Molecular Docking: Novel Drug Discovery And Drug Designing Tool

Vaibhav Amilkanthawar¹,

Dr.Shubhangi Daswadkar²,

Vasanti patil³,

Lalita Joshi⁴

Dr. D.Y. Patil College of Pharmacy Akurdi, Pune ,India

Mail Id- mailvaibhav.s.a@gmail.com

Index

Sr.	Content	Page No.
No.	Content	1 age 110.
1	Abstract	2
2	Keywords	2
3	Introduction	3
4	Drug discovery procedure	4
5	Objective	5
6	Type of docking	6
7	Types of interaction	6
8	Strategies	8
9	Virtual screening	12
10	Limitations	12
11	Applications	13
12	Docking programs	14
13	Conclusion	15
14	References	17

Abstract

Molecular docking is making model of the structure which is formed by interaction of two or more structure by computational means. The main goal of the molecular docking is to predict the 3D structure of molecule with maximum site of attachment. Docking makes the various probable and acceptable structure of the molecule and then they are rank accordingly how they are available in nature with the help of scoring function. This review contains the molecular docking approaches, different types of targets, drug discovery procedure, different types of interaction, different search algorithms and the scoring functions, different strategies, applications and limitations are also presented in this article. Molecular docking is been developed and improving from so many years, but bringing it to a medicine or to the drug market effectively is still generally a big question. In most of the cases, the structures of drug and the interacting drug residues on the target protein are also been showed. It provides us the confidence that the docking will be applied in the industry and basic research. Moreover, we can apply molecular docking and related technologies to create new therapies for disease.

Key Words

Computational drug design, Molecular docking, scoring function, Virtual screening, docking algorithm.

1. Introduction

Molecular docking is the one of the type of computer aided drug design that can be used in drug discovery and drug designing [1-4]. It is a form of structure-based drug design that helps in deciding the binding affinities between micromolecular and macromolecular targets i.e. proteins. The first step in molecular docking is selection of drug target. We can use any macromolecule as a target. Commonly used targets (fig.1) are enzymes, protein binding sites, receptors, transporter and regulatory elements. Next step is the determination and prediction of three-dimensional structure. Structures can be determined using X-rays, NMR, or electron microscopy (EM). Thousands of popular target structures are available in the protein data bank (PDB)^[5]. Many drug targets have known binding sites; if not, software that can predict potential binding sites for different ligands. Docking studies can be performed using known and novel ligands. Virtual screening is nothing but identifying novel ligands with molecular docking which is an extremely useful but time-consuming method of drug discovery. Here we design molecules having high binding affinity to a specific site. Docking studies are more validated using further computational methods, such as molecular dynamic simulation. The most successful candidates from computational trials can be tested in-vitro or in-vivo, for the progress to clinical trials [6].

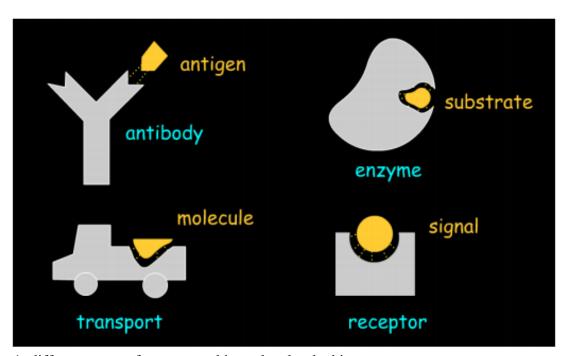


Fig. 1: different types of targets used in molecular docking

2. Novel Drug Discovery Procedure:

Molecular docking is a part of Novel Drug Discovery Procedure which contains drug discovery along with the drug development and make it available in market for treatment ^[7]. The process is,

- 1) **Drug target identification:** Drug target identification is the process of identifying the effective site for the drug molecule. It might be active site of any molecule for example protein and nucleic acid for having maximum pharmacological action.
- 2) Drug target validation: It is also known as High throughput screening (HTS) perform using robotics, data processing software, liquid handling devices, and sensitive detectors. This permit researchers to quickly conduct various genetic, chemical or pharmacological tests, through which we can rapidly identify active compounds.
- 3) Lead identification: It is also called as Lead Generation it is a stage in early drug discovery where the small molecule hits from a High throughput screen (HTS) are evaluated and undergo optimization to identify lead compound.
- 4) Lead optimisation: These lead compounds undergo optimization in a subsequent step of drug discovery called lead optimisation. This optimization is done through chemical modification of the hit structure, with modifications chosen by employing knowledge of the structure-activity relationship (SAR) as well as structure-based design if structural information about the target is available.
- 5) **Preclinical testing:** In drug development the preclinical studies is a stage of research that begins before clinical trial. The main goal is to determine the safe dose for first in men and to determine safety profile.
- 6) Clinical testing: This includes safety, toxicity, Pharmacokinetics and metabolism study of new drug in humans. It involves four types: Phase I, Phase II, Phase III and Phase IV

