

Review Paper

AN OVERVIEW ON MOLECULAR DOCKING

Gaba Monika^{1*}, Gaba Punam¹, Singh Sarbjot², and Gupta G. D¹

¹Department of Pharmaceutical Chemistry, ASBASJSM College of Pharmacy, BELA (Ropar)-140111, Punjab, India

²Biology Research, Drug Discovery Research, Panacea Biotec Pvt. Ltd., Mohali, Punjab, India

ABSTRACT

Molecular Docking is the computational modeling of the structure of complexes formed by two or more interacting molecules. The goal of molecular docking is the prediction of the three dimensional structures of interest. Docking itself only produces plausible candidate structures. These candidates are ranked using methods such as scoring functions to identify structures that are most likely to occur in nature. The state of the art of various computational aspects of molecular docking based virtual screening of database of small molecules is presented. This review encompasses molecular docking approaches, different search algorithms and the scoring functions used in docking methods and their applications to protein and nucleic acid drug targets. Limitations of current technologies as well as future prospects are also presented.

Keywords: *Molecular docking, scoring functions, virtual screening, docking algorithm.*

INTRODUCTION

In the field of molecular modeling, molecular docking is a method which predicts the preferred orientation of one molecule to a second when bound to each other to form a stable complex^[1]. Knowledge of the preferred orientation is used to predict the strength of association or binding affinity between two molecules using scoring functions. The associations between biologically relevant molecules such as proteins, nucleic acids, carbohydrates, and lipids play central role in signal transduction. Furthermore, the relative orientation of the two interacting partners may affect the type of signal produced (e.g. agonism/ antagonism). Therefore docking is useful for predicting both the strength and type of signal produced. Docking is

frequently used to predict the binding orientation of drug candidates to their protein targets in order to predict the affinity and activity of the small molecule. Hence docking plays an important role in the rational design of drugs^[2]. The aim of molecular docking is to achieve an optimized conformation for both the protein and ligand and relative orientation between protein and ligand so that the free energy of the overall system is minimized. Molecular recognition plays a key role in promoting fundamental biomolecular events such as enzyme-substrate, drug-protein and drug-nucleic acid interactions. Detailed understanding of the general principles that govern the nature of the interactions (van der Waals, hydrogen bonding, electrostatic) between the ligands and their protein or nucleic acid targets may provide a framework for designing the desired potency and specificity of potential drug leads for a given therapeutic target. Practical application of this knowledge requires structural data for the target of interest and a procedure for evaluating candidate ligands. A variety of

***Corresponding Author: Monika Gaba**
Department of Pharmaceutical Chemistry
ASBASJSM College of Pharmacy, BELA -140 111
(Ropar) India
+91-1881-263108, +919872390321
E mail: gaba_monika12@yahoo.co.in
Fax No. +911881263108

computational docking methods are available [3-7]. These provide the ranking of potential ligands with respect to their ability to interact with given target. Docking of a small molecule to a biological target involves efficient sampling of possible poses of the former in the specified binding pocket of the latter in order to identify the optimal binding geometry, measured by a user-defined fitness or score function. X-ray crystallography and NMR spectroscopy continue to be the primary source of 3D structural data for protein and nucleic acid targets. When proteins of unknown structure have high sequence homology to known structures, homology modeling can provide a viable alternative by generating a suitable starting point for 'in silico' discovery of high affinity ligands. Databases of drug like molecules such as MDDR [8] or CMC [9], as well as other small molecule databases including ACD [10], CSD [11] and NCI [12] are available. During computational docking, a pose is generated, scored and compared to the previous pose(s). The current pose is then accepted or rejected on the basis of the score for that pose. A new pose is then generated, and the search process iterates to an endpoint. Thus, searching and scoring can be tightly coupled in docking. Reliable rank ordering of the ligands based on their docked scores such that the scores correlate with experimental binding affinities appears to be even more challenging than searching the conformation and orientation space [13-15]. A recent trend has been to employ consensus scoring (apply a number of score functions to the same docked pose identified by docking) to eliminate false positives [15, 16]. 'In silico' approaches need to be robust and fast in order to have a major impact on lead identification. The standard test that has emerged for docking-based virtual screening protocols evaluates the ability of the docking method to prioritize known active molecules from a database comprised largely of molecules known to be inactive. Over the

last few years a vast amount of effort has been directed for developing efficient docking methods and scoring functions as tools for the identification of lead compounds. Considerable progress has been made in the computational prediction of ligand-target binding modes. A number of review articles in this emerging area of research have been recently published [17-21]. This review highlights current computational docking technology, approaches, successes, failures, limitations, challenges and future prospects.

MOLECULAR DOCKING APPROACHES

Two approaches are particularly popular within the molecular docking community. One approach uses matching technique that describes the protein and the ligand as complementary surfaces [22, 23]. The second approach simulates the docking process in which the ligand-protein pair wise interaction energies are calculated [24]. Both approaches are outlined below.

Shape complementarity

Geometric matching/ shape complementarity methods describe the protein and ligand as a set of features that make them dockable (Fig. 1) [25]. These features may include molecular surface/ complementary surface descriptors. In this case, the receptor's molecular surface is described in terms of its solvent-accessible surface area and the ligand's molecular surface is described in terms of its matching surface description. The complementary between the two surfaces amounts to the shape matching description that may help in finding the complementary pose of docking the target and the ligand molecules. Another approach is to describe the hydrophobic features of the protein using turns in the main-chain atoms. Yet another approach is to use a Fourier shape descriptor technique [26-28]. Whereas the shape complementarity based

approaches are typically fast and robust, they cannot usually model the movements or dynamic changes in the ligand/ protein conformations accurately, although recent developments allow these methods to investigate ligand flexibility. Shape complementary methods can quickly scan through several thousand ligands in a matter of seconds and actually figure out whether they can bind at the protein's active site, and are usually scalable to even protein-protein interactions. They are also much more amenable to pharmacophore based approaches, since they use geometric descriptions of the ligands to find optimal binding.

