Journal of Innovations in Pharmaceuticals and Biological Sciences
www.jipbs.com
JIPBS
e-ISSN: 2349-2759 p-ISSN: 2395- 1095

Review article

In-silico expectations of pharmaceutical industry to design of new drug molecules

Ram Babu Tripathi*, Milind Pande, Gaurav Garg, Deepanshu Sharma

Bioinformatics Lab, Department of Pharmacy, Mangalayatan University, Beswan, Aligarh -202145, Uttar Pradesh, India

Abstract

Increasing costs of drug development and reduced number of new chemical entities have been a growing concern for new drug development in recent years. A number of potential reasons for this outcome have been considered. One of them is a general perception that applied sciences have not kept pace with the advances of basic sciences. Therefore, there is a need for the use of alternative tools to get answers on efficacy and safety, with more certainty and at lower cost. One such alternative tool is the in silico drug design or the computer aided drug design (CADD). In Silico drug designing is a form of computer-based modeling whose technologies are applied in drug discovery processes. This approach has given tremendous opportunity of pharmaceutical industry to identify many new potential drugs than the conventional approaches. It emphasizes on how we can develop better and competitive drugs with the use of software and wet lab synchronization and the hope of developing better tools to facilitate human life the comfort and disease competitive.

Key words: Drug Designing, *In Silico*, Molecular Modeling, Drug Discovery.

***Corresponding Author: Ram Babu Tripathi,** Bioinformatics Lab, Department of Pharmacy, Mangalayatan University, Beswan, Aligarh -202145, Uttar Pradesh, India.

1. Introduction

Drugs are essential for the prevention and treatment of disease. Human life is constantly threatened by many diseases. Therefore, ideal drugs are always in great demand. To meet the challenges of ideal drugs, an efficient method of drug development is demanding. But the process of drug design, development and commercialization is a tedious, time-consuming and cost-intensive process [1]. To fulfil these challenges, several multidisciplinary approaches are required for the process of drug development; collectively these approaches would form the basis of *In-Silico* approach in drug design [2].

CADD were established in the early 1970s with the use of structural biology to modify the biological activity of insulin and to guide the synthesis of human haemoglobin ligands. At that time, X-ray crystallography was expensive and time-consuming, rendering it infeasible for large-scale screening in industrial

laboratories [3]. Over the years, new technologies such as comparative modeling based on natural structural homologues have emerged and began to be exploited in lead design [4]. These, together with advances in combinatorial chemistry, high-throughput screening technologies and computational infrastructures, have rapidly bridged the gap between theoretical modeling and medicinal chemistry. Numerous successes of designed drugs were reported, including Dorzolamide for the treatment of cystoid macular edema, Zanamivir for therapeutic or prophylactic treatment of influenza infection, Sildenafil for the treatment of male erectile dysfunction, and Amprenavir for the treatment of HIV.
In Silico approach can be classified mainly in two different categories viz. a) Structure based drug design (SBDD) and b) Ligand based drug design (LBDD)[5].

2. Structure based drug design (SBDD)

Structure-based drug design (or direct drug design) relies on knowledge of the three dimensional structure of the biological target obtained through methods such as x-ray crystallography or NMR spectroscopy. If an experimental structure of a target is not available, it may be possible to create a homology model of the target based on the experimental structure of a related protein. Using the structure of the biological target, candidate drugs that are predicted to bind with high affinity and selectivity to the target may be designed using interactive graphics and the intuition of a medicinal chemist. Alternatively, various automated computational procedures may be used to suggest new drug candidates.
In structure, based drug design for a particular target has been developed on the basis of known structural information of the drug target like receptor structure (mostly protein). If the structure of receptor is not available, the receptor structure can be predicted by homology modeling. Homology modeling usually refers to as comparative modeling in which on the basis of known amino acid sequences of a protein, a model of protein can be constructed and the structure is comparable with the 3D-structure of similar homologous protein (template) [6].

Docking

Molecular docking is one of the *in-silco* methods (computational technique) to study the configuration of intermolecular complexes of one smaller molecule (ligands or drug) with a larger molecule (receptor or enzyme), and a certain score (usually referred to as 'docking score') has been given to each orientation a ligand docked in the active site. This score can then be used to evaluate the potential of ligand-protein affinity, which ultimately leads to prediction of biological effectiveness of a ligand against the particular protein [7].