7) Marketing: The full cost of bringing a new drug in the market (from discovery to clinical trials approval) It is complex and controversial process where companies spend ten to hundreds of millions of U.S. dollars.

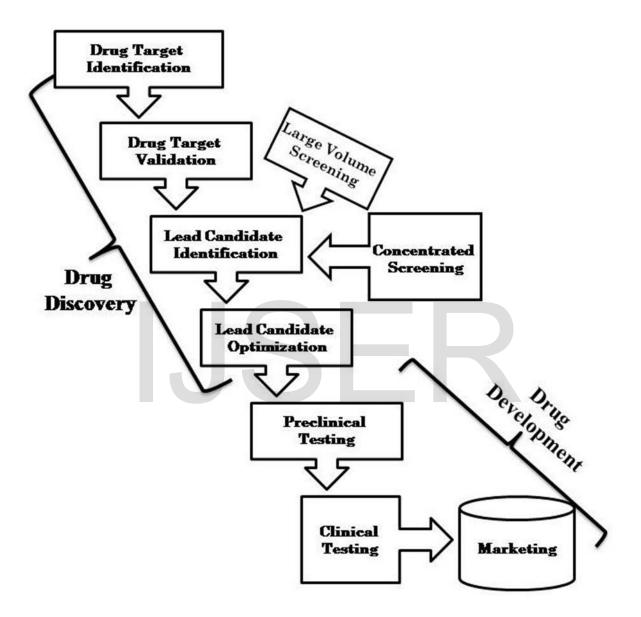


Figure 2: A brief flowchart of novel drug discovery procedure

3. OBJECTIVES

- To synthesize lead compound.
- New analogues with improved potency.
- Reduced off-target activities.
- Physiochemical properties of drug by *in vivo* pharmacokinetics.

4. TYPES OF DOCKING

- 1) **Flexible docking**: Freedom of rotation to the ligand or receptor or both, depending upon this they have two types^[8]
 - (a) **Flexible ligand docking**: where the receptor is held rigid, but the ligand is treated as flexible.
 - **(b) Flexible docking**: where both receptor and ligand flexibility is considered.
- 2) **Rigid body docking:** Does not allow the freedom of rotation where, both the receptor and small molecule are treated as rigid [8].

The commonly preferred docking algorithms uses a rigid receptor/flexible ligand model. The principle docking methods that are extensively use search algorithms based on Monte Carlo, genetic algorithm, fragment-based and molecular dynamics.

Some programs that are well-suited for docking of a large database of molecules include: DOCK ^{[9] [10]}, FlexX ^[11], GOLD^[12], and ICM^[13].

5. DIFFERENT TYPES OF INTERACTIONS

Interaction can be defined as the force of attraction between particles or the target site and the molecule. The various types of interaction in docking given below,

A) Electrostatic forces: The electrostatic forces form due to the charges on the matter. (fig.3) They are charge-charge, charge-dipole and dipole-dipole interactions.

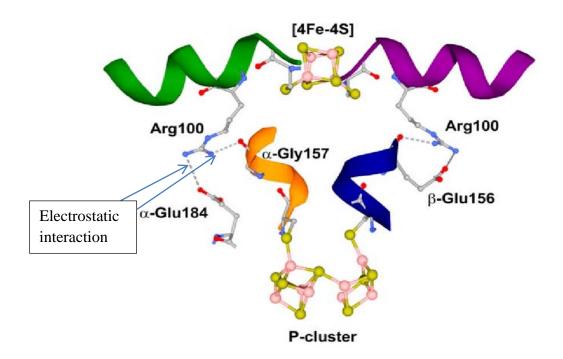


Fig.3:Electrostatic interaction

B) Electrodynamics forces: This interaction is known as the Van der Waals interactions. It is a distance-dependent interaction between atoms or molecules. These attractions do not result from a chemical electronic bond and they are comparatively weak (fig.4)

