Annals of R.S.C.B., ISSN: 1583-6258, Vol. 25, Issue 6, 2021, Pages. 20763 – 20775
Received 05 September 2021; Accepted 20 October 2021.

Interaction of Protein- Ligand: Molecular Docking, A novel computational biology tool

Avinash kumar[1*], Arundhiti sharma[2], Harpreet Kaur[3], Shubham Punera[4], Pawan Tanwar[5], Palukuriyashwanth Kumar[6]

[1,4,5,6]Galgotias University, Greater Noida, India
[2,] AIIMS, New Delhi, India
[3,] Amity University, Noida, India
*avinash.kumar@galgotiasuniversity.edu.in

ABSTRACT

Computer-assisted drug discovery development (CADD) is a rapidly growing field in life sciences that has seen a number of breakthroughs in recent years. Numerous pharmaceutical companies and firms, as well as educative studies and scholastic studies employ this development investigation. The progress of CADDD was assisted by advances in genomics, proteomics, and structural knowledge. To excite the protein interactions and its simulated score for the analysis, several tools and algorithms were developed. Molecular Docking roots the computational screened approach that relies on structure and allows tiny molecules to be placed in a chosen model in a diversity of locations, conformations, and alignments.

Keywords
Investigation; Docking; Numerous; Pharmaceutical, Proteomics

Introduction

Molecular Docking is a method for shaping a stable structure by predicting the optimum route of a ligand toward a receptor (protein). Preferential alignment can be employed to forecast the probability of interactivity or binding affinity among the ligand as well as the protein employing scoring algorithms. Preferential alignment can be employed to anticipate the probability of interaction or binding affinity among the ligand as well as the protein applying scoring algorithms. Docking is a technique for predicting medication target affinity as well as activity by projecting their binding orientation towards protein target[1].

Resultingly, docking is essential during the development as well as the discovery of drug process. Computational techniques are used to reproduce the molecular identification process via computation and get an optimum configuration that reduces the system's overall free energy. The task of developing a new medication is quite difficult. Recent drug development frequently employs in- silico methods. In the drug, discovery processes the use of computer-aided approaches is increasing in popularity, recognition, and acceptability[2].

Interaction types in molecular docking

In docking, there are four different types of interactions:

- Hydrogen bond and hydrophobic interactions (Solvent-related forces)
- Van der Waals interaction (Electrodynamics forces)
- Charge- charge; Dipole-dipole along with charge-dipole (Electrostatic forces)
- Entropy (Steric forces)

COMPUTER AIDED DRUG DEVELOPMENT

Computational ability is employed in CADD to streamline the development of drug process. Chemical as well as biological understanding of ligands and/or targets aids in the discovery and

Annals of R.S.C.B., ISSN: 1583-6258, Vol. 25, Issue 6, 2021, Pages. 20763 – 20775
Received 05 September 2021; Accepted 20 October 2021.

refinement of novel medicines. Chemical compounds with undesired characteristics (activity is reduced and/or poor Absorptivity, Distributiveness, Metabolic activity, Excretion, and Toxicology, or ADMET) are screened out using in-silico scanners, and the best candidates are identified. To find prospective targeted therapies, scientists use databases of aim protein structures, such as the Protein Data Bank, to search for them[3]. Virtual sampling is intended to find new drug compounds from various chemical nanostructures by browsing databases[4].

MECHANISM OF DOCKING

Since it accurately predicts the configuration of ligands inside the target binding site, molecular docking is by far the most often utilised drug design rooted upon structure approach. Molecular docking methods estimate binding energetics and rank docked molecules according to their receptor binding affinity[5]. In molecular docking, particular scoring algorithms are used to evaluate ligand conformation, and this process is repeated until a low-energy solution is found[6].
In molecular docking there are two sections,

(i) Score algorithm
(ii) Score function

Score Algorithm

The structural properties of the ligands, namely the torsion, translation, also the rotational i.e. the degrees of freedom, change in the very first stage. To do this, conformational search algorithms use stochastic and systematic search approaches[7].