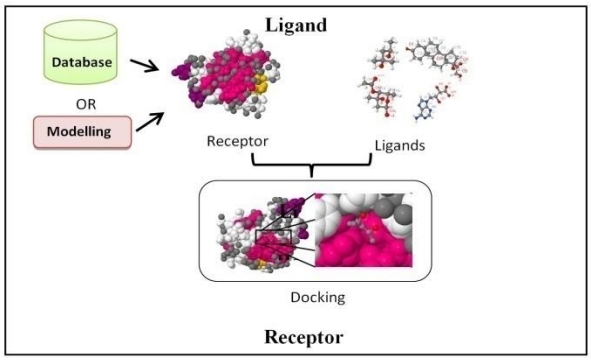

Figure 1. Molecular Docking

Classification of Docking

On the basis of flexibility of proteins and ligands, the molecular docking can be classified in the following categories [8-9].

a) ***Rigid body docking*** in which both receptor and ligand are rigid.

b) ***Flexible ligand docking*** in which receptor is rigid while ligand is flexible.
c) ***Flexible docking*** in which both receptor and ligand are flexible and it is most commonly used docking method.

In-silco Tools Available to Performed Docking Experiment

There are a number of computational programs available to do this such as M-ZDOCK, AutoDock, GLIDE, GOLD, and SLIDE etc. Each of these programs has distinct advantages. The main issue of these programs address is what conformations and orientations of ligands screened are most likely to bind to a specific receptor. An important factor in determining a program's success is its ability to duplicate the experimentally determined interactions between a protein and its ligand. A prediction is generally found acceptable if the RMSD between the docked ligand and the experimentally determined ligand in under 2À.

Different algorithms used in the process of docking are Monte Carlo, genetic algorithm, fragment-based, molecular dynamics etc and on the basis of these algorithms, different programs were developed (free and commercial purpose)[10-11].

Steps Involved in Docking Experiment:

Protein data bank (PDB) files may have a variety of problems that need to be corrected before they can be used in AutoDock. These potential problems include missing atoms, added waters; remove water, more than one molecule, and chain breaks etc. [12-13].

Table 1. Computational Docking Tools

S. NO.	Docking Software	Designer/Company	License term	Supported platforms	Docking approach	Scoring function
1	*AutoDock*	Scripps research institute	Free for academic use	Unix, Mac, window	Lamarckian genetic algorithms	Force field method
2	*DOCK*	I. Kuntz university of California San Francisco	Free for academic use	Unix, Mac, window	Shape fitting	Chem Score, salvation scoring
3	*FlexX*	T. Lengauer and M. RareyBioSolve IT	Commercial free for evaluation (6 week)	Unix, Linux, Windows	Incremental construction	FlexX Score
4	*FRED*	OpenEye Scientific Software	Free for academic use	Unix, Linux, Windows	Shape fitting (Gaussian)	Screen score
5	*Glide*	Schrödinger Inc.	Commercial	Unix, Linux	Monte Carlo Sampling	Glide Score
6	*GOLD*	Cambridge Crystallographic Data center	Commercial free for evaluation (2 month)	Unix, Linux, Windows	Genetic algorithm	Gold score Chem score User defined
7	*Ligand Fit*	Accelrys	Commercial	Linux, IBM	Monte Carlo Sampling	Lig Score

a) Ligand Preparation

The ligand must be prepared before starting the docking experiment, like energy minimization of ligand, protonation state of ligand etc. The protonation of ligand depends upon the pH of receptor environment.

b) **Receptor Preparation**

Structures of protein/biomolecule evaluated by X-ray crystallographic technique are available on protein data bank and it could easily be downloaded in text format from their website [http://www.rcsb.org]. The selected chains of the biomolecules were prepared for docking by using following steps:

- Adding Gasteiger charges
- Adding polar hydrogen atoms
- Checking whether total charge per residue is integer
- Choosing flexible residues
- Removal of water and/or ions molecules as per requirement
- Minimizing the receptor, if necessary

c) Receptor grid generation

From the prepared protein/biomolecule, the co-crystallized ligand was separated from its active site. The active site is generally represented as an enclosing box at the centroid of work space ligand. Following this protocol, a grid centered on the ligand was generated using the default settings of desired software. All ligands were docked into this grid structure.

Docking and scoring

On a defined receptor grid, flexible docking was performed using appropriate module of desired software. The module analysis the protein ligand interaction on the basis of different interactions between them like vander waals, hydrogen bonding and electrostatic interactions.

Validation of Docking Methods

Validation is the process by which we can predict the reliability of docking method. A number of validation techniques can be used to simulate the predicative ability of docking experiment.

a) **Alignment Method**

In this method the docked structure of the molecules is superimposed on the reference molecule by using software's like *MMP™*, *Field Align™*

b) **Chemical reasonableness**

It is another approach of validation technique in which the amino acids, which show the binding interaction with ligand are mutated and further predict their binding affinity.