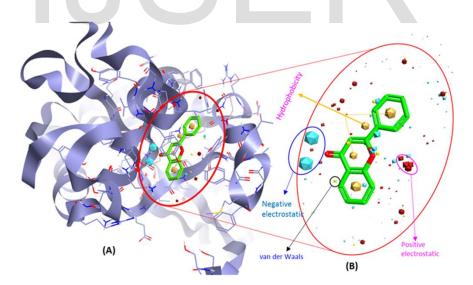


Fig.4Electrodynamics interaction

C) Steric forces/Hindrance: These forces are due to the atoms of different molecules come into a very close contact with each other and start affecting the reactivity of each other. This force can affect chemical reactions and the free energy of a system. (fig.5)

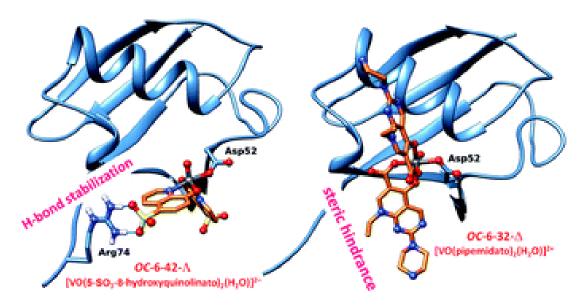


Fig 5 Stearic hindrance

D) Solvent-related forces: These are the forces formed due to the chemical reactions between the solvent and the protein or ligand.

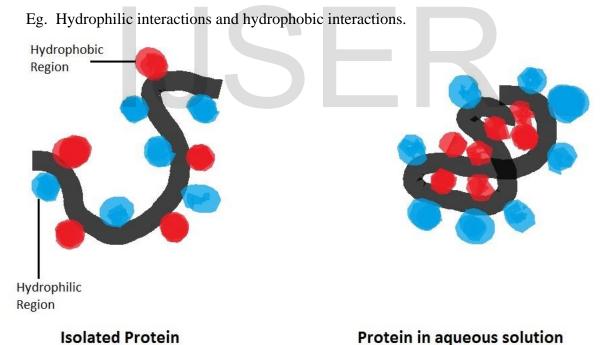


Fig6: hydrophilic and Hydrophobic interaction.

6. Strategies Of Molecular Docking

A) Search Algorithm/Strategy

1. Shape Matching:

This is simplest and commonly used method where we know the active site and accordingly the ligands were made [8].

2. Systamatic Search Method

A systematic search method does the slight variations in the structural parameters, gradually changing the conformation of the ligands ^[15]. The algorithm probes the energy landscape of the conformational space converges to the minimum energy solution relating it to to the most likely binding mode (Figure 7). In spite of the method is functional in exploring the conformational space, it can approach other to a local minimum rather than the global minimum. This type of drawback can be surmount by performing simultaneous searches starting from distinct points of the energy landscape ^[16].

This method also has the three subtypes,

I) Exhaustive

This method is only for rigid body, we cannot perform the exhaustive systematic search on flexible body. It is the mostly used straightforward method and all higher and lower energy confirmation will be detected [8].

I.i) Conformation

This method is used for both rigid and flexible body. This method has overcome on the problems of exhaustive method ^[8].

I.ii) Fragment Based

In this method the molecule is fragmented separately and then dock them separately and see the perfect match of molecule with the ligand^[8].

3. Stochastic Algorithm

This method is completely based on keep on fixing the ligand on the active site unless and until the perfectly fit ligand is not found. The stochastic algorithm is the continuous shape matching method where we develop the ligand and dock it for checking whether it is fit at active site if it is not fit then modified it and re-docks it until the perfect match is not found [8]. Genetic Algorithm is a stochastic search method which is shown below [8]. All confirmation will rank by the survival of the fittest method means the higher compatible will prefer first [8].

A stochastic method is done by randomly modifying the structural parameters of the ligand ^[17]. For instance, systematic search methods explore all combinations of the structural parameters. The number of possible combinations grows rapidly as the degrees of freedom connected with the ligand increases which results in a phenomenon called as combinatorial explosion.