Systematic search

Small changes in the structural properties of the ligands, as well as gradual changes in their conformation, are encouraged by these techniques[8]. The program examines the conformational space's energy landscape and, after several searches or even evaluation cycles, finds the lowest energy solution that corresponds to one of the most likely binding mode[9]. This technique efficiently searches conformational space, but rather of a global least it might converge to a local minimum. This constraint can be addressed by conducting parallel searches from various places around the energy landscape[10].

Stochastic search

Transforms the structural properties of the ligands at random. It accomplishes this by generating a variety of molecular conformations as well as a diverse energy environment. The ultimate answer at the global energy minimum is investigated in this method. Due to the extreme extensive coverage of the energy landscape that this algorithm provides, the cost connected with this technique is a significant disadvantage[11]. Monte Carlo simulation, Tabu search, and Genetic algorithm are a few illustrations of stochastic search methods.
For molecular docking, both systematic and stochastic techniques are commonly employed, and each takes a distinct approach to solving a problem.

Systematic search strategies aim to uncover all of a molecule's degrees of freedom. As the degrees of freedom rise, the number of viable configurations grows tremendously. As a result of the huge number of conformations[12], systematic techniques eventually face the challenge of combinatorial explosion. Surflex, DOCK, and FRED are docking programmes that use an incremental construction approach to address the combinatorial explosion issue. The ligand is continually generated in the binding site using this method[13]. The ligand's structure is split down into numerous pieces in this approach. Furthermore, one of the fragments is utilised as an anchor

Annals of R.S.C.B., ISSN: 1583-6258, Vol. 25, Issue 6, 2021, Pages. 20763 – 20775
Received 05 September 2021; Accepted 20 October 2021.

fragment as well as docked in a binding site, as thus the remaining fragments are introduced in a sequential order until the ligand is finished. As a consequence, the conformational search is limited to the specified pieces, avoiding a combinatorial explosion[14].

The Genetic algorithm, which is an instance of stochastic search, is used in molecular docking systems like Gold and Auto-Dock. Natural selection but also biological evolution is at the heart of the GA algorithm. Various factors such as translation, rotation, as well as conformation of the ligand with regard to the target protein dictate the organization of the ligand and protein. These factors are referred to as 'state variables' since they correlate to the gene in the GA. The set of chromosomes is generated using a random search method that covers a large portion of the energy landscape[15]. This set is assessed, as well as the highly acclimated (lowest energy values) chromosomes were chosen for the following population's generation. By transferring structural features from one population to the next, this approach lowers the average energy of the chromosomal ensemble, decreasing the spatial conformation space to be searched[16]. The genetic algorithm is repeated until it converges on a conformation (chromosome) that corresponds to the global energy minimum after several conformational searches and assessment cycles[17].

Score Function

The energetics of binding of ligand-receptor complexes are estimated using scoring functions. The energy fluctuation caused by the development of the ligand-receptor structure is provided by Gibbs free energy (ΔGL) and binding constant (Kd)[18]. Physical-chemical factors including the intermolecular interactions, entropic effects, and desolvation are used to calculate the binding energy. By improving the physical-chemical parameters, the scoring function's accuracy can be improved. Incorporating a larger number of variables, however, raises the computing cost[19].

Scoring functions are of three forms: **established upon force-field, empirically, and knowledge functions rooted**.

The binding energy is calculated using a force-field-dependent scoring function that adds the inputs of non-bonded (van der Waals and electrostatic interactions) but also bonded components (dihedral variation, bond stretching, and angle bending). Employing classical mechanics equations, an ab initio functional scoring technique is employed to determine the energy corresponding with each term of the function. Its main flaw is the inaccuracy in predicting entropic contributions that is caused by the lack of a physical model to characterise entropic contributions.

The binding energy is calculated using an empirical scoring function which accounts the bonds of hydrogen, ionic and apolar interactions, and also desolvation as well as entropic effects. Multiple linear regression techniques are commonly used to fit the coefficients of the scoring function. The empirical scoring function's drawback is that it is reliant on the accuracy of the evidence used to construct the model. Empirical functions, on the other hand, are quicker as compared to the techniques relied upon force-field. The molecular docking programmesFlexX and Surflex are commonly used and employ empirical scoring[20].

