Review Article

Molecular Docking Simulation with Special Reference to Flexible Docking Approach

Tahmeena Khan[1]*, Alfred J. Lawrence[2], Iqbal Azad[1], Saman Raza[2], and Abdul Rahman Khan[1]

[1]*Department of Chemistry, Integral University, India*
[2]*Department of Chemistry, Isabella Thoburn College, India*

Copyright

Abstract

Handling flexibility in molecular docking is a major challenge in chemical biology research. It has been observed that in most of the cases incorrect binding energies and poses are predicted by the docking algorithms. Neglecting receptor flexibility results in incorrect ligand binding scores. Conformational rearrangements of the receptor binding pocket while predicting binding pose and binding score is crucial for improving structure-based drug design and virtual ligand screening techniques, focusing on the algorithms and their rational. Direct modeling of protein binding site flexibility is a big challenge due to large conformational space that is to be sampled, and difficulties remain in predicting the accurate energy function. The review is an attempt to summarize the different approaches used in flexible docking analysis leading to correct predictions.

INTRODUCTION

Molecular docking

Most of the cellular processes require protein-protein interaction. Accurate prediction of three-dimensional structure of complexes as an outcome of protein-protein interaction may shed light on their functional mechanisms and their roles in the cell [1]. Molecular docking studies help in understanding the interaction of drug molecule with biological target molecules. The study involves binding of the ligand (suggested molecule) with the preferred binding site (active site) of the target protein or DNA which is referred to as receptor. The interaction takes place in a covalent manner leading to potential specificity [2]. Crucial molecular mechanisms, ligand binding modes and factors effecting ligand-receptor complex can be studied through the docking results. The results can predict the binding energy as related to stability of complexes. The binding energy can be predicted beforehand using the molecular docking.

Theory of Molecular Docking

Docking studies are performed to determine the interaction between two molecules and to find out the best orientation of ligand which would form a complex with overall minimum energy. Scoring functions are analyzed which converts interacting energy into docking score. The interacting energy is also calculated. The 3D pose of the bound ligand can be visualized using tools like Pymol, Rasmol etc.

Types of Interactions

The different interactions studied between the ligand and protein are-

1. Electrostatic forces: These forces arise due to presence of charge. Charge-charge, dipole-dipole and charge-dipole are common electrostatic interactions.

2. Electrodynamic forces: Van der Waals' interactions are the most common types of Electrodynamic forces.

3. Steric forces: Steric forces are generated due to close proximity of molecules and affect the reactivity and the chemical reactivity.

4. Solvent-related forces: These forces are the outcome of a chemical reaction between solvent and protein or ligand.

MOLECULAR DOCKING ALGORITHMS

The docking algorithms help in determination of all possible optimal conformations for a said complex. Binding energy can also be calculated for the resultant complex. The common algorithms used for docking analysis are-

1. Molecular dynamics

2. Monte Carlo methods

3. Generic algorithm

4. Fragment-based methods

⊘SciMedCentral

5. Point complementary methods

6. Distance geometry methods

7. Systematic searches

TYPES OF ENERGIES EVALUATED

A molecular docking program uses scoring functions to estimate the binding energetics of the predicted ligand-receptor complexes. The energy variation, due to the formation of the ligand-receptor structure, is given by the binding constant (Kd) and the Gibbs free energy (ΔGL) [3]. The primary aim of molecular docking is to have the stable ligand-receptor complex which has an optimized conformation with least binding free energy. The net binding energy (ΔGbind) is calculated in terms of hydrogen bond (ΔGHbond), electrostatic (ΔGelec), torsional free energy (ΔGtor), dispersion and repulsion (ΔGvdw), desolvation (ΔGdesolv), total internal energy (ΔGtotal) and unbound system's energy (ΔGunb) [4]. The software usually predicts the scoring functions to estimate the binding energetics of a ligand-receptor complex. The energy is given as the binding constant (K_d) and in terms of Gibbs' free energy (ΔGL) [3]. Intermolecular interactions, desolvation and entropic effects are some of the important factors considered while predicting the ligand-receptor binding .Greater is the number of physico-chemical parameters evaluated; greater is the accuracy of the scoring function. Optimized conformation is obtained by molecular docking between target and pre-defined ligand's conformation in a specific groove. The recognition of the most likely binding conformation requires-