Different docking programmes like FRED, Surflex and DOCK solve this problem by applying an increasing construction algorithm in which the ligand is gradually built in the binding site (Figure8)^[18–20]. Here, the chemical structure is initially broken into several fragments (Figure 7A). And then, one of these parts is selected as an anchor fragment and is docked in a complementary region of the binding site (Figure 7B) while the remaining fragments are sequentially added (Figure 7C–E). The process continues until the entire ligand is been constructed. The algorithm performs the conformational search only for the fragments being added, it reduces the degrees of freedom to be explored, and thereby it avoids the combinatorial explosion ^[21].

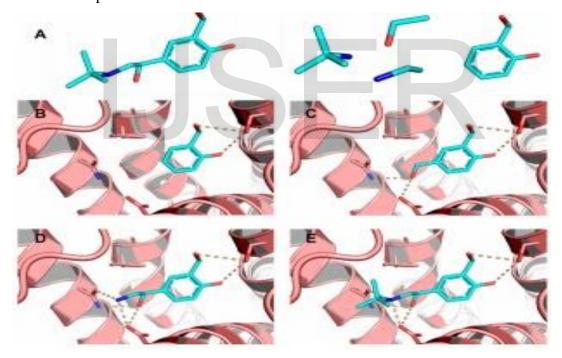


Fig. 7The incremental construction method

A)The ligand which consist stick representation include carbon in cyan is broken into various fragments;

B)The first is anchor fragment (cartoon representation, carbon in salmon)

C)Another fragment is docked after the anchor fragment; (D and E). The other fragments are docked orderly to build the entire ligand in its binding conformation. Residues are shown as carbon in salmon. Hydrogen bonds are represented as dash lines.

Genetic algorithms (GA) are an interesting application of the random search, which are used in molecular docking programs such as Auto Dock and Gold [22][23]. The GA algorithm has the high computational cost associated with stochastic methods by applying concepts of the theory of evolution and natural selection. In the first step the algorithm encodes all of the structural parameters of the initial structure in a chromosome, which is shown by a vector. Starting from this chromosome, the random search algorithm generates an initial population of chromosomes covering a wide area of the energy. This population is evaluated and the most adapted chromosomes (i.e., those with the lowest energy values) are selected as templates for the generation of the next population. [24]

This procedure decreases the average energy of the chromosome by transmitting favourable structural characteristics from one population to another, reducing therefore, the conformational space to be explored. The GA routine is repeatedly executed and, after a reasonable number of conformational search-and-evaluation cycles [24].

B) Scoring Function (binding affinity of ligand)

The scoring function takes a pose as input and returns a number indicating that the pose shows a favourable binding interaction. The scoring functions are physics based molecular mechanics force fields which determine the energy of the pose i.e. low or negative energy and shows stable system also an favourable binding interaction. An non-traditional approach is to derive a statistical potential for the interactions from a extent of database of protein-ligand complexes, such as the Protein Data Bank, and assesed the fit of the pose according to their potential.

There are a considerable number of structures obtained from X-ray crystallography for the complexes between proteins and high affinity ligands, but comparatively they are not definite for low affinity ligands as the later complexes tend to be less stable and therefore they are more difficult to crystallize. Scoring functions trained with this data can easily dock high affinity ligands not only correctly and effectively but also give possible docked conformations for ligands which do not bind. This gives a large number of false positive hits which includes ligands predicted to bind to the proteins that actually do not bind when they placed together in a test tube. The one of the way to reduce number of false positives is to recalculate the energy of the top scoring poses by using Generalized Born or Poisson-Boltzmann methods.

Scoring functions can be generalised into three distinct categories: knowledge-based, empirical and force field-based. The first one is Knowledge-based scoring functions which depends on statistical means to extract rules on both preferred, and non-preferred atom pair interactions from experimentally decided protein-ligand complexes. The rules are also explain as pair- potentials which are subsequently used for score ligand binding poses. The PMF score ^[25] 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 the training set of protein-ligand complexes. The weights are selected by regression methods. Examples of empirical scoring functions include PLP ^[26], ChemScore ^[27] and the FlexX scoring function.

Force field scoring functions also used to 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 ^[28], the score function used for single ligand docking DOCKVISION^[29,30], Autodock, LigScore, PLP, PMF, LUDI, FlexX, GOLD, DOCK, Chem Score, DrugScore and X Score ^[31]. Other docking score functions of interest includes GLIDE ^[32], DockVision^[29,30], ICM^[7], SurFlex^[33].