This method uses doublet energy potentials due to known ligand-receptor complexes to compute a function of scoring. Instead of a binding energy, scoring functions are established upon observations of intermolecular interactions recognized from databases. The frequency of finding two distinct atoms within a specified distance in the dataset is used to generate these potentials. The total of these individual encounters determines the final score. Knowledge-based functions

Annals of R.S.C.B., ISSN: 1583-6258, Vol. 25, Issue 6, 2021, Pages. 20763 – 20775
Received 05 September 2021; Accepted 20 October 2021.

provide an excellent mix of speed as well as accuracy since they are neither relied on recreating binding affinities nor ab initio calculations[21].

.

Various types of docking

The foregoing are the most often used docking methods:

(1) The Rigid Lock and Key Docking entails keeping the receptor as well as the ligand stationary while docking takes place.

(2) Fit that has been induced. The Ligand and the receptor are conformationally flexible in Flexible Docking, either of the ligand and the receptor are conformationally flexible. The surface cell possessions as well as energy are computed for each rotation, and then the highest optimal posture is chosen[22].

Main steps incorporated in mechanics of molecular docking:-

It is the technique of investigating intermolecular interactions between two molecules in a computer simulation. A protein receptor is usually the macromolecule, whereas a ligand molecule is the micromolecule

The steps incorporated are as described below:

Preparing the Protein: Protein data bank is employed to fetch the 3D structure of the protein (PDB). Following that, the cavity's water molecules are evacuated, charges are stabilised, and missing residues are filled.

Active site prediction: It is predicted that the protein's active site is a molecule of water or a heteroatom if present.

Preparation of ligand: A range of databases including PubChem, ZINC, and the ChemSketch tool are available for retrieving ligands.

While in ligand selection, the LIPINSKY'S RULE OF 5 should be employed. The Lipinski rule of 5 helps in discerning amongst drug like and non-drug like candidates. Therefore, the chances of success become higher.

LIPINSKY'S RULE includes:

(1) Hydrogen bond donors should be less than five

(2) Hydrogen bond acceptors should be less than ten

(3) The molecular mass should be less than 500 Da

(4) The lipophilicity (expressed as LogP) should be less than 5

(5) The molar refractivity should be between 40-130.

Docking: The ligand is docked to the target protein, as well as the interaction is evaluated using a scoring function. Among the other possibilities, the best docked ligand complex is chosen[23].

Protocol for Performing Molecular Docking using Auto Dock Vina

1.1 Obtaining the Requisite Ligand as well as Target.pdb information via significant databases.

https://www.rcsb.org/

- Type the query protein
- Select Protein (e.g., AKT-1 human)

Annals of R.S.C.B., ISSN: 1583-6258, Vol. 25, Issue 6, 2021, Pages. 20763 – 20775
Received 05 September 2021; Accepted 20 October 2021.

- Select upon the download option
- Select upon the PDB file and download the file
- Make a copy of the above PDB file and rename it as protein
- Open it in WordPad
- Select all heteroatoms and delete them
- Save the file

1.2 Ligand.pdb files are retrieved from main ligand libraries.

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

- Cast around your Ligand (Tetracycline)
- Select upon the Ligand (Tetracycline)
- Select 3D image
- Open SDF
- Save 3D SDF

2 Preparing PDBQT format for Target and ligand (Target.pdbqt, Ligand.pdbqt), Grid and Docking Parameter file using Auto Dock 4.2

- Open Auto Dock present on desktop

2.1 Preparing the Target.pdbqt file

- Go to File
- Read Molecule
- Open and select Protein.pdb (*Created in first step)
- On the screen, the target molecule will display.
- Select Edit
- Select Hydrogens
- Select Add
- Select Polar Only
- Select OK
- Open the Grid
- Select on Macromolecules
- Select on Choose

Annals of R.S.C.B., ISSN: 1583-6258, Vol. 25, Issue 6, 2021, Pages. 20763 – 20775
Received 05 September 2021; Accepted 20 October 2021.

- Select the Protein
- Select the Select Molecule option
- Select OK
- Save the molecule

2.2 Preparing the Ligand.pdbqt file

- Transform the non-PDBQT file for ligand to PDBQT format by Open Babel.
- Open ligand.sdf file in Open Babel.
- Choose 3D coordinate, and Select transform.
- Choose the ligand.Pdbqt format to save at C:\Autodock.