1. Exploration of a large conformational space which represents potential binding modes

2. Prediction of the interaction energy associated with predicted binding conformations [5].

SOME EFFICIENT MOLECULAR DOCKING SOFTWARES

AutoDock is automated software used for prediction of interactions of ligands with bio macromolecular targets. The software provides the minimum energy between the substrate and the target protein, exploring all available degrees of freedom (DOF) for the system. AutoDock uses the Lamarckian Genetic algorithm and empirical scoring function. It provides reproducible docking results for ligands with approximately 10 flexible bonds. Docking is usually carried out using one of the several search methods out of which Lamarckian genetic algorithm (LGA) is most efficient. Traditional genetic algorithm and simulated annealing are also used sometimes. AutoDock is run several times to give several docked conformations and analysis of the predicted energy and the consistency of results are combined to identify the best solution (Figure 1).

AutoDock uses a semi-emperical free energy force field to evaluate conformations during docking simulations. The force field evaluates binding in two steps. In the first step, the ligand and protein start in an unbound conformation. In the second step, evaluation of intermolecular energetics of combining the ligand and protein in their bound conformation is made. The force field includes six pair-wise evaluations (V) and an estimate of the conformational entropy lost upon binding (Figure 2).
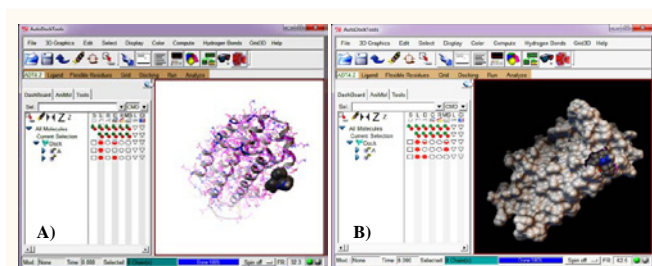
Figure 1 AutoDock windows showing docking interactions.

$$\Delta G = (V^{L\text{-}L}_{bound} - V^{L\text{-}L}_{unbound}) + (V^{P\text{-}P}_{bound} - V^{P\text{-}P}_{unbound}) + (V^{P\text{-}L}_{bound} - V^{P\text{-}L}_{unbound} + \Delta S_{conf})$$

Figure 2 Force field equation showing six pair wise evaluations: where L refers to the "ligand" and P refers to the "protein" in a ligand-protein docking calculation.

Analysis in AutoDock can be divided into following categories, (1) Initializing molecules (2) Running Auto Grid (3) Running AutoDock (4) Analyzing Interaction energy.

iGEMDOCK

iGEMDOCK is a graphical-automated software for integration docking, screening and post-analysis. To do docking in the software, protein structure file and ligand file has to be prepared. The protein structure file can be obtained from Protein Data Bank (PDB) (http://www.rscb.org/). The ligand files are available from ZINC (http://zinc.docking.org/) or PubChem (http://www.ncbi.nlm.nih.gov/sites/entrez?db=pccompound). The input format of receptor structure for iGEMDOCK is PDB format. GEMDOCK computes a ligand conformation and orientation relative to the binding site of target protein based on generic evolutionary method (GA) (Figure 3,4). Therefore, the GA parameters are directly related to the docking performance. After generating a set of poses, iGEMDOCK recalculates the energy of each pose. iGEMDOCK recalculates the energy of each pose The interaction 25 data includes summarized energy and individual energy terms. Fitness is the total energy of a predicted pose in the binding site. The empirical scoring function of iGEMDOCK is estimated as:

Fitness = vdW + Hbond + Elec

Here, the vdW term represents van der Waal energy whereas Hbond and Elect terms are hydrogen bonding and electro statistic energies, respectively. The ligand can be docked with the binding site of each PDB files using accurate docking function (slow docking). Finally, the post analysis tool visualizes and ranks the compound to be screened by merging the pharmacological interactions and energy-based scoring function (Figure 5-7).

