aqueous solution is generally appreciated but not yet adequately accounted for in molecular docking procedures. The placement of water molecules and the fast prediction of protonation states in binding pockets will provide a more satisfactory solution.
- It is necessary to consider the sufficient degree of protein flexibility during docking procedures.
- Polar interactions should be treated adequately.
- Fast scoring functions cover only part of the whole receptor-ligand binding process. A more detailed picture could be obtained by taking into account properties of the unbound ligand, that is, solvation effect and energetic differences between low-energy solution conformations & the bound conformation.

17. References

1. Jurs, P. 'In: Handbook of Chem-informatics' Gasteiger, J. Ed.; Wiley-VCH: Weinheim, 2003, Vol. 3, 1314-1335.
2. Ivanciuc, O. 'In: Encyclopedia of Complexity and Systems Science' Meyers, R.A. Ed., Springer-Verlag: Berlin, 2009, 2113-2139.
3. Bagchi, M. C.; Maiti, B. C.; Mills, D.; Basak S. C. 'Usefulness of graphical invariants in quantitative structure–activity correlations of tuberculostatic drugs of the isonicotinic acid hydrazide type'. J. Mol. Model., 2004, 10, 102–111.
4. Mon, J.; Flury, M.; Harsh, J. B. 'A quantitative structure–activity relationships (QSAR) analysis of triarylmethane dye tracers'. J. Hydrology, 2006, 316, 84–97.
5. Sabljic´, 'A. Quantitative modeling of soil sorption for xenobiotic chemicals'. Environ. Health Perspect, 1989, 83, 179–190.
6. http://www.mediapeta.com/petauk/PDF/QSAR,QSARs and REACH. Expertise for REACH.pdf-accessed on August 18, 2012.
7. Hansch, C. 'A quantitative approach to biochemical structure activity relationships'. Acc. Chem. Res., 1969, 2, 232-239.
8. Ivanciuc, O. 'Weka Machine Learning for Predicting the Phospholipidosis Inducing Potential'. Curr. Topics Med. Chem., 2008, 8, 1691-1709.
9. Modi, S. 'Positioning ADMET in silico tools in drug discovery'. Drug Discov. Today, 2004, 9, 14–15.
10. Ivanciuc, O. 'Machine Learning Quantitative Structure–Activity Relationships (QSAR) for Peptides Binding to the Human Amphiphysin–1 SH3 Domain', Current Proteomics, 2009, 6, 289-302.
11. Mjolsness, E.; DeCoste, D. 'Machine learning for science: State of the art and future prospects'. Science, 2001, 293, 2051-2055.
12. Duch, W.; Swaminathan, K.; Meller, 'J. Artificial Intelligence Approaches for Rational Drug Design and Discovery'. Cur. Pharm. Des., 2007, 13, 1-12.
13. Kitchen, D. B.; Decornez, H.; Furr, J. R.;Bajorath, J. Docking and Scoring In Virtual Screening for Drug Discovery:Methods and Applications. Nat. rev. Drug Discov. 2004, 3, 935-946.
14. Kitchen, D. B.; Decornez, H.; Furr, J. R.;Bajorath, J. Docking and Scoring In Virtual Screening for Drug Discovery:Methods and Applications. Nat. rev. Drug Discov. 2004, 3, 935-946.

15. Sotriffer, C.; Klrbe, G.; Stahl, M.;Bohm, H.-J., Docking and Scoring Functions/Virtual Screening. In Burger's Medicinal Chemistry and Drug Discov, 6th ed.; Abraham, D. J., Ed. 2003; Vol. 1, 281-323.
16. Rognan, D.;Folkers, G., Protein-based virtual screening. 2nd ed.; WILEY-VCH GmbH & Co.KGaA: 2003; Vol. 5, 145-167.
17. Krovat, E. M.; Steindl, T.;Langer, T. Recent Advances in Docking and Scoring. J. Curr. Comput.-Aided Drug. Des. 2005, 1, 93-102
18. Halperin, I.; Ma, B.; Wolfson, H.;Nussinov, R. Principles of Docking: An Overview of Search Algorithms and a Guide to Scoring Functions. Proteins. 2002, 47, 409-443.
19. Chen, H.; Lyne, P. D.; Giordanetto, F.;Lovell, T. On Evaluating Molecular-Docking Methods for Pose Prediction and Enrichment Factors. J. Chem. Inf. Model. 2006, 46, 401-415
20. Wang, R.; Lu, Y.;Wang, S. Comparative Evaluation of 11 Scoring Functions for Molecular Docking. J. Med. Chem. 2003, 46, 2287-2303.
21. Muegge, I. PMF Scoring Revisited. J. Med. Chem. 2006, 49, 5895-5902.
22. Taylor, R. D.; Jewsbury, P. J.;Essex, J. W. A review of protein-small molecule docking methods J. Comput.-Aided Mol. Des. 2002, 16, 151-166.
23. Liu, M.;Wang, S. MCDOCK: A Monte Carlo simulation approch to the molecular docking problem. J.Comput.-Aided Mol.Des. 1999, 13, 435-451.
24. David, L.; Luo, R.;Gilson, M. k. Ligand-recepter docking with the Mining Minima Optimizer. J. Comput.-Aided Mol. Des. 2001, 15, 157-171.
25. Ewing, T. J. A.; Makino, S.; Skillman, A. G.;Kuntz, I. D. DOCK 4.0: Search strategies for automated molecular docking of flexible molecular databases. J. Comput.-Aided Mol. Des. 2001, 15, 411-428.
26. Jain, A. N. Surflex: Fully Automatic Flexible Molecular Docking Using a molecular Similarity-Based Search Engine. J. Med. Chem. 2003, 46, 499-511.
27. Sherman, W.; Day, T.; Jacobson, M. P.; Friesner, R. A.;Farid, R. Novel Procedure for Modeling Ligand / Receptor Induced Fit Effects. J. Med. Chem. 2006, 49, 534-553.
28. TONDEL, K.; Anderssen, E.;Drablos, F. Protein Alpha Shape (PAS) Dock: A new gaussian-based score function suitable for docking in homology modelled protein. J. Comput.-Aided Mol. Des. 2006, 20, 131-144.
29. Thomsen, R.;Christensen, M. H. MolDock: A New Technique for High-Accuracy Molecular Docking. J. Med. Chem. 2006, 49, 3315-3321.
30. Degen, J.;Rarey, M. Flex Novo: Structure-Based Searching in Large Fragment Spaces. J. Chem. Med. Chem. 2006, 1, 854-868.
31. Ragno, R.; Frasca, S.; Manetti, F.; Brizzi, A.;Massa, S. HIV-Reverse Transcriptase Inhibition: Inclusion of Ligand-Induced Fit by Cross-Docking Studies. J. Med. Chem. 2005, 48, 200-212.
32. Lyne, P. D.; Lamb, M. L.;Saeh, J. C. Accurate Prediction of the Relative Potencies of Members of a Series of Kinase Inhibitors Using Molecular Docking and MM-GBSA Scoring. J. Med. Chem. 2006, 49, 4805-4808.
33. Sun, C.; Zhang, X.; Huang, H.;Zhou, P. Synthesis and evaluation of a new series of substituted acyl (thio) urea and thiadiazolo [2,3-a] pyrimidine derivatives as potent inhibitiors of influenza virus neuraminidase. Bioorg. Med. Chem. 2006, 14, 8574-8581.
34. http/www. phillips-lab.biochemwisc.edu/pdfs/108-docking.pdf