2.3 Preparing the Grid Parameter File

- Open Grid
- Select Grid Box

*As seen below, we employed the X, Y, and Z dimensions. The X, Y, and Z centre (Center Grid Box) may also be adjusted to suit your needs.

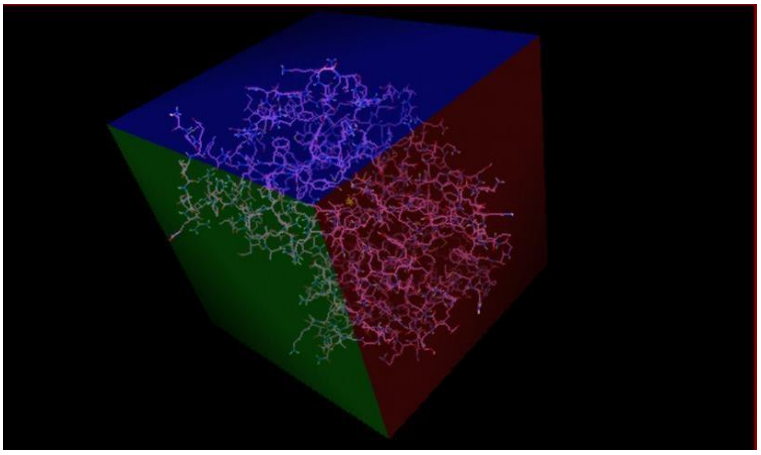


Figure 1: Centre grid box

- Select File
- Select Close saving current
- Create an Auto Dock Vina setup file consisting the PDBQT files for the ligand as well as receptor as well as the docking settings.

Annals of R.S.C.B., ISSN: 1583-6258, Vol. 25, Issue 6, 2021, Pages. 20763 – 20775
Received 05 September 2021; Accepted 20 October 2021.

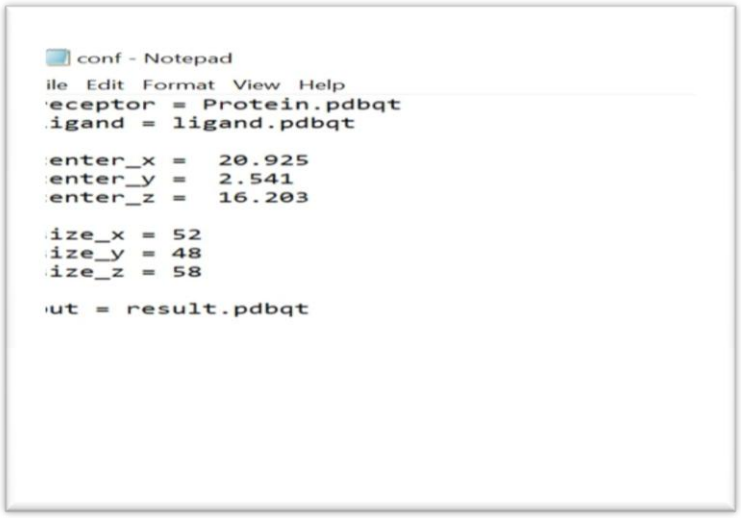


Figure 2: Conf.txt

Copy your file (protein.pdbqt, ligand.pdbqt, conf.txt) in vina folder

3 To start AutoDock Vina from the command line:

- Go to the docking folder and copy the address
- Open CMD
- Type cd (paste the address) and press enter
- Now type the command vina.exe –config conf.txt and press enter