7. Virtual Screening

Virtual screening is a widely accepted method in lead discovery because it is useful 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 pockets. Docking may help in this case to identify the active site in detail and uncovered binding pockets or interaction points. This approach is carried out by using various de novo design tools for e.g. ligand construction and docking in GROWMOL [34]. These applications can result in the synthesis of new strategies. Docking of virtual combinatorial libraries can also forms the innovative ligands [35]. The aspect of privileged motive design can be implemented by using the innovative software tool focus to generate virtual libraries for the desired target [36,37].

This program can build libraries of drug-like organic molecules for rational lead structure discovery. The Compounds are created by combining user-defined fragments according to their state-of-the-art chemical knowledge. The technique of virtual screening is used for biological assays. It is cost as well as time efficient and has also contributed important advances for the lead discovery programs in various pharmaceutical companies. Often virtual screening techniques are used in combination with HT screening lead discovery tools [38].

The docking of huge molecule database against a specific target may give the new candidates for further lead development. In addition to this docking can also help to compare the experimentally determined biological activities, ligand poses, and predicted binding affinities by the docking program, also to evaluate the scoring functions and to find out a good score for the target protein. A frequently used method is the re-docking of a complex ligand in order to verify the validity of the docking and scoring algorithms.

8. Limitation Of Docking

- 1) Accurate prediction is not possible all the time.
- 2) Process of choosing an appropriate scoring function or algorithm for specific target.
- 3) Protein flexibility in docking program is not taking in count but it should be consider.

9. Application Of Docking

- 1) Docking is most commonly used in the field of drug design. Most drugs are organic molecules, and docking is use for the purpose of Hit identification. The docking is combined with a scoring function which is used for quickly screening of large databases of potential drugs for identifying the molecules that are bind to target of interest.
- **2**) Lead optimization and docking can be used to predict that where and in which relative orientation a ligand binds to a protein (i.e. binding mode or pose). This information is used for designing more potent and selective analogues. Bioremediation and Protein ligand docking can also be used to predict pollutants that can be degraded by enzymes ^[39].
- 3) 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) [40].
- 4) The variability in DNA structures is relatively small. The folds observed in RNA structures such as ribozymes and ribosomes [41,42], comparable in complexity to those of proteins, make RNA more attractive as drug targets[43,44]. There are little effort has been given for the rational design of ligands for the 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 very important difference between protein and RNA targets is their binding pocket location. The protein lies in the interior region separated from the solvents in comparison to the RNA targets the binding pocket which is located along the surface and hence is relatively exposed to the solvent. The highly charged nature of the target RNA is 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 [45].

10. Some Docking Programs Are

1) **Protein Ligand Docking:** AutoDock, DOCK, Gold, Glide FlexX, Fred, MOE, Surflex

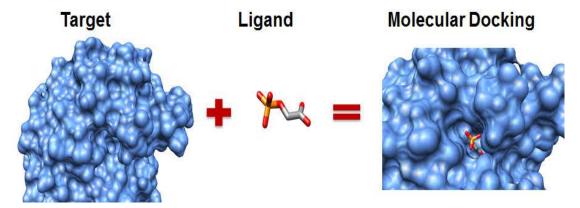


Fig. 8: Protein Ligand Docking

2) Protein-Protein Docking: ClusPro, ZDOCK, GRAMM-X, RosettaDock, DOT.

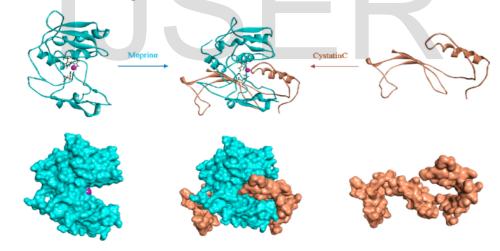


Fig.9: Protein-Protein Docking

3) Protein-Nucleic Acid Docking: ParaDock, HADDOCK, YASARA DOCK, DOT

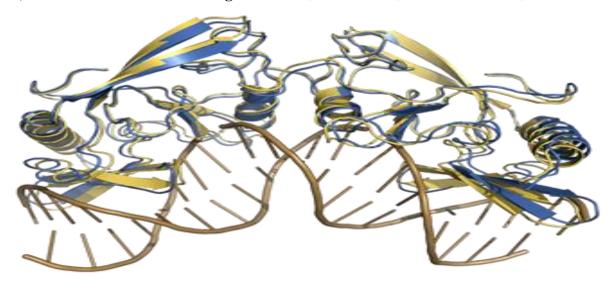


Fig. 10: Protein-Nucleic acid docking

4) Nucleic Acid-Ligand Docking: rDock, DNA Ligand Docking.