BASIC PROCESSES INVOLVED IN DOCKING

Target Selection or protein preparation

Selection of a target is the first step in molecular docking. Target selection includes identification of binding sites. If the 3D structure of the target protein is known, its binding potential with the target can be identified. The target preparation involves removal of solvent, addition of hydrogen atoms. Preparation of
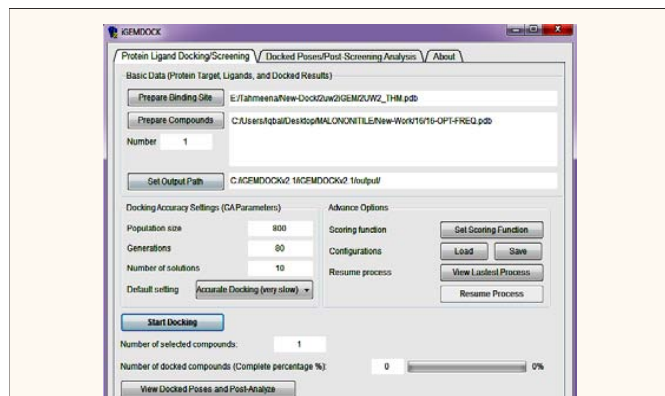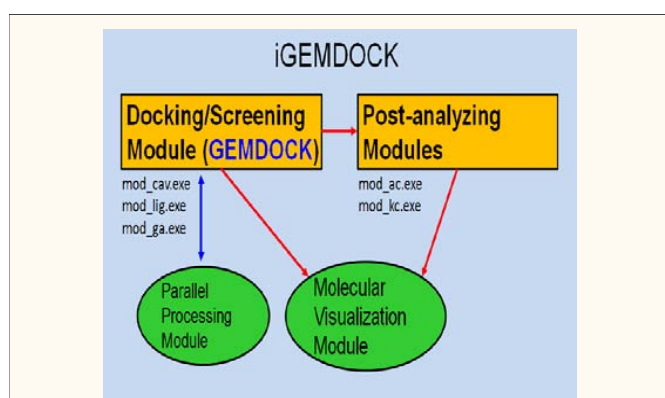
Figure 3 iGEMDOCK window.

Figure 4 iGEMDOCK functions.

protein generally effects the refining of the final results of the virtual screening.

Active Site Prediction

The active sites in the target protein can be identified using CASTp (Computed Atlas of Surface Topography of proteins) server. The server identifies all the feasible binding pockets in the protein structure. The server identifies area and volume of each probable binding cavity [6]. After the receptor is built, the active site should be identified. From the active site water molecules and heteroatom's should be removed.

Structural Cleaning and Energy Minimization of Receptor

Protein cleaning is done to insert missing atoms in missing residues, removing water molecules, bound DNA and standardization of atoms etc. The PDB files of target proteins can be downloaded from RCB. The structure can be visualized in Discovery studio. Receptor energy minimization can be carried out by using default constraint of 0.3Å RMSD (root mean square) and Assisted Model Building with Energy Refinement (AMBER) force field 14SB using Chimera. Energy calculations are made after structural cleaning and removing structural inconsistencies. Minimization routine can be performed by Molecular Modeling Toolkit (MMTK) which is included with Chimera [7].

