Computational Drug Designing and Development: An insight

Rajnish Kumar, Anju Sharma, Rajesh Kumar Tiwari*

Amity Institute of Biotechnology, Amity University Uttar Pradesh, Lucknow Campus, Lucknow-226028, Uttar Pradesh, India

---***---

Abstract - *Computational developments have enabled researchers and scientists to switch to rational drug designing. It has significantly reduced the attrition rate and total time taken in drug development. Development of robust direct and indirect computational methods such as De novo drug design and Quantitative structure activity relationship respectively have enhanced the success rate of ligands to become drug. Current review is focused on the computational drug design methods and their effect on drug design and development.*

Key Words: Drug, Drug designing, De novo, Ligand, QSAR, Receptor

1. INTRODUCTION

Computational drug designing and development can roughly follow two different approaches; direct drug designing e.g. De novo drug design and indirect drug designing e.g. Quantitative structure activity relationship (QSAR). Further, drug designing can be broadly classified into two types; Target based and physiology based. The main difference between these two classes depends upon the time at which the drug target is identified. Physiology-based approach follows physiological indications, e.g. the study and detection of a disease phenotype in cell-based assay or an animal model. Compounds are profiled and screened based on physiological readout. A purely physiology-based approach initially ignores the target identification and validation. The approach directly moves into screening process. Depending upon the pharmacological properties of lead molecules, target identification and mechanism of action is derived later on. Whereas, in case of target based approach, the drug designing initiates with target identification, validation and derivation of its role in disease. Thousands of pathogen and human genes and their products make it an extremely difficult task. The revolution in genomics has played a crucial role over last twenty years.

Direct drug designing is also known as structure based drug design and the indirect drug designing is also referred as ligand based drug designing. Direct drug designing is based on information of target/receptor. As this approach is dependent on the structure of target, it is also named as Structure based drug designing. De novo ligand designing is classical example of structure based drug designing. It is used as an alternative method to the screening experiment especially when lead structure is not available. The major challenge with this approach is to design ligands which can be easily synthesizable. In generally, fragment-based approach is preferred to design ligands that can be synthesized easily in laboratories in comparison to other available connection methods of De novo drug design. The main step in this approach is to dock a library of smaller ligands into the active site of target/receptor to find out best fit orientation between the two molecules [1].

Indirect drug design is based on the information of ligand. This approach of drug designing can produce a model of biological target on the basis of information binding ligand to it. Pharmacophore modelling comes under indirect drug designing where basic knowledge of ligands/drug candidates is available.

2. COMPUTATIONAL DRUG DESIGN APPROACHES

2.1 De novo Drug Design

De novo means from the beginning. This method of drug designing fall into one of three categories: (a) Methods that analyze the active site, (b) Methods that dock whole molecule, (c) Methods that connect molecular fragments or atoms together to produce a ligand.

Methods that analyze active site are not considered as true De novo drug design. However, it is an important prerequisite for De novo drug design. Active site defines the probable function of protein. While prediction of active site of a protein, two assumptions are made. One, protein structure is already modeled and second the three dimensional structure of protein is known. Some of the famous methods for prediction of active site of proteins include geometrical method, physicochemical approach and machine learning approach. Geometrical method includes, probe and ball method and Arc method etc. Probe and ball method measures the volume of active site. The arc method measures the slop and depth of active site geometrically.

The physicochemical approach includes secondary structure composition, active residues, phylogenetic analysis etc. According to the secondary structure composition theory, it has been observed that most of the coiled regions are present on surface in comparison to helixes. Protein active site is present on surface. Therefore, active site is present on coiled region than helix.

There are certain amino acids which are very frequently present on active sites. Such amino acids are Aspartic acid, Arginine, Hystidene, Glutamine, Serine, Cystene, Lysine etc. Mapping of such amino acids on a protein may give an idea of active site. In phylogenetic analysis, it has been observed that how two functions are similar to each other. It is most frequently used method. Five to six residues of active site are known to carry out multiple sequence alignment and quantify how their function is similar to each other. This method is also known as evolutionary trace method. Machine learning approaches use Support Vector Machine (SVM) and Artificial Neural Network (ANN) to predict the active site of a protein.