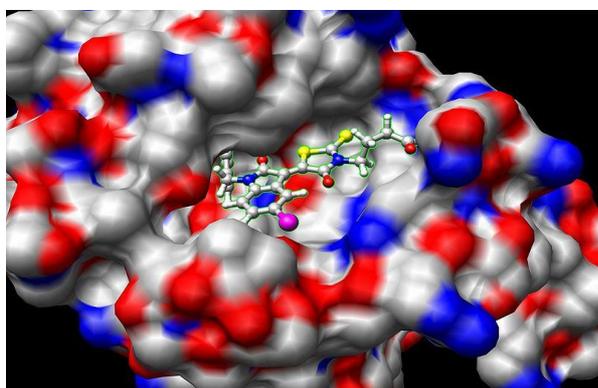


Figure: Molecular docking of ligand to a protein receptor to produce a complex.

Simulation

The simulation of docking process as such is much more complicated process. In this approach, the protein and the ligand are separated by some physical distance, and the ligand finds its position into the protein's active site after a certain number of "moves" in its conformational space. The moves incorporate rigid body transformations such as translations and rotations, as well as internal changes to the ligand's structure including torsion angle rotations. Each of these moves in the conformation space of the ligand induces a total energetic cost of the system, and hence after every move the total energy of the system is calculated. The advantage of this method is that it is more amenable to incorporate ligand flexibility into its

modeling whereas shape complementary techniques have to use some ingenious methods to incorporate flexibility in ligands. Another advantage is that this process is physically closer to what happens in reality, when the protein and ligand approach each other after molecular recognition. The disadvantage of this technique is that it takes longer time to evaluate the optimal pose of binding since they have to explore a rather large energy landscape. However grid-based techniques as well as fast optimization methods have significantly ameliorated these problems.

TYPES OF DOCKING

- (a) Rigid body docking, where both the receptor and small molecule are treated as rigid.
- (b) Flexible ligand docking, where the receptor is held rigid, but the ligand is treated as flexible; and
- (c) Flexible docking, where both receptor and ligand flexibility is considered.

The most commonly docking algorithms use the rigid receptor/flexible ligand model. The principle docking methods that are used extensively employ search algorithms based on Monte Carlo, genetic algorithm, fragment-based and molecular dynamics. Some programs that are well-suited for high throughput docking of a large database of molecules include: DOCK^[3, 4], FlexX^[5], GOLD^[6], and ICM^[7].

BASIC REQUIREMENTS FOR MOLECULAR DOCKING

The setup for a ligand docking approach requires components a target protein structure, 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. Most docking algorithms assume protein to be rigid; the ligand is mostly regarded as

flexible. Beside the conformational degrees of freedom the binding pose in protein's binding pocket must be taken into consideration. Docking can be performed by placing rigid molecules or fragments into protein's active site using different approaches like the clique-search, geometric hashing, or pose clustering.

Structural data

Ligand Representation

Typically, the structure most likely to be dominant at neutral pH is generated. The structures can be further adjusted by adding or removing hydrogens provided approximate pKa values. It is important to make sure that 'accurate' atom typing occurs. The wrong definition of donor and acceptor properties of heteroatoms may lead to serious docking errors. For example, Watson and Crick originally assigned the wrong tautomeric formulae (enol forms) for nucleic acid bases and thus could not build a helical model with purine-pyrimidine hydrogen bonded base pairs. However, once the correct tautomeric structures (keto forms) of the bases were assigned, all the key features of 3D structure of double helical DNA were readily accounted [29]. In cases where stereochemistry of synthesized compound is unknown, it is beneficial to generate all possible diastereoisomers and dock them individually to the receptor. Commercial software programs for the enumeration of all possible diastereoisomers of a given compound include: Stergen [30], Stereoplex [31] and PipelinePilot [32].

Receptor Representation

The quality of receptor structure employed plays central role in determining the success of docking calculations [33-35]. In general, the higher the resolution of the employed crystal structure better will be the observed docking results. Schapira *et al.* using ICM for docking showed that they could reproduce the known binding modes of ligands to

within 1 Å of the bound conformations, in cases where the resolution of the employed co-crystal structures were better than 2.0 Å [35]. A recent review for accuracy, limitations and pitfalls of the structure refinement protocols of protein ligand complexes in general provided a critical assessment of the available structures [36]. The importance of the pH dependence of ligand binding modes was highlighted. Uncertainties in locating the ligands ('mistaken identity') in the co-crystal structures as well as the subjective nature of deriving good quality protein models were emphasized in the context of published structures. The reliability of the ligand structures found in co-complexes has been questioned also. Even at high resolution, the difficulties in defining ligand atomic positions unambiguously can be attributed to the disparity between the high-quality dictionaries of bond lengths, bond angles and torsions available for proteins and nucleic acids structure refinement and those available for small organic molecules [37, 38]. Regardless of the possible ambiguities, success has been reported for numerous high throughput docking studies using X-ray receptor structures. Recent examples of this type of study include: kinesin [39], thymidylate synthase [40], phosphoribosyl transferase [41], farnesyltransferase [42], HIV protease [43], and beta-lactamase [44].