Ligand preparation and energy minimization

The 2D structure of the ligand can be prepared in Chem Draw professional. The SMILES & MOL file of the ligand was prepared in Chem Draw 3D. Biovia Discovery Studio version 2017 R2 is usually used to prepare SDF and PDB files of the ligand. The PDB file can be transported to Gaussian View and Gaussian 9W for energy minimization. A database of ligands is constructed prior to ligand selection. The database consists of experimentally known data. The commonly used databases are as follows: 1) NCI has more than 25,000 compounds. The database is usually used in industry. 2) ZINC is a free database with more than 35 million compounds' structures. 3) MDDR includes more than 120,000 drug candidates. The database also includes information on biological activity. 4) ADC, contains all the data of the marketed compounds and chemicals. For ideal selection of ligand, Lipinsky's rule of 5 should be applied [8].

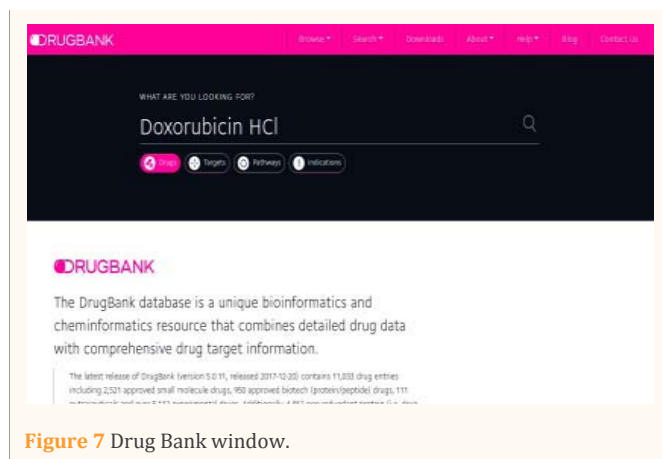## TYPES OF DOCKING

Rigid ligand and rigid receptor docking

Only three translational and rotational degrees of freedom

Figure 5 PDB window.

Figure 6 Pubchem window.

Khan et al. (2018)
Email: tahminakhan30@yahoo.com

⊘SciMedCentral

Figure 7 Drug Bank window.

are possible in this type of docking methodology. With the help of a pre-computed set of ligand, its' flexibility can be addressed. DOCK and FTDOCK have been commonly used for rigid docking [9].

Flexible ligand and rigid receptor docking

Ligand and receptor are found to change their conformations in order to obtain a minimum energy perfect fit complex. To take into account the flexibilities of the ligand and receptors since they tend to modify their conformations. Though the methodology is a bit expensive, yet softwares like AutoDock and FlexX can be used for flexible docking.

Flexible Docking

In bound state proteins undergo conformational changes including backbone and side-chain movements. Ignoring flexibility could prevent docking algorithms from recovering native associations. Accounting for flexibility is also essential for the accuracy of the solutions [10]. The protein flexibility analysis methods can be classified into three major groups-

1. For generation of discrete conformations. The different conformations can be assigned by analyzing experimentally deduced protein structures or by using Molecular Dynamics (MD) simulation.

2. For determination of a continuous protein conformational space. Many flexible docking methods sample this pre-calculated conformational space in order to generate a set of discrete conformations including Normal Modes Analysis (NMA) and Essential Dynamics.

3. For identifying rigid and flexible regions in the protein including the rigidity theory and hinge detection algorithms.

Challenges and requirements in flexible approach: Though ligand flexibility has been dealt with by a variety of algorithms, receptor flexibility is still a challenge. Direct modeling of protein movements associated with binding site flexibility represents a major glitch because of the twin challenges of high dimensionality of the conformational space and of the complexity of energy function. A typical ligand binding site for a drug-like molecule may have ten to twenty amino acid-side chains