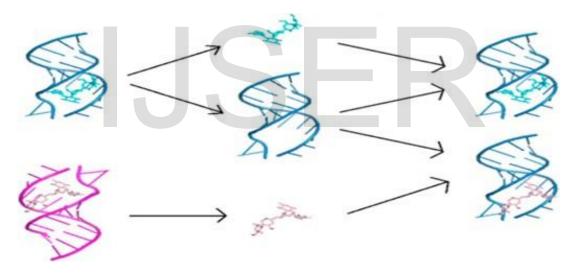


Fig.11: Nucleic acid-ligand Docking

Conclusion

In this review we focused on molecular docking and it's applications. The aim of a docking procedure is to discover new lead candidates and dentification of an overall reliable scoring function is one of the main challenges to be happened in the future. Rational algorithms find out new solutions and helpful to overcome the limitations in recent docking. Especially the issue of protein flexibility and induced-fit motions of the protein will have the importance over the coming years. Docking of small rigid molecules to receptor structures is straightforward. Whereas, by applying intelligent filters the number of molecules that

actually needs to be docked can be significantly decreased. For a given target it is not clear how many receptor conformations need to be included in a docking calculation. The good news is that search methods are improving with Better scoring schemes. Currently, there is no reliable way to account for the energy differences between receptor-bound and unbound or free ligands Despite of all the indicated limitations, significant progress in docking methodology has been made in the recent few years. Computational docking calculations are now performing at various stages of the drug discovery process. The power of docking calculations has been well-recognized by the pharmaceutical industry.

REFERENCES:

- [1]Pagadala NS, Syed K, Tuszynski J. Software for molecular docking: Biophysical Reviews. 2017; 9 (2):91-102
- [2]Wong CF. Flexible receptor docking for drug discovery. Expert Opinion on Drug Discovery. 2015; 10 (11):1189-1200
- [3]Ferreira LG, Dos Santos RN, Oliva G, Andricopulo AD. Molecular docking and structurebased drug design strategies. 2015; 20(7):13384-13421
- [4]Yuriev E, Ramsland PA. Latest developments in molecular docking: 2010–2011 in review. Journal of Molecular Recognition. 2013; 26 (5):215-239
- [5]Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The protein data bank. Nucleic Acids Research. 2000; 28(1):235-242
- [6]Mark Andrew Phillips, Marisa A. Stewart, Darby L. Woodling and Zhong-Ru Xie. Has Molecular Docking Ever Brought us a Medicine?. 2006;10(6):71-95
- [7]Gaba M, Singh A, Juihar S. An overview on Molecular Docking.International Journal of Drug Development & Research. 2010;2(2): 231-242
- [8]e-pathshala academy https://youtu.be/quXFgP_p5bE.
- [9]Kuntz ID, Blaney JM, Oatley SJ, Langridge R, Ferrin TE. A geometric approach to macromolecule-ligand interactions. Mol Biol. 1982;161:269-288.
- [10]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. Med Chem, 1988;31:722-729.
- [11]Rarey M, Kramer B, Lengauer T, Klebe G. A. fast flexible docking method using an incremental construction algorithm. Mol Biol, 1996;261:470-89.
- [12]Jones G, Willett P, Glen RC, Leach AR, Taylor R. Development and validation of a genetic algorithm for flexible docking. Mol Biol, 1997;267:727-748.