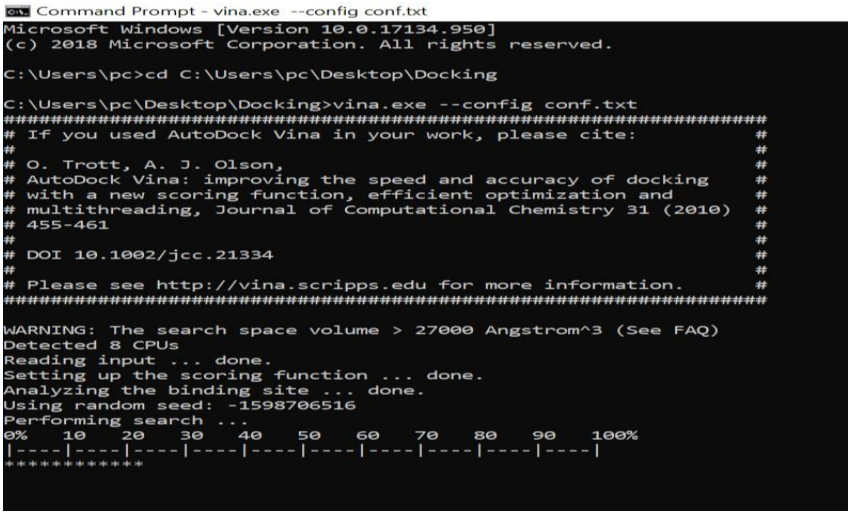


Figure 3: Command Prompt

4 Trying to retrieve the Ligand Interaction Complex.pdb as well as analysing the findings

Annals of R.S.C.B., ISSN: 1583-6258, Vol. 25, Issue 6, 2021, Pages. 20763 – 20775
Received 05 September 2021; Accepted 20 October 2021.

4.1 Analyzing Results

- Open AutoDock
- Analyze ➞ Dockings ➞ Open AutoDock vina result➞ Open Multiple molecules.
- Display ➞ Show Interactions
- file ➞ Save ➞ as image

Docking Result & Discussion

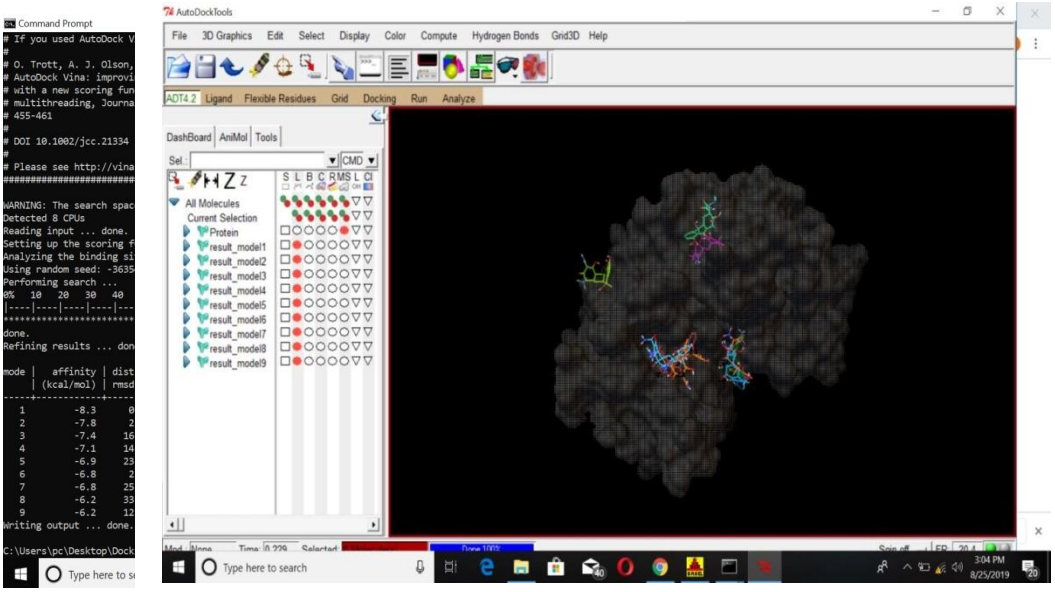
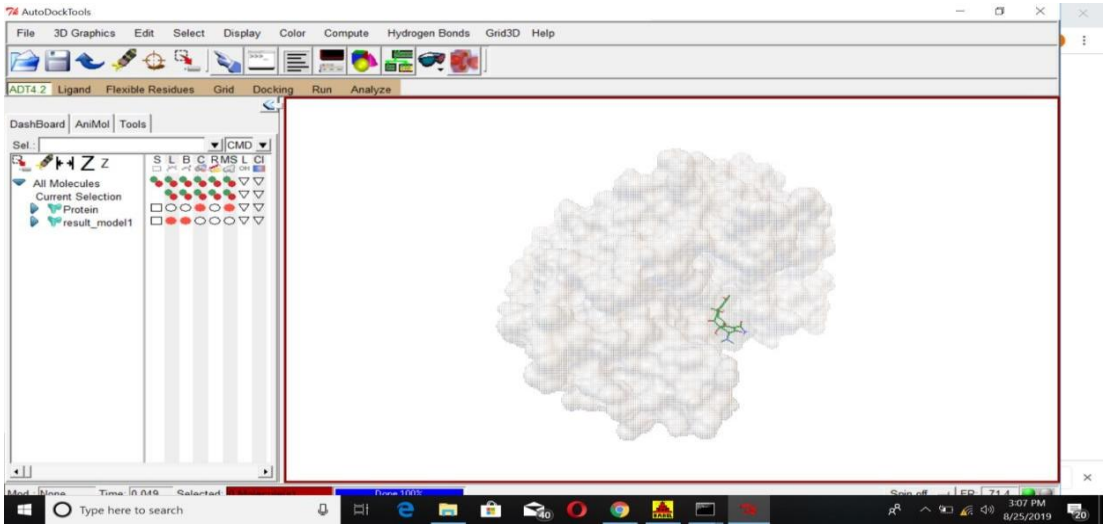