Current Challenges in Molecular Docking

Much work has been invested in the making of better docking programs and scoring functions over the past years and, although much progress has been made but enhancement of docking program is still necessary [14-15].

a) Docking into Flexible Receptors

The most challenging problems in docking and scoring is the treatment of flexible receptors. Numerous examples have become known where the same protein adopts different conformations depending on nature of ligands. In flexible docking both ligand and receptor are considered flexible. However, there are still some limitations such as only side chains are set flexible while backbone is rigid in nature.

b) **Water Interaction**

Water molecules frequently play a key role in drug-receptor interactions; if one ignores water-mediated interactions during docking then the calculated

interaction energy of a given ligand conformation may be too low. It is notoriously difficult to treat water adequately, as first one need to identify possible positions for water molecules where they could interact with the protein and ligand, and subsequently one must be able to predict whether a water molecule is indeed present at that position.

c) **Tautomers formation of ligands**

Another challenge of docking is the formation of various tautomeric and protomeric states of the molecules that can adopt during drug-receptor interaction. A molecule such as acids or amines are stored in their neutral forms but they are ionized under physiological conditions means it is necessary to ionize them earlier to docking experiment. One approach to this would be to generate all possible forms, subsequently to dock all of them, and to choose the relevant form based on the scores. However, it remains to be seen whether such an approach would be beneficial or just generate a large number of tautomers.

3. Ligand based drug design (LBDD)

Ligand-based drug design (or indirect drug design) relies on knowledge of other molecules that bind to the biological target of interest . This approach is particularly useful when 3D structure of the receptor is not available and it relies on the knowledge of ligands that bind to the desired target[16]. The most prominent techniques used in this approach is quantitative structure activity relationships (QSAR).

In quantitative structure activity relationship (QSAR), a correlation between experimentally determined biological activity and calculated properties of molecules is derived. These exhibit a particular squared predictive correlation coefficient (r^2); and model with r^2 value close to 1 will be designated as best model. These QSAR models relationships can then be used to predict the activity of new analogous.

Basic Requirements for QSAR Analysis

Some basic requirements are essential for the development of best QSAR model to predict the biological activity [17]. Out of which some of them are mentioned below

- All analogues belong to a congeneric series (classical QSAR studies) exerting the same mechanism of action
- The set of compounds have same mechanism of action
- Biological response should be distributed over a wide range
- Biological activity should be in specific units (concentration in molar units or IC_{50} or percentage inhibition).

Approaches in QSAR Studies

There are different widely used approaches in QSAR studies. Following are the commonly used ones[18].

- **Hansch analysis (linear free energy relationship or extra thermodynamic approach)**

It is one of the most promising approaches to the quantification of the interaction of drug molecules with biological system given by Corwin Hanschin 1969. It is also known as linear free energy (LFER) or extra thermodynamic method, which assumes additive effect of various substituents inelectronic, steric, hydrophobic, and dispersion data in the non-covalent interaction of a drug and macromolecules. Hansch analysis relates the biological activity within a homologous series of compounds to a set of theoretical molecular parameters, which describe essential properties of the drug molecules. Hansch

proposed that the action of a drug is depending on two processes.

- Journey from point of entry in the body to the site of action which involves passage of series of membranes and therefore it is related to partition coefficient log P (lipophilic) and can be explained by random walk theory.
- Interaction with the receptor site depends on,
  a) Bulk of substituent groups (steric)
  b) Electron density on attachment group (electronic)

This approach was originally coined as Linear Free Energy Relationships (LFER) and later changed, more appropriately, to extra thermodynamic approach and expressed by the following equation.

$$\log 1/C = a \log P + b(\log P)^2 + c$$

Where, a and b are the coefficients of the *log P* and *(log P)²* terms, respectively, and c is a constant term.

- **Free and Wilson analysis**

The Free-Wilson approach is truly a structure-activity based methodology because it incorporates the contribution made by various structural fragments to the overall biological activity. Indicator variables are used to denote the presence or absence of a particular structural feature. It is represented by equation

$$BA = \Sigma\, ai\, xi + \mu$$

Where, BA is the biological activity, μ is the overall activity, ai is the contribution of each structural feature, xi denotes the presence (xi = 1) or absence (xi = 0) of particular structural fragment.

- **Quantum mechanical methods**

The information provided by QM is more accurate than Free and Wilson analysis therefore more robust QSAR models and/or QSPR models are expected with QM descriptors. Partial charges are the most common descriptors in QSAR/QSPR models due to their simplicity and informative content.