RECEPTOR FLEXIBILITY

It is well known that macromolecules often undergo conformational change or induced fit on ligand binding in order to maximize energetically favorable interactions with the ligand or solvent [45, 46]. The driving force behind most induced fit mechanisms is hydrophobic interactions or hydrophobic collapse of the receptor around the bound ligand [47]. There are varying degrees of receptor flexibility. Conformational flexibility does not necessarily need to involve domain, tertiary, or secondary structure

motions but may consist of side-chain adjustments. It has been noted that the successes and failures of docking simulations have been explained on the basis of thermodynamic properties determined from equilibrium simulations and the shape of the underlying binding energy landscape, funnel-like for rigid docking and rugged for flexible docking [48]. The most docking approaches are the rigid receptor hypothesis [49, 50]. The major drawback of the rigid receptor docking approach is that it may lead to incorrect ligand binding modes or poor docking scores, thus overlooking prospective drug leads. This, coupled with the fact that conformational changes within the receptor may have important implications with respect to ligand selectivity, illustrates the importance of incorporating receptor flexibility in computational drug design. The degree of flexibility one could incorporate in a given experiment is directly proportional to computational complexity and cost. A few of the more elegant methods simulating receptor flexibility are described below.

Soft Docking

Soft docking algorithms attempt to allow for flexibility of the receptor and ligand structures by using a relaxed representation of the molecular surface. An efficient scheme to handle receptor flexibility is to use additional energy (i.e. van der Waals) in the empirical scoring function. The soft docking concept, originally proposed by Jiang and Kim, describes the molecular surface and volume as a “cube representation” [51]. This cube representation implies conformational changes by way of size/shape complementarity, close packing and, most importantly, liberal steric overlap. In recent years the soft docking concept has evolved primarily toward use in protein-protein docking [52-56] and protein-receptor modeling combined with experimental NMR data [57-59].

Side-Chain Flexibility

Side-chain flexibility is another way to provide receptor flexibility. One method, originally proposed by Leach [60], uses pre-generated side-chain rotamer libraries that subsequently are subjected to optimization during a ligand docking procedure via the dead-end elimination algorithm. The optimized ligand/side-chain orientations are then scored in order to rank lowest energy combination of side-chain and ligand conformers [61]. Gilson has recently enhanced the Mining Minima optimizer [62, 63] by incorporating side-chain flexibility [64]. The algorithm allows conformations of user-selected side-chains in the active site to be optimized along with the conformation and position of the ligand. This is accomplished by computing energies associated with the selected side-chains as if they belonged to the ligand. Another docking procedure has automated the ‘userdefined’ selection of active-site, flexible residues by way of SOFTSPOTS algorithm [65]. SOFTSPOTS makes use of a knowledge-based function that identifies active site residues most likely to undergo conformational change upon ligand binding. Usually only a few hydrophobic residues are selected, depending on location relative to ligand position and secondary structure assignment. The use of PLASTIC algorithm [65] generates the side-chain rotamers, or minimal conformations, prior to docking calculations. Molecular dynamics and Monte Carlo simulations have also been used to explicitly model side-chain movement [66].

Experimental and Theoretical Ensemble Docking

Ensemble docking has gained considerable attention as a method of incorporating protein flexibility in computational drug design. In most cases full receptor ensembles are generated by molecular dynamics, Monte Carlo simulation or homology modeling methods. The ensembles can be generated

experimentally by NMR solution structure determination or in few cases, multiple x-ray crystal structures. Comparisons have shown that there is significant overlap of dynamic information content between theoretically derived molecular dynamics ensembles and experimentally derived NMR ensembles^[67].

Hybrid Techniques

The hybrid techniques for incorporating into docking experiments are beginning to appear in the literature. A combined method of soft docking and side-chain optimization has recently been reported for protein-protein docking^[68]. This procedure, DISCO, combines the use of soft receptor potentials as defined with the ICM docking engine, and interface residue side-chain optimization. Another hybrid methodology refines the "Relaxed Complex" approach to take advantage of MM/PBSA scoring^[69]. FDS (Flexible ligand and receptor Docking with a continuum Solvent model) is another hybrid technique^[70]. This method initially docks the ligand into the protein active site based on satisfying possible sets of hydrogen bonds using graph theory. The docked compounds are then filtered by cluster analysis to reduce the number of structures submitted to a modified Monte Carlo algorithm that uses a Generalized Born Surface Area (GBSA) solvent model^[71].

VIRTUAL SCREENING TECHNIQUES AND DOCKING

Virtual screening is a widely accepted method in lead discovery because it is advantageous in the elimination of undesired molecules from compound libraries and the reduction of cost and time in drug discovery projects. In structure-based design ligands are modelled regarding the demand of the protein binding pocket. Docking may help in this case to investigate the active site in detail and detect

uncovered binding pockets or interaction points. This approach is carried out by several de novo design tools, e.g. ligand construction and docking in GROWMOL^[72]. These applications provide new scaffolds and can result in new synthesis strategies for the medicinal chemist. Docking of virtual combinatorial libraries may yield innovative ligands as well^[73]. The aspect of privileged motive design^[74] can be implemented by using the innovative software tool *ilib diverse* to generate focused virtual libraries for the desired target^[75, 76]. This program builds libraries of drug-like organic molecules for rational lead structure discovery. Compounds are generated by combining user-defined fragments according to state-of-the-art chemical knowledge. The technique of virtual screening is used to prioritise molecules for biological assays. It is cost and time efficient and has contributed important advances to lead discovery programs in many pharmaceutical companies. Often virtual screening techniques are used in combination with HT screening lead discovery tools^[77]. The docking of huge molecule database against a specific target may yield new candidates for further lead development. Furthermore, docking can help to correlate the experimentally determined biological activities, ligand poses, and predicted binding affinities by the docking program, to evaluate the scoring functions and to identify a good score of the target protein. A frequently used method is the redocking of a complexed ligand in order to verify the validity of the docking and scoring algorithms.