- [13] 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. Comput Chem, 1994;15: 488-506.
- [14]Moman E. Introduction to Molecular Docking.Pharmaceutical and Medicinal Chemistry. 2013;12:76-87
- [15]Sousa, S.F.; Fernandes, P.A.; Ramos, M.J. Protein-ligand docking: Current status and future challenges. Proteins Struct. Funct. Bioinform. 2006;65:15–26.
- [16]Zsoldos Z, Reid D, Simon A, Sadjad SB, Johnson AP. eHiTS: A new fast exhaustive flexible ligand docking system. Mol Graph Model. 2007;26:198–212.
- [17]Gorelik B, Goldblum A. High quality binding modes in docking ligands to proteins, Proteins Struct. Funct. Bioinform. 2008;71:1373–1386.
- [18]McGann, M. FRED and HYBRID docking performance on standardized datasets. Comput. Aided Mol. Des. 2012,26,897–906.
- [19]Jain AN. Surflex: Fully automatic flexible molecular docking using a molecular similarity-based search engine. Med. Chem. 2003;46:499–511.
- [20]Ewing TJ, Makino S, Skillman AG, Kuntz ID. DOCK 4.0: Search strategies for automated molecular docking of flexible molecule databases. Comput. Aided Mol. Des. 2001;15:411–428.
- [21]Dias R, de Azevedo J, Walter F. Molecular docking algorithms. Curr. Drug Targets. 2008; 9:1040–1047.
- [22]Morris GM, Goodsell, DS, Huey R, Olson AJ. Distributed automated docking of flexible ligands to proteins. Parallel applications of AutoDock 2.4. Comput. Aided Mol. Des.1996;10:293–304.
- [23]Jones G, Willett P, Glen RC, Leach AR, Taylor R. Development and validation of a genetic algorithm for flexible docking. Mol. Biol. 1997; 267: 727–748.
- [24]Krovat EM, Steindl T, Langer T. Recent advances in docking and scoring. Curr. Computer Aided Drug Des. 2005;1: 93–102.
- [25] 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.
- [26]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. Bio.1995; 2: 317-324.
- [27] 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 af inity of
- ligands in receptor complexes. J Comput Aided Mol Des 1997; 11: 425-445.

- [28]Meng EC, Shoichet BK, Kuntz ID. Automated docking with gridbased energy evaluation. J Comput Chem 1992; 13: 505-524.
- [29] Hart TN, Read RJ. A multiple-start Monte Carlo docking method. Proteins 1992;13:206-222.
- [30] Hart TN, Ness SR, Read RJ. Critical evaluation of the research docking program for the CASP2 challenge. Proteins 1997;1: 205-209.
- [31]Wang R, Lu Y, Wang S. Comparative evaluation of 11 scoring functions for molecular docking. J Med Chem 2003; 46: 2287-2303.
- [32]Waszkowycz B. Current Drug Discovery and Development. Computer aided drug discovery.2002;5:407-413.
- [33]Jain AN. Surflex: fully automatic flexible molecular docking using a molecular similarity-based search engine. Med Chem. 2003; 46: 499-511.
- [34] Suresh PS, Kumar A, Kumar R, Singh VP. An in silico approach to bioremediation. Mol Graph Model. 2008; 26 (5): 845-849
- [35] Demeunynck M, Bailly C, Wilson WD. DNA and RNA Binders. Germany Wiley-VCH. 2002:12:104-117.
- [36] 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: 340-348.
- [37]Batey RT, Rambo RP, Lucast L, Rha B, Doudna JA. Crystal structure of the ribonucleoprotein core of the signal recognition particle. Science 2000; 287: 1232-1239.
- [38]Chen Q, Shafer RH, Kuntz ID. Structure-based discovery of ligands targeted to the RNA double helix. Biochemistry.1997; 36: 11402-11407.
- [39]Filikov AV, Mohan V, Vickers TA, Grifey RH, Cook PD, Abagyan RA, et al. Identification of ligands for RNA targets via structure-based virtual screening: HIV-1 TAR. Comput Aided Mol Design 2000; 14: 593-610.
- [40]Klebe G, Bohm HJ. Energetic and entropic factors determining binding afinity in protein-ligand complexes. Recept Signal Transduct Res. 1997; 17: 459-473.
- [43]Bohacek R, Mcmartin C, Glunz P, Rich DH. IMA Volumes in Mathematics and Its applications. Rational Drug Design. 1999;13:103-114.
- [44] Krovat E M, Langer T J. Chem Information and Computer Science.2004;44:1123-1129.
- [45] Bleicher KH, Bahm HJ, Miller K. Natural Rev Drug Discovery Alainine. 2003;2:369-378.
- [46] Wolber G, Langer T. Drug discovery and development. Computer aided drug discovery. 2004; 7:475-482.

[47] Wolber G, Langer T. Proceedings of the European Symposium on Quantitative Structure-Activity Relationships. Rational Approaches to Drug Design. 2000;10: 390-399.

IJSER