De novo uses whole molecular docking as a central step for designing the active ligands. It is the process by which best match between the two molecules is derived [2-3]. A drug has to have the affinity as well as intrinsic activity to be an agonist. Affinity is inherent property of the ligands. However, intrinsic activity is dependent upon the fact that how well the two molecules are docked. Depending upon the size of two molecules being docked, the docking can be divided into two types; micro-molecular docking (e.g. ligand-protein docking, ligand-DNA docking) and macromolecular docking (e.g. Protein –protein docking). Available computational tools for docking are Dock, GOLD, HEX, PRO_LEADS, AutoDock etc [4-7].

Connection methods are truly de novo ligand design methods. These methods are further divided into four types [8];

a) Site-point connection methods: It determines desirable places of individual atoms in the active site and then place suitable fragments/ functional groups at those locations. Programs used for site point connection method are CLIX and LUDI.

b) Fragment connection methods: These methods start with previously positioned fragments and linkers are used to connect those fragments i.e. individual fragments that are selected in different ways are connected. Methods that use fragment connection method are NEWLEAD, HOOK, CAVEAT, PROLIGAND etc.

c) Sequential buildup methods: This method sequentially constructs a ligand fragment-by-fragment. In this construction each new piece may be added anywhere on the existing ligand and need not to be linear. Programs that use sequential buildup method are GROW, GROWMOL etc.

d) Random connection methods: A special class of connection methods which is amalgamation of site point connection method, fragment connection method and sequential buildup method. Softwares available under this method are MCDNLG, CONCEPTS, CONCERTS, LigBuilder etc.

2.2 Quantitative Structure Activity Relationship (QSAR)

Biological activity is function of its molecular structure. This makes the principle of QSAR. Further, the similarity principle states that set of compounds will typically display an understandable structure-activity relationship. It has probably led down the foundation of principle of QSAR. The QSAR attempts to find out the relationship between activity and structure in the form of mathematical model. It is very difficult to find out direct relation of a single property with molecular structure but structural factors known as descriptors and which have influence on molecular property can be identified. In other words, descriptors can be considered as connecting link between molecular structure and molecular property (figure 1).

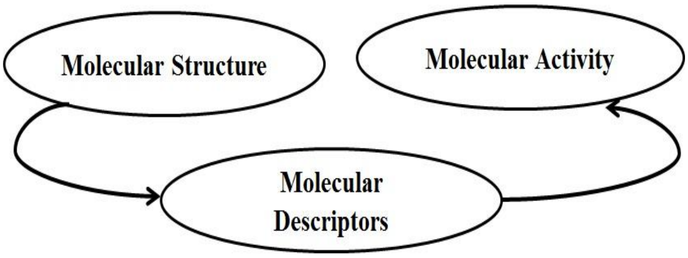

Figure 1: Relationship between molecular structure and property

The basic equation of QSAR can be considered as follows:

Biological activity = (Molecular structure)

QSAR equations are nothing but molecular properties expressed in the form of function of molecular descriptors. These QSAR equations are different from each other in the molecular property used in correlation, descriptors used and mathematical expression used by them. Example:

Cell permeability = (descriptor set 1)
Toxicity = (descriptor set 2)

There are two most common parameters that are correlated to molecular activity; electronic and lipophilicity i.e. "σ" is electronic parameter (Hammett equation), "π" is lipophilicity parameter (developed specifically for QSAR by Hansch) [9-10]. These two parameters are not exclusive. Various other parameters are also tested but σ and π have wide acceptance.
A typical QSAR equation is:

$$\log\left(\frac{1}{C}\right) = -K_1(\log P)^2 + K_2 \log P + K_3\Pi + K_4 MR + K_5 E_s + K_6$$

Where, C is the concentration needed to carry out the desired effect, logP and (logP)2 terms are used to show the parabolic relationship of lipophilicty and activity, K1, K2, K3, K4, K5, K6 are all regression coefficients. Figure 2 shows the simplified steps to perform traditional QSAR.

First and foremost step to perform the traditional QSAR is compound selection. The selected compounds should be diverse enough. They should be selected on following parameters while selecting compounds.