PRE-DOCKING COMPOUND FILTERING

Prior to carrying out docking calculations, it is beneficial to pre-select the database of compounds to be docked by applying hierarchical filters to produce 'focused' database. Such filters often drastically reduce the number of compounds that need to be evaluated in the more demanding

docking calculation. An illustration of such an approach has been recently reported in a lead discovery study involving human carbonic anhydrase II ^[78]. Starting from a set of 90, 000 compounds, the successive application of 2-D substructure queries to identify known metal-binding groups, 3-D pharmacophore based queries, and flexible superposition reduced the database to 100 compounds. Docking calculations of the selected 100 compounds with FlexX identified four potent inhibitors with activities in the nanomolar range. Subsequent crystallographic studies confirmed the predicted docking poses of two of these hits ^[78]. In general case, a suitable set of pharmacophores can be derived by performing a binding site analysis to identify regions of favorable protein-ligand interactions. An excellent review has been published describing the application of pharmacophore based modeling methods in discovering new leads in the absence of structural data ^[79]. A hybrid approach in which initial pharmacophore-based filtering was followed by subsequent docking of a small subset of compounds yielded novel inhibitors of alanine racemase ^[80] and dihydrofolate reductase ^[81].

MECHANICS OF DOCKING

To perform a docking screen, the first requirement is a structure of the protein of interest. Usually the structure has been determined using a biophysical technique such as x-ray crystallography, or less often, NMR spectroscopy. This protein structure and a database of ligands serve as inputs to a docking program. The success of a docking program depends on two components such as search algorithm and scoring function.

Searching Conformational Space

The search space consists of all possible orientations and conformations of the protein paired

with ligand. With present computing resources, it is impossible to exhaustively explore the search space this would enumerating all possible distortions of each molecule and all possible rotational and translational orientations of the ligand relative to the protein at a given level of granularity. Most docking programs in use account for flexible ligand, and several are attempting to model a flexible protein receptor. Each "snapshot" of the pair is referred as a pose.

Scoring Functions

The scoring function takes a pose as input and returns a number indicating the likelihood that the pose represents a favorable binding interaction. Most scoring functions are physics based molecular mechanics force fields that estimate the energy of the pose; a low (negative) energy indicates stable system and thus a likely binding interaction. An alternative approach is to derive a statistical potential for interactions from a large database of protein-ligand complexes, such as the Protein Data Bank, and evaluate the fit of the pose according to this inferred potential. There are a large number of structures from X-ray crystallography for complexes between proteins and high affinity ligands, but comparatively fewer for low affinity ligands as the later complexes tend to be less stable and therefore more difficult to crystallize. Scoring functions trained with this data can dock high affinity ligands correctly, but they will also give plausible docked conformations for ligands that do not bind. This gives a large number of false positive hits, i.e. ligands predicted to bind to the proteins that actually don't when placed together in a test tube. One way to reduce the number of false positives is to recalculate the energy of the top scoring poses using more accurate but computationally more intensive techniques such as Generalized Born or Poisson-Boltzmann methods. Scoring functions can

be classified into three distinct categories: knowledge-based, empirical and forcefield-based. Knowledge-based scoring functions rely on statistical means to extract rules on preferred, and nonpreferred, atom pair interactions from experimentally determined protein-ligand complexes. The rules are interpreted as pair-potentials that are subsequently used to score ligand binding poses. The PMF score^[82] is a well known knowledge-based scoring function. Empirical scoring functions sum enthalpic and entropic interactions with the relative weights of the terms based on a training set of protein-ligand complexes. The weights are assigned by regression methods that are used to fit the experimentally determined affinities. The interaction terms often include Van der Waals, electrostatic interactions and hydrogen bonds. Examples of empirical scoring functions include PLP^[83], ChemScore^[84] and the FlexX^[5] scoring function. Force field scoring functions predict binding free energy of a protein-ligand complex by adding up individual contributions from different types of interactions. Examples of force field scoring functions in docking programs include DOCK^[85], the score function used for single ligand docking DOCKVISION^[86, 87]. Authors have also provided a very useful appendix comprising the details of the score functions: Autodock, LigScore, PLP, PMF, LUDI, FlexX, GOLD, DOCK, Chem Score, DrugScore and X Score^[88]. Other docking score functions of interest includes GLIDE^[89], DockVision^[86, 87], ICM^[7], SurFlex^[90].

APPLICATIONS OF MOLECULAR DOCKING

A binding interaction between a small molecule ligand and an enzyme protein may result in activation or inhibition of the enzyme. If the protein is a receptor, ligand binding may result in agonism or antagonism. Docking is most commonly used in

the field of drug design. Most drugs are organic molecules, and docking may be applied for:

Hit identification – docking combined with a scoring function can be used to quickly screen large databases of potential drugs in silico to identify molecules that are likely to bind to protein target of interest.

Lead optimization – docking can be used to predict in where and in which relative orientation a ligand binds to a protein (i.e. binding mode or pose). This information may in turn be used to design more potent and selective analogs.