The QSAR is based on structure activity relation (SAR) approach. It uses physicochemical properties (parameters) to represent drug properties that are believed to have a major influence on drug action. Some of the common pharmacophoric features include hydrophobic, aromatic, hydrogen bond acceptor, hydrogen bond donor, positive ionizable, and negative ionizable groups. These parameters are properties that are capable of being represented by a numerical value which are used to produce a general equation correlating activity with relevant physicochemical properties.

Parameters or Descriptors

Descriptors can be defined as a numerical representation of chemical information encoded within a molecular structure via mathematical procedure. Descriptor can be classified in following categories [19-20].

a) Lipophilic parameters

Partition coefficient (P) and the lipophilic substituent Constant (p) are the two most important lipophilic parameter use in QSAR analysis. The former parameter refers to the whole molecule whilst the latter is related to substituent groups. A drug has to pass through a number of biological membranes in order to reach its site of action. Partition coefficients were the obvious parameter to use as a measure of the movement of the drug through these membranes.

c) Electronic parameters

The distribution of the electrons in a drug molecule will have an influence the activity

of a drug. When the the drug reaches its target site the distribution of electrons in its structure will control the type of bonds it forms with that target, which in turn affects its biological activity. In other words, the electron distribution in a drug molecule will have an effect on how strongly that drug binds to its target site, which in turn affects its activity. The distribution of electrons within a molecule depends on the nature of the electron withdrawing and donating groups found in that structure.

d) Steric parameters

Drug to bind effectively to its target site the dimensions of the pharmacophore of the drug must be complementary to those of the target site. The Taft steric parameter (Es) was the first attempt to show the relationship between a measurable parameter related to the shape and size (bulk) of a drug and the dimensions of the target site and a drug's activity. This has been followed by Charton's steric parameter, Verloop's steric parameters and the molar refractivity (MR), amongst others. The most used of these additional parameters is probably the molar refractivity. However, in all cases the required parameter is calculated for a set of related analogues and correlated with their activity using a suitable statistical method such as regression analysis.

Advantage of QSAR Studies

- Refinement of synthetic targets
- Reduction or replacement of animal tests, thus reducing animal use
- To predict the biological activities of untested and sometimes yet unavailable compounds
- To developed new model for a biological systems
- To optimized the existing leads so as to improve their biological activities
- QSAR models act virtual screening tools for predicting ADME and toxicity studies
- To elucidate the phenomena and nature of drugs and receptor interactions

Pitfalls in QSAR Studies

Despite the fact that number of successful QSAR application, there are several pitfalls in their proper application like [21]

a) Multi-conditionality

Drug action is based on a sequence of complicated physiochemical events (delivery, targeting, metabolism, and excretion) that are either still unknown or not fully understood on a molecular level. For this reason and because of hardware and software limitations *in silico* studies can only fragmentally reproduce real world observations. QSAR and QSPR are used to describe quantitatively ADMET processes in living cells, e.g. protein binding (plasma enzymes etc.).

b) Common Mode of Action and Multiple Binding Modes

An important prerequisite of QSAR is the use of a series of congeners with a common target structure. Chemical similarity is not a guaranty for a common action mechanism of all congeners. A complication is the occurrence of multiple binding modes (MBM) of the very same ligand to its target molecule. QSAR is conducted under the silent assumption that no MBM is present when comparing molecular similarities with ligand binding analysis (LBA) or protein binding analysis (PBA) techniques.

c) Multiple Targets and Multi-potency

QSAR mainly work with cell-free data is not affected by drug binding to multiple targets. Such multi-potencies occur *in vivo* when a molecule in lower doses binds to a biomolecule with higher affinity, while in higher doses the same ligand may bind to other targets with lower affinity.

d) Prodrug Function

The molecules considered in QSAR studies are not necessarily the ones responsible for the biological response as in case of Prodrug.

e) Over- and Under-Determined Equations

In QSAR studies, over fitting occurs if too many independent variables relative to the number of data points are included in a regression equation. In such cases, regression equations tend to fit the "noise" or errors in the data and, in general, do not yield robust predictions.

Conclusion

Various compounds have occupied researchers in recent years and numerous computational models have been drawn up. Many of these models have been generated by means of Ligand-based approaches, mainly QSAR studies. Such models were capable to predict a potent drug for new drug discovery. Most of these models have has also been successfully applied to the design of new ligands or to the optimization of known active compounds. It is generally recognized that drug discovery and development are very time and resources consuming processes. There is an ever growing effort to apply computational power to the combined chemical and biological space in order to streamline drug discovery, design, development and optimization. In biomedical arena, computer-aided or in silico design is being utilized to expedite and facilitate hit identification, hit-to-lead selection, optimize the absorption, distribution, metabolism, excretion and toxicity profile and avoid safety issues. The development of any potential drug begins with years of scientific study to determine the biochemistry behind a disease, for which pharmaceutical intervention is possible. The result is the determination of specific receptors (targets). In the post genomic era, computer-aided drug design (CADD) has considerably extended its range of applications, spanning almost all stages in the drug discovery pipeline, from target identification to lead discovery, from lead optimization to preclinical or clinical trials.