Figure 4: Docking reuslt

Interactions involved



Annals of R.S.C.B., ISSN: 1583-6258, Vol. 25, Issue 6, 2021, Pages. 20763 – 20775
Received 05 September 2021; Accepted 20 October 2021.

Figure 5: Interaction Involved

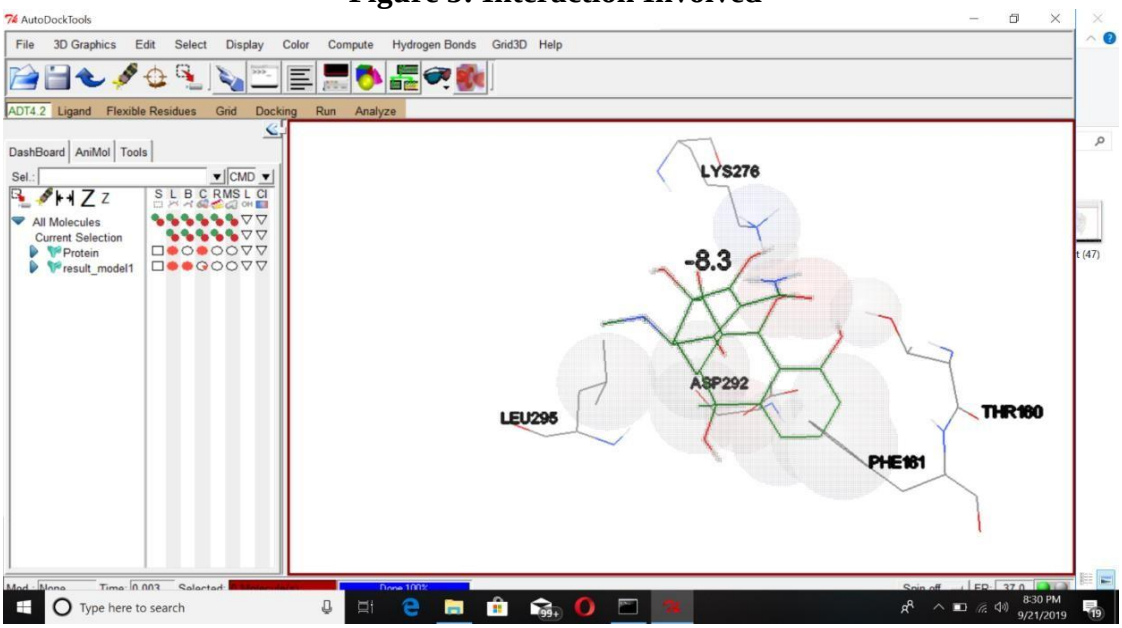


Figure 6: Interaction result

Using Auto dock Vina, the AKT-1 protein was docked with the antibiotic drug tetracycline. The highest-ranking energy value achieved when AKT-1 was docked with tetracycline was (-8.3 kcal/mol). Leucine, aspartic acid, phenylalanine, lysine, and threonine are protein residues that interact with the tetracycline (ligand).

Applications of Molecular Docking

In certain fields, molecular docking has transformed research. For example, molecular docking can be used to assess the feasibility of a biological process before it is carried out. It may also be used to forecast whether an enzyme will be inhibited or activated. Furthermore, this type of data may be utilised to create logical medication designs.

The following are some of the most common uses of molecular docking: -

Lead optimization (Drug discovery)

Molecular docking can forecast the optimal orientation of a ligand on its target protein. It aids in the prediction of various ligand binding mechanisms in the protein groove (target). This knowledge may be applied to the development of more effective and efficient medication candidates.

Hit identifications (Virtual Screening)

A molecular docking approach can be used to analyze large databases and determine the potent drugs that can specifically target a given molecule [24].

Drug-DNA interaction

Molecular docking aids in the prediction of a drug's nucleic acid binding characteristics. Medicinal chemists are attempting to unravel the underpinning anticancer mechanism by studying drug-nucleic acid interactions. This knowledge is useful in the creation of novel

Annals of R.S.C.B., ISSN: 1583-6258, Vol. 25, Issue 6, 2021, Pages. 20763 – 20775
Received 05 September 2021; Accepted 20 October 2021.

anticancer drugs as well as the structural alteration of current drugs to enhance their target selectivity.

Bioremediation

Bioremediation is the technique of using microorganisms to alleviate pollution in the environment, mostly through enzymatic activities. Docking is used to investigate protein-pollutant interactions and determine if the enzyme's active site can accept pollutant molecules [25]. The inference is predicated on the active site's nature as well as the residues' stearic hindrance. As a result, molecular docking has been effectively used in environmental remediation to comprehend the biodegradation process.

Prediction of KA (Biological activity)

On the foundation of their chemical characteristics, a validated QSAR model may be used to predict the biological activity of novel substances [28]. Furthermore, the compounds can be additionally changed to increase their action. Ordinarily, all newly developed compounds are rated as afterwards designated for experimental testing based on QSAR projections [26].

Binding site prediction (Blind docking)

Blind Docking is the process of docking a ligand to the whole surface of a protein without having any prior understanding of a particular target pocket. Until arriving at a favorable protein-ligand complex posture, blind docking requires several attempts including energy computations [29].

De-orphaning of protein

De-orphanization is the process in which ligands are identified that are highly selective for orphan targets. Advances in computational tools have revolutionized de-orphaning of target. Molecular docking calculations can be used for the prediction of the binding conformation of ligands inside the target's binding pocket [27]. Hence, it provides an essential binding information for the identification of ligands for the orphaned targets [30].

Discussion and Conclusion

For drug analysis as well as design molecular docking seems to be a very helpful tool. The structural libraries as well as visualization of molecules are simple to use and have proven to be an important part of the pharmaceutical outcomes. The basic user interface continues to be expanded upon in commercialized software programs. New commercial and research algorithms are swiftly incorporated into the large products. In terms of performance, public domain software is increasingly more dependable but also skillful of going head-to-head with commercial software. Every year and a half, the computers almost double rate of change increases, while visual projections grow progressively complex as well as intuitive. Molecular docking is becoming an essential part of drug development as a result of all of these causes.

References

[1] Jorgensen, W.L. The many roles of computation in drug discovery. Science, 2004, 303, 1813-1818.

[2] Bajorath, J. Integration of virtual and high-throughput screening. Nat. Rev. Drug Discov., 2002, 1, 882-894.

[3] Walters, W.P.; Stahl, M.T.; Murcko, M.A. Virtual screening - an overview. Drug Discov. Today, 1998, 3, 160-178

[4] Langer, T.; Hoffmann, R.D. Virtual screening: an effective tool for lead structure discovery? Curr. Pharm. Des., 2001, 7, 509-527.

[5] 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, 935-949.

[6] Gohlke, H.; Klebe, G. Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors. Angew. Chem. Int. Ed. Engl., 2002, 41, 2644-2676.

[7] Moitessier, N.; Englebienne, P.; Lee, D.; Lawandi, J.; Corbeil, C.R. Towards the development of universal, fast and highly accurate docking/scoring methods: a long way to go. Br. J. Pharmacol., 2008, 153(Suppl 1), S7-26.

[8] Shoichet, B.K.; McGovern, S.L.; Wei, B.; Irwin, J.J. In: Hits, leads and artifacts from virtual and high throughput screening., Molecular Informatics: Confronting Complexity, 2002.

[9] Bailey, D.; Brown, D. High-throughput chemistry and structure based design: survival of the smartest. Drug Discov. Today, 2001, 6, 57-59.

[10] Kuntz, I.D.; Blaney, J.M.; Oatley, S.J.; Langridge, R.; Ferrin, T.E. A geometric approach to macromolecule-ligand interactions. J. Mol. Biol., 1982, 161, 269-288.

[11] 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.

[12] Coupez, B.; Lewis, R.A. Docking and scoring--theoretically easy, practically impossible? Curr. Med. Chem., 2006, 13, 2995-3003.

[13] Kontoyianni, M.; Madhav, P.; Suchanek, E.; Seibel, W. Theoretical and practical considerations in virtual screening: a beaten field? Curr. Med. Chem., 2008, 15, 107-116.

[14] Brooijmans, N.; Kuntz, I.D. Molecular recognition and docking algorithms. Annu. Rev. Biophys. Biomol. Struct., 2003, 32, 335- 373.

[15] ten Brink, T.; Exner, T.E. Influence of protonation, tautomeric, and stereoisomeric states on protein-ligand docking results. J. Chem. Inf. Model., 2009, 49, 1535-1546.

[16] Cross, J.B.; Thompson, D.C.; Rai, B.K.; Baber, J.C.; Fan, K.Y.; Hu, Y.; Humblet, C. Comparison of several molecular docking programs: pose prediction and virtual screening accuracy. J. Chem. Inf. Model., 2009, 49, 1455-1474.

[17] Li, X.; Li, Y.; Cheng, T.; Liu, Z.; Wang, R. Evaluation of the performance of four molecular docking programs on a diverse set of protein-ligand complexes. J. Comput. Chem., 2010, 31, 2109- 2125.

[18] Plewczynski, D.; Lazniewski, M.; Augustyniak, R.; Ginalski, K. Can we trust docking results? Evaluation of seven commonly used programs on PDBbind database. J. Comput. Chem., 2010 doi: 10.1002/jcc.21643.

[19] McConkey, B.J.; Sobolev, V.; Edelman, M. The performance of current methods in ligand-protein docking. Curr. Sci., 2002, 83, 845-855.

[20] Goodford, P.J. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. J. Med. Chem., 1985, 28, 849-857.

[21] Kastenholz, M.A.; Pastor, M.; Cruciani, G.; Haaksma, E.E.; Fox, T. GRID/CPCA: a new computational tool to design selective ligands. J. Med. Chem., 2000, 43, 3033-3044.

[22] Levitt, D.G.; Banaszak, L.J. POCKET: a computer graphics method for identifying and displaying protein cavities and their surrounding amino acids. J. Mol. Graph., 1992, 10, 229-234.

[23] Laskowski, R.A. SURFNET: a program for visualizing molecular surfaces, cavities, and intermolecular interactions. J. Mol. Graph., 1995, 13, 323-330, 307-328.

[24] Glaser, F.; Morris, R.J.; Najmanovich, R.J.; Laskowski, R.A.; Thornton, J.M. A method for localizing ligand binding pockets in protein structures. Proteins, 2006, 62, 479-488.

[25] Brady, G.P. Jr.; Stouten, P.F. Fast prediction and visualization of protein binding pockets with PASS. J. Comput. Aided Mol. Des., 2000, 14, 383-401.

[26] Mezei, M. A new method for mapping macromolecular topography. J. Mol. Graph. Model., 2003, 21, 463-472.

[27] Fischer, E. Einfluss der configuration auf die wirkungderenzyme. Ber. Dt. Chem. Ges., 1894, 27, 2985-2993.

[28] Koshland, D.E. Jr. Correlation of structure and function in enzyme action. Science, 1963, 142, 1533-1541.

[29] Hammes, G.G. Multiple conformational changes in enzyme catalysis. Biochemistry, 2002, 41, 8221-8228.

[30] Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. A fast flexible docking method using an incremental construction algorithm. J. Mol. Biol., 1996, 261, 470-489.