Bioremediation – Protein ligand docking can also be used to predict pollutants that can be degraded by enzymes^[91].

In contrast to proteins, nucleic acids have received much less attention as drug targets. Drugs known to interact with DNA include: groove binders (daunomycin), intercalators (actinomycin) and alkylating agents (cisplatin)^[92]. The variability in DNA structures is relatively small. The folds observed in RNA structures such as ribozymes and ribosomes^[93, 94], comparable in complexity to those of proteins, make RNA's attractive as drug targets^[95-98]. Very little effort has been devoted to the rational design of ligands for RNA targets. In the last few years a number of crystal and NMR structures of interesting RNA drug targets have appeared in the literature. An important difference between protein and RNA targets relates to binding pocket location. In case of proteins the binding pocket typically lies rather deep in the interior region and the cavity is well separated from solvent. In RNA targets the binding pocket is located along the surface and is therefore relatively exposed to solvent. The highly charged nature of the target RNA's phosphate backbone requires that electrostatic interactions be handled more accurately than typically needed for proteins. Based on DOCK screening aminoglycosides are identified and

capable of binding with RNA duplex but not the DNA [95].

CONCLUSIONS

In this review we focused on molecular docking and scoring by the description of several applications. The aim of a docking procedure is often the discovery of new lead candidates. 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. Novel algorithms will arise to find new solutions to docking problems and overcome the limitations of recently developed scoring functions. Especially the issue of protein flexibility and induced-fit motions of the protein will gain in importance over the coming years in the design and discovery of novel lead candidates by means of protein-ligand docking and scoring. It is important to point out that the end-user should pay attention to the documented validation studies performed at various levels of development of a given docking program. Docking small rigid molecules to receptor structures is straightforward (e.g. staurosporine to kinases, steroids to the estrogen receptor etc.). Ligand flexibility, permutations and combinations of stereoisomers and possible protonation states pose additional challenges to the docking problem, enormously increasing the total number of structures that need to be sampled in a virtual screening experiment. Considering the magnitude of the problem at hand (having to dock millions of molecules for any given receptor), one could justifiably be intimidated. However, by applying intelligent filters the number of molecules that actually needs to be docked can be substantially reduced. The sensitivity of docking programs to the initial ligand conformation is still an open question. In addition, for a given target it is not clear how many discrete receptor conformations need to be included in a docking calculation. The

good news is that search methods are improving. Better scoring schemes, adequate incorporation of solvent effects and methods to reliably accommodate receptor flexibility are areas of active research that hold much promise. Specific entropic contributions are largely ignored. Currently, there is no reliable way to account for the energy differences between receptor-bound and unbound (free) ligand conformations. An indirect way of including this effect has been achieved by an additional 'ad hoc' term in the scoring function that correlates with the number of rotatable bonds in the ligand [96, 99]. Despite all the indicated limitations, significant progress in docking methodology has been made in the recent past. Computational docking calculations are routinely being performed at various stages of the drug discovery process. The power of docking calculations has been well-recognized by interdisciplinary teams in the pharmaceutical industry. As the field of docking-based virtual screening matures, this recognition will undoubtedly increase. It is hoped that appropriate and widely accepted sets of test data will become established, that the methods will evolve to facilitate the comparisons required to define the new frontiers.

ABBREVIATIONS

ACD = Available Chemicals Directory
CMC = Comprehensive Medicinal Chemistry database
CSD = Cambridge Structural Database
MDDR = MDL Drug Data Report
MM/PBSA = Molecular Mechanics/ Poisson-Boltzmann Surface Area
NCI = National Cancer Institute database
NMR=Nuclear magnetic resonance

REFERENCES

- 1) Lengauer T, Rarey M. *Computational methods for biomolecular docking. Curr Opin Struct Biol* 1996; **6** (3): 402-6.
- 2) Kitchen DB, Decornez H, Furr JR, Bajorath J. *Docking and scoring in virtual screening for drug discovery: methods and applications. Nature reviews Drug discovery* 2004; **3** (11): 935-49.

- 3) Kuntz ID, Blaney JM, Oatley SJ, Langridge R, Ferrin TE. A geometric approach to macromolecule-ligand interactions. *J Mol Biol* 1982; 161: 269-88.
- 4) DesJarlais RL, Sheridan RP, Seibel GL, Dixon JS, Kuntz ID, Venkataraghavan R. Using shape complementarity as an initial screen in designing ligands for a receptor binding site of known three-dimensional structure. *J Med Chem* 1988; 31: 722-9.
- 5) Rarey M, Kramer B, Lengauer T, Klebe G. A fast flexible docking method using an incremental construction algorithm. *J Mol Biol* 1996; 261: 470-89.
- 6) Jones G, Willett P, Glen RC, Leach AR, Taylor R. Development and validation of a genetic algorithm for flexible docking. *J Mol Biol* 1997; 267: 727-48.
- 7) Abagyan R, Totrov M, Kuznetsov D. ICM - a new method for protein modelling and design. Applications to docking and structure prediction from the distorted native conformation. *J Comput Chem* 1994; 15: 488-506.
- 8) MDDR, www.mdll.com/products/knowledge/drug_data_report/index.jsp.
- 9) CMC, www.mdll.com/products/knowledge/medicinal_chem/index.jsp.
- 10) ACD, www.mdll.com/products/experiment/available_chem_dir/index.jsp.
- 11) CSD, www.ccdc.cam.ac.uk/prods/csd/csd.html.
- 12) NCI, www.mdll.com/products/knowledge/cancer_db/index.jsp.
- 13) Ajay, Murcko MA. Computational methods to predict binding free energy in ligand receptor complexes. *J Med Chem* 1995; 38: 4953- 67.
- 14) Bohm HJ, Stahl M. Rapid empirical scoring functions in virtual screening. *Med Chem Res* 1999; 9: 445-462.
- 15) Tame JR. Scoring functions: a view from the bench. *J Comput Aided Mol Des* 1999; 13: 99-108.
- 16) Charifson PS, Corkery JJ, Murcko MA, Walters WP. Consensus scoring: A method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins. *J Med Chem* 1999; 42: 5100-9.
- 17) Taylor RD, Jewsbury PJ, Essex JW. A review of protein-small molecule docking methods. *J Comput Aided Mol Des* 2002; 16: 151-66.
- 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: 409-43.
- 19) Shoichet BK, McGovern SL, Wei B, Irwin JJ. Lead discovery using molecular docking. *Curr Opin Chem Biol* 2002; 6: 439-46.
- 20) Muegge I, Rarey M. In Lipkowitz KB, Boyd DB. Ed *Reviews in Computational chemistry*, New York, John Wiley & Sons 2001, 17: pp 1-60.
- 21) Abagyan R, Totrov M. High-throughput docking for lead generation. *Curr Opin Chem Biol* 2001; 5: 375 82.
- 22) Meng EC, Shoichet BK, Kuntz ID. Automated docking with grid-based energy evaluation. *J Comput Chem* 2004; 13 (4): 505-524.
- 23) Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, Olson AJ. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J Comput Chem* 1998; 19 (14): 1639-1662.
- 24) Feig M, Onufriev A, Lee MS, Im W, Case DA, Brooks CL. Performance comparison of generalized born and Poisson methods in the calculation of electrostatic solvation energies for protein structures. *J Comput Chem* 2004; 25 (2): 265-84.
- 25) Shoichet BK, Kuntz ID, Bodian DL. "Molecular docking using shape descriptors. *J Comput Chem* 2004; 13 (3): 380-397.
- 26) Cai W, Shao X, Maigret B. Protein-ligand recognition using spherical harmonic molecular surfaces: towards a fast and efficient filter for large virtual throughput screening. *J Mol Graph Model* 2002; 20 (4): 313-28.
- 27) Morris RJ, Najmanovich RJ, Kahraman A, Thornton JM. Real spherical harmonic expansion coefficients as 3D shape descriptors for protein binding pocket and ligand comparisons. *Bioinformatics* 2005; 21 (10): 2347-55.
- 28) Kahraman A, Morris RJ, Laskowski RA, Thornton JM. Shape variation in protein binding pockets and their ligands. *J. Mol. Biol.* 2007; 368 (1): 283-301.
- 29) http://www.phy.cam.ac.uk/camphy/dna/dna13_1.htm.
- 30) Molecular Networks GmbH, Erlangen, Germany.
- 31) Tripos: St. Louis, MO, USA.
- 32) SciTegic: San Diego, CA, USA.
- 33) Muegge I. The effect of small changes in protein structure on predicted binding modes of known inhibitors of influenza virus neuraminidase: PMF-scoring in DOCK4. *Med Chem Res* 1999; 9: 490-500.
- 34) McGovern SL, Shoichet BK. Information decay in molecular docking screens against holo, apo, and modeled conformations of enzymes. *J Med Chem* 2003; 46: 2895-907.

- 35) Schapira M, Abagyan R, Totrov M. Nuclear hormone receptor targeted virtual screening. *J Med Chem* 2003; 46: 3045-59.
- 36) Davis AM, Teague SJ, Kleywegt GJ. Application and Limitations of X-ray Crystallographic Data in Structure-Based Ligand and Drug Design. *Angew Chem Int Ed Engl* 2003; 42: 2718-36.
- 37) Kleywegt GJ, Henrick K, Dodson EJ, van Aalten DM. Pound-wise but penny-foolish: How well do micromolecules fare in macromolecular refinement? *Structure (Camb)* 2003; 11: 1051-9.
- 38) Bostrom J. Reproducing the conformations of protein-bound ligands: a critical evaluation of several popular conformational searching tools. *J Comput Aided Mol Des* 2001; 15: 1137-52.
- 39) Hopkins SC, Vale RD, Kuntz ID. Inhibitors of kinesin activity from structure-based computer screening. *Biochemistry* 2000; 39: 2805-14.
- 40) Tondi D, Slomczynska U, Costi MP, Watterson DM, Ghelli S, Shoichet BK. Structure-based discovery and in-parallel optimization of novel competitive inhibitors of thymidylate synthase. *Chem Biol* 1999; 6: 319-31.
- 41) Aronov AM, Munagala NR, Ortiz De Montellano PR, Kuntz ID, Wang CC. Rational design of selective submicromolar inhibitors of *Tritrichomonas foetus* hypoxanthine-guanine-xanthine phosphoribosyltransferase. *Biochemistry* 2000; 39: 4684-91.
- 42) Perola E, Xu K, Kollmeyer TM, Kaufmann SH, Prendergast FG, Pang YP. Successful virtual screening of a chemical database for farnesyltransferase inhibitor leads. *J Med Chem* 2000; 43: 401-8.
- 43) Olson AJ, Goodsell DS. Automated docking and the search for HIV protease inhibitors. *SAR QSAR Environ Res* 1998; 8: 273-85.
- 44) Powers RA, Morandi F, Shoichet BK. Structure-based discovery of a novel, noncovalent inhibitor of AmpC beta-lactamase. *Structure (Camb)* 2002; 10: 1013-23.
- 45) Bursavich MG, Rich DH. Designing non-peptide peptidomimetics in the 21st century: inhibitors targeting conformational ensembles. *J Med Chem* 2002; 45: 541-58.
- 46) Leulliot N, Varani G. Current topics in RNA-protein recognition: control of specificity and biological function through induced fit and conformational capture. *Biochemistry* 2001; 40: 7947-56.
- 47) Davis AM, Teague SJ. Hydrogen bonding, hydrophobic interactions and failure of the rigid receptor hypothesis. *Angew Chem Int Ed Engl* 1999; 38: 736-749.
- 48) Verkhivker GM, Bouzida D, Gehlhaar DK, Rejto PA, Freer ST, Rose PW. Complexity and simplicity of ligand-macromolecule interactions: the energy landscape perspective. *Curr Opin Struct Biol* 2002; 12: 197-203.
- 49) Koshland DE. Application of a theory of enzyme specificity to protein synthesis. *Proc Natl Acad Sci USA* 1958; 44: 98-104.
- 50) Jorgensen WL. Rusting of the lock and key model for proteinligand binding. *Science* 1991; 254: 954-5.
- 51) Jiang F, Kim SH. "Soft docking": matching of molecular surface cubes. *J Mol Biol* 1991; 219: 79-102.
- 52) Schneidman-Duhovny D, Inbar Y, Polak V, Shatsky M, Halperin I, Benyamini H, et al. Taking geometry to its edge: fast unbound rigid (and hinge-bent) docking. *Proteins* 2003; 52: 107-12.
- 53) Li CH, Ma XH, Chen WZ, Wang CX. A protein-protein docking algorithm dependent on the type of complexes. *Protein Eng* 2003; 16: 265-9.
- 54) Li CH, Ma XH, Chen WZ, Wang CX. A soft docking algorithm for predicting the structure of antibody-antigen complexes. *Proteins* 2003; 52: 47-50.
- 55) Palma PN, Krippahl L, Wampler JE, Moura JJ. BiGGER: a new (soft) docking algorithm for predicting protein interactions. *Proteins* 2000; 39: 372-84.
- 56) Fernandez-Recio J, Totrov M, Abagyan R. Soft protein-protein docking in internal coordinates. *Protein Sci* 2002; 11: 280-91.
- 57) Morelli X, Czjzek M, Hatchikian CE, Bornet O, Fontecilla-Camps JC, Palma NP, et al. Structural model of the Fehydrogenase/ cytochrome c553 complex combining transverse relaxation-optimized spectroscopy experiments and soft docking calculations. *J Biol Chem* 2000; 275: 23204-10.
- 58) Morelli X, Dolla A, Czjzek M, Palma PN, Blasco F, Krippahl L, et al. Heteronuclear NMR and soft docking: an experimental approach for a structural model of the cytochrome c553-ferredoxin complex. *Biochemistry* 2000; 39: 2530-7.
- 59) Morelli XJ, Palma PN, Guerlesquin F, Rigby AC. A novel approach for assessing macromolecular complexes combining softdocking calculations with NMR data. *Protein Sci* 2001; 10: 2131-7.
- 60) Leach AR. Ligand docking to proteins with discrete side-chain flexibility. *J Mol Biol* 1994; 235: 345-56.
- 61) Schaffer L, Verkhivker GM. Predicting structural effects in HIV-1 protease mutant complexes with flexible ligand docking and protein side-chain optimization. *Proteins* 1998; 33: 295-310.

- 62) David L, Luo R, Gilson MK. Ligand-receptor docking with the Mining Minima optimizer. *J Comput Aided Mol Des* 2001; 15: 157-71.
- 63) Head MS, Given JA, Gilson MK. "Mining Minima": Direct computation of conformational free energy. *J Phys Chem* 1997; 101: 1609-18.
- 64) Kairys V, Gilson MK. Enhanced docking with the mining minima optimizer: acceleration and side-chain flexibility. *J Comput Chem* 2002; 23: 1656-70.
- 65) Anderson AC, O'Neil RH, Surti TS, Stroud RM. Approaches to solving the rigid receptor problem by identifying a minimal set of flexible residues during ligand docking. *Chem Biol* 2001; 8: 445-57.
- 66) Nakajima N, Higo J, Kidera A, Nakamura H. Flexible docking of a ligand peptide to a receptor protein by multicanonical molecular dynamics simulation. *Chem Phys Lett* 1997; 278: 297-301.
- 67) Abseher R, Horstink L, Hilbers CW, Nilges M. Essential spaces defined by NMR structure ensembles and molecular dynamics simulation show significant overlap. *Proteins* 1998; 31: 370-82.
- 68) Fernandez-Recio J, Totrov M, Abagyan R. ICM-DISCO docking by global energy optimization with fully flexible side-chains. *Proteins* 2003; 52: 113-7.
- 69) Lin JH, Perryman AL, Schames JR, McCammon JA. The relaxed complex method: Accommodating receptor flexibility for drug design with an improved scoring scheme. *Biopolymers* 2003; 68: 47-62.
- 70) Taylor RD, Jewsbury PJ, Essex JW. FDS: flexible ligand and receptor docking with a continuum solvent model and soft-core energy function. *J Comput Chem* 2003; 24: 1637-56.
- 71) Still WC, Tempezyk A, Hawley RC, Hendrickson T. Semianalytical treatment of solvation for molecular mechanics and dynamics. *J Am Chem Soc* 1990; 112: 6127-29.
- 72) Bohacek, R.; McMartin, C.; Glunz, P.; Rich, D. H. *IMA Volumes in Mathematics and Its Applications, Rational Drug Design*, 1999; 103-114.
- 73) Krovat, E. M.; Langer, T. *J Chem Inf Comput Sci* 2004, 44, 1123-1129.
- 74) Bleicher, K. H.; Böhm, H.-J.; Müller, K.; Alainine, A. I. *Nat Rev Drug Discov* 2003, 2, 369-378.
- 75) Wolber, G.; Langer, T. 2004, <http://www.inteligand.com/>.
- 76) Wolber, G.; Langer, T. *Rational Approaches to Drug Design, Proceedings of the European Symposium on Quantitative Structure-Activity Relationships, 2000*, 390-399.
- 77) Waszkowycz, B. *Curr Opin Drug Discov Devel* 2002, 5, 407-13.
- 78) Gruneberg S, Stubbs MT, Klebe G. Successful virtual screening for novel inhibitors of human carbonic anhydrase: strategy and experimental confirmation. *J Med Chem* 2002; 45: 3588-602.
- 79) Van Drie JH. Pharmacophore discovery--lessons learned. *Curr Phar Des* 2003; 9: 1649-64.
- 80) Mustata GI, Briggs JM. A structure-based design approach for the identification of novel inhibitors: application to an alanine racemase. *J Comput Aided Mol Des* 2003; 16: 935-53.
- 81) Rastelli G, Pacchioni S, Sirawaraporn W, Sirawaraporn R, Parenti MD, Ferrari AM. Docking and database screening reveal new classes of Plasmodium falciparum dihydrofolate reductase inhibitors. *J Med Chem* 2003; 46: 2834-45.
- 82) Muegge I, Martin YC. A general and fast scoring function for protein-ligand interactions: a simplified potential approach. *J Med Chem* 1999; 42: 791-804.
- 83) Gehlhaar DK, Verkhivker GM, Rejto PA, Sherman CJ, Fogel DB, Fogel LJ, et al. Molecular recognition of the inhibitor AG-1343 by HIV-1 protease: conformationally flexible docking by evolutionary programming. *Chem Biol* 1995; 2: 317-24.
- 84) Eldridge MD, Murray CW, Auton TR, Paolini GV, Mee RP. Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J Comput Aided Mol Des* 1997; 11: 425-45.
- 85) Meng EC, Shoichet BK, Kuntz ID. Automated docking with gridbased energy evaluation. *J Comput Chem* 1992; 13: 505-24.
- 86) Hart TN, Read RJ. A multiple-start Monte Carlo docking method. *Proteins* 1992; 13: 206-22.
- 87) Hart TN, Ness SR, Read RJ. Critical evaluation of the research docking program for the CASP2 challenge. *Proteins* 1997; Suppl 1: 205-9.
- 88) Wang R, Lu Y, Wang S. Comparative evaluation of 11 scoring functions for molecular docking. *J Med Chem* 2003; 46: 2287-303.
- 89) <http://www.schrodinger.com/Products/glide.html>
- 90) Jain AN. Surflex: fully automatic flexible molecular docking using a molecular similarity-based search engine. *J Med Chem* 2003; 46: 499-511.
- 91) Suresh PS, Kumar A, Kumar R, Singh VP. An in silico [correction of insilico] approach to bioremediation: laccase as a case study. *J Mol Graph Model* 2008; 26 (5): 845-9.

- 92) Demeunynck M, Bailly C, Wilson WD. Ed. *DNA and RNA Binders*, Weinheim, Germany, Wiley-VCH. 2002.
- 93) Carter AP, Clemons WM, Brodersen DE, Morgan-Warren RJ, Wimberly BT, Ramakrishnan V. *Functional insights from the structure of the 30S ribosomal subunit and its interactions with antibiotics. Nature* 2000; 407: 340-8.
- 94) Batey RT, Rambo RP, Lucast L, Rha B, Doudna JA. *Crystal structure of the ribonucleoprotein core of the signal recognition particle. Science* 2000; 287: 1232-9.
- 95) Chen Q, Shafer RH, Kuntz ID. *Structure-based discovery of ligands targeted to the RNA double helix. Biochemistry* 1997; 36: 11402-7.
- 96) Filikov AV, Mohan V, Vickers TA, Griffey RH, Cook PD, Abagyan RA, et al. *Identification of ligands for RNA targets via structure-based virtual screening: HIV-1 TAR. J Comput Aided Mol Des* 2000; 14: 593-610.
- 97) Du Z, Lind KE, James TL. *Structure of TAR RNA complexed with a Tat-TAR interaction nanomolar inhibitor that was identified by computational screening. Chem Biol* 2002; 9: 707-12.
- 98) Lind KE, Du Z, Fujinaga K, Peterlin BM, James TL. *Structurebased computational database screening, in vitro assay and NMR assessment of compounds that target TAR RNA. Chem Biol* 2002; 9: 185-93.
- 99) Klebe G, Bohm HJ. *Energetic and entropic factors determining binding affinity in protein-ligand complexes. J Recept Signal Transduct Res* 1997; 17: 459-73.

Author information: Gaba MONIKA is working in Department of Pharmaceutical Chemistry, ASBASJSM College of Pharmacy, BELA (Ropar)-140111, Punjab, India

Article History:-----

Date of Submission: 12-02-10

Date of Acceptance: 18-04-10

Conflict of Interest: None

Source of Support: Nil