involving much potentially rotatable torsion. This number can be larger than the number of degrees of freedom for the ligand (up to 6-12). The backbone movements may worsen the situation, as each backbone movement affects multiple side-chains. Thus, fully flexible receptor/ligand docking simulation involves sampling of an order of magnitude higher number of degrees of freedom than typical rigid-receptor/flexible ligand simulations routinely used in current structure-based virtual screening processes. Side-chain flexibility alone may or may not be sufficient for adequate modeling. Conformational variation in the HIV protease binding site is well described in terms of movements of several side-chains and a water molecule [11]. On the contrary, many kinases exhibit loop rearrangements delimiting the active site [12]. Diversity in ligand binding mechanisms and the frequent unpredictability of receptor movement types makes the use of pre-determined (by experimental or computational means) multiple receptor conformations (MRC) an interesting area to work upon. Approaches used for the purpose. To deal with side chain flexibility is easier and possible for small ligands [13]. A 'minimum rotation hypothesis' was proposed by Zavodsky and Kuhn. Their docking algorithm, SLIDE, attempts to resolve ligand–receptor steric clashes by a minimal number of side-chain rotations, with the cost of side-chain movement evaluated as a product of the rotation angle and the number of atoms moved

One of the easiest ways to include multiple conformations of receptor in a docking experiment is to run multiple independent simulations. Integration of MRC sampling into the docking algorithm may also offer advantages in terms of calculation speed as well as helps in data management. Such 'ensemble docking' extensions of original rigid-receptor algorithms have been reported for AUTODOCK [13] or ICM [14]. Extension of the popular FlexX algorithm, FlexE not only utilizes MRC individually, but also extends the search space beyond the input set of conformations by detecting distinct dissimilar parts and joining them combinatorially [15]. New potentially accessible receptor conformations are thus generated during the search. However, consideration of too many conformations can lead to reduced performance. In a recent critical evaluation of FlexE on two targets of pharmaceutical interest, b-secretase and JNK-3, the algorithm was unable to handle large loop movements and could not match enrichment factors obtained by running multiple independent FlexX runs on each receptor structure [16]. FLIP Dock is another algorithm using the AutoDock force field that introduces a highly sophisticated data structure for the MRC representation, termed Flexibility Tree (FT) [17]. A hierarchical and multi resolution description of the pocket structure and flexibility provides a framework for incorporating various types of flexibility into AutoDock.

FITTED algorithm, a recently developed algorithm allows two receptor flexibility modes [18]. The first mode, termed 'semi-flexible', is essentially an MRC ensemble docking. The second 'fully flexible' mode allows genetic algorithm (GA) to generate different combinations of side-chain rotamers and backbone conformations found in the input ensemble. In addition, the algorithm is capable of simulating replaceable interface water molecules by a combination of special functional form for water interaction and sampling absence/presence of waters in GA. Ensemble methods may offer significant advantages over

SciMedCentral

sequential docking to multiple conformations by conventional rigid-receptor algorithms. The efficiency of ensemble methods should depend on the diversity of the receptor conformations, if the ensemble only involves minor structural variations, its exploration may contribute only additively to the overall computational cost; however, if highly dissimilar binding site conformations are included, each of them will have to be explored virtually independently, potentially multiplying the search time by the number of conformations. Post-docking optimization may help to further improve both docking pose and its score. Nabuurs, Wagener, and de Vlieg demonstrated a robust performance of a combination of FlexX-Ensemble docking combined with a post-docking explicit receptor ligand optimization on a benchmark of 35 ligand–receptor complexes [19]. Post-docking optimization may help to further improve both docking pose and its score. Nabuurs, Wagener, and de Vlieg demonstrated a robust performance of a combination of FlexX-Ensemble docking combined with a post-docking explicit receptor ligand optimization on a benchmark of 35 ligand–receptor complexes [19]. The authors used a combination of FlexX-Ensemble docking with the Yasara/Yamber2 program for conformational generation and full atom refinement of the high-ranking complexes. The 'flexible' residues were pre-selected using a set of rules. The protocol was tested on 20 cross docking ligand–protein pairs

CONCLUSION

The principles and methods in this review highlight the strategies by which flexible docking can be applied in the identification of novel bioactive compounds. With the exceptional growth in the number of protein structures in PDB, improved understanding of the flexible docking approach through MCR method has made it an interesting approach to explore. This MRC approach is comparatively less time taking and still suitable for virtual ligand screening as long as the number of fixed receptor conformations is relatively small and carefully chosen. Still the challenge remains for issues involving the accuracy of the available scoring functions. Handling of solvent effects, entropic effects and dealing with receptor flexibility are some of the important issues that need more exploration. Successful molecular docking protocol requires clear use of fundamental methods. Understanding these principles is necessary for obtaining meaningful results.