Acknowledgements

We gratefully acknowledge the financial support from the Department of Biotechnology (DBT) government of India to work on different drug deigning softwares, our principal, Department faculty, for the source of inspiration in the review of the paper done. We hope to have near advancements in this field of study to make our community better in health care and disease competitive. We are also thankful to the library of university to provide us the full help in studying different aspects of drug discovery.

References

1. Kuhlman J.: Drug research: from the idea to the product. Int. J. Clinical Pharm. and Therp. 1997; 35:541.
2. Fuller JC, Burgoyne NJ, Jackson RM.: Predicting druggable binding sites at the protein-protein interface. Drug Discovery Today 2009; 14:155-161.
3. Debouck C, Metcalf B: The impact of genomics on drug discovery. Annual Rev Pharmacology and toxicology 2004; 40:193-208.
4. Burbaum J, Tobal G M: Proteomics in drug discovery. Current opinion chem. Bio. 2002; 6:427-433.
5. Gatto, JG: The changing face of bioinformatics. Drug discovery today 2003; 8:375-376.
6. Greer J, Erickson WJ, Baldwin JJ, Varney MD.: Application of the three dimensional structures of protein target molecules in structure-based drug design. J. Med. Chem. 1994; 37:1035-1054.
7. Lengauer, T. and Rarey, M.: Computational methods for biomolecular docking.

Current Opinion in Structural Biology 1996; 6:402-406.
8. Halperin, I., Wolfson, H. and Nussinov, R.: Principles of docking: An overview of search algorithms and a guide to scoring functions. Proteins 2002; 47:409-443.
9. Wang, R., Lu, Y. Fang, X. and Wang, S.: An Extensive Test of 14 Scoring Functions Using the PDB bind Refined Set of 800 Protein-ligand Complexes. Journal of Chemical Information and Modeling 2004; 44:2114-2125.
10. Kontoyianni, M., McClellan, L.M. and Sokol, G.S.: Evaluation of docking performance: comparative data on docking algorithms. Journal of Medicinal Chemistry 2004; 47:558-65.
11. Mizuno, C. S., Chittiboyina, A. G. Shah, F. H. Patny, A. Kurtz, T. W. Speth, R. C. and Karamyan, V. T.: "Design, synthesis, and docking studies of novel benzimidazoles for the treatment of metabolic syndrome. Journal of Medicinal Chemistry 2010; 53:1076-1085.
12. Nazreen, S., Alam ,M.S Hamid, H. Yar, M.S. Shafi, S. Dhulap, A. Alam, P. and Pasha, M.A.: Design, synthesis, in silico molecular docking and biological evaluation of novel oxadiazole based thiazolidine-2, 4-diones bis-heterocycles as PPAR-γ agonists. European Journal of Medicinal Chemistry 2014; 16:175-185.
13. Teague, S.J.: Implications of protein flexibility for drug discovery. Nature Reviews Drug Discovery 2003; 2:527–541.
14. Murray, C.W., Baxter, C.A. and Frenkel, A.D: The sensitivity of the results of molecular docking to induced fit effects: application to thrombin, and neuraminidase. Journal of Computer-Aided Molecular Design 1999; 13:547–562.
15. Hansch, C., Maloney, P.P. and Fujita, T.: Correlation of biological activity of phenoxyacetic acids with Hammett substituent constants and partition coefficients. Nature 1962; 194:178–180.
16. Gupta, S.P., "QSAR Studies on Enzyme Inhibitors", Chemical Reviews; 1987,87: 1183-1253.
17. Free, S.M. and Wilson, J.W.A.: Mathematical contribution to structure–activity studies. Journal of Medicinal Chemistry 1964; 7:395–399.
18. Fujita, T., Iwasa, J. and Hansch C: A new substituent constant, π, derived from partition coefficients. Journal of American Chemical Society 1964; 86:5175–5180.
19. Yasri, A. and Hartsough, D. J.: Toward an optimal procedure for variable selection and QSAR model building. J Chem Inf Comput Sci. 2001; 41:1218-1227.
20. Roy, P.P., Paul, S. Indrani, M. and Roy, K.: On two novel parameters for validation of predictive QSAR models. Molecules 2009; 14:1660-1701.