ACKNOWLEDGEMENTS

REFERENCES

1. Han K, Zhang L, Wang M, Zhang R, Wang C, Zhang C. Molecules. 2018; 23: 2303.

2. Rohs R, Bloch I, Sklenar H, Shakked Z. Molecular flexibility in ab initio drug docking to DNA: binding-site and binding-mode transitions in all-atom Monte Carlo simulations. Nucleic Acids Res. 2005; 33: 7048-7057.

3. Foloppe N, Hubbard R. Towards predictive ligand design with free-energy based computational methods? Curr Med Chem. 2006; 13: 3583-3608.

4. Agarwal S, Chadha D, Mehrotra R. Molecular modeling and spectroscopic studies of semustine binding with DNA and its comparison with lomustine-DNA adduct formation. J Biomol Str and Dyn. 2015; 33: 1653-1668.

5. Kapetanovic IM. Computer-aided drug discovery and development (CADDD): In silico-chemico-biological approach. Chem Biol Interac. 2008; 171: 165-176.

6. Paramashivam SK, Elayaperumal K, Natarajan BB, Ramamoorthy MD, Balasubramanium S, Dhiraviam KN. *In silico* pharmacokinetic and molecular docking studies of small molecules derived from Indigoferaaspalathoides Vahl targeting receptor tyrosine kinases. Bioinformation. 2015; 11: 73-84.

7. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem. 2004; 25: 1605-1612.

8. Chao J, Dai Y, Varpoote R, Lam W, Cheng YC, Pao LH, et al. Major achievements of evidence-based traditional Chinese medicine in treating major diseases. Biochem Pharmol. 2017; 139: 94-104.

9. Meng XY, Zhang HX, Mezei M, Cui M. Molecular docking: a powerful approach for structure-based drug discovery. Curr Comput Aided Drug Des. 2011; 7: 146-157.

10. Andrusier N, Mashiach E, Nussinov R, Wolfson HJ. Principles of flexible protein-protein docking. Proteins. 2008; 73: 271-289.

11. Osterberg F, Morris GM, Sanner MF, Olson AJ, Goodsell DS. Automated docking to multiple target structures: incorporation of protein mobility and structural water heterogeneity in AutoDock. Proteins. 2002; 46: 34-40.

12. Cavasotto CN, Abagyan RA. Protein flexibility in ligand docking and virtual screening to protein kinases. J Mol Biol. 2004; 337: 209-225.

13. Zavodszky MI, Kuhn LA. Side-chain flexibility in protein-ligand binding: the minimal rotation hypothesis. Protein Sci. 2005; 14: 1104-1114.

14. Schapira M, Abagyan R, Totrov M. Nuclear hormone receptor targeted virtual screening. J Med Chem. 2003; 46: 3045-3059.

15. Claussen H, Buning C, Rarey M, Lengauer T. FlexE: efficient molecular docking considering protein structure variations. J Mol Biol. 2001; 308: 377-395.

16. Polgar T, Keseru GM. Ensemble docking into flexible active sites. Critical evaluation of FlexE against JNK-3 and beta-secretase. J Chem Inf Model. 2006; 46: 1795-1805.

17. Zhao Y, Sanner MF. FLIP Dock: docking flexible ligands into flexible receptors. Proteins. 2007; 68: 726-737.

18. Corbeil CR, Englebienne P, Moitessier N. Docking ligands into flexible and solvated macromolecules. 1. Development and validation of FITTED 1.0. J Chem Inf Model. 2007; 47: 435-449.

19. Nabuurs SB, Wagener M, de Vlieg J. A flexible approach to induced fit docking. J Med Chem. 2007; 50: 6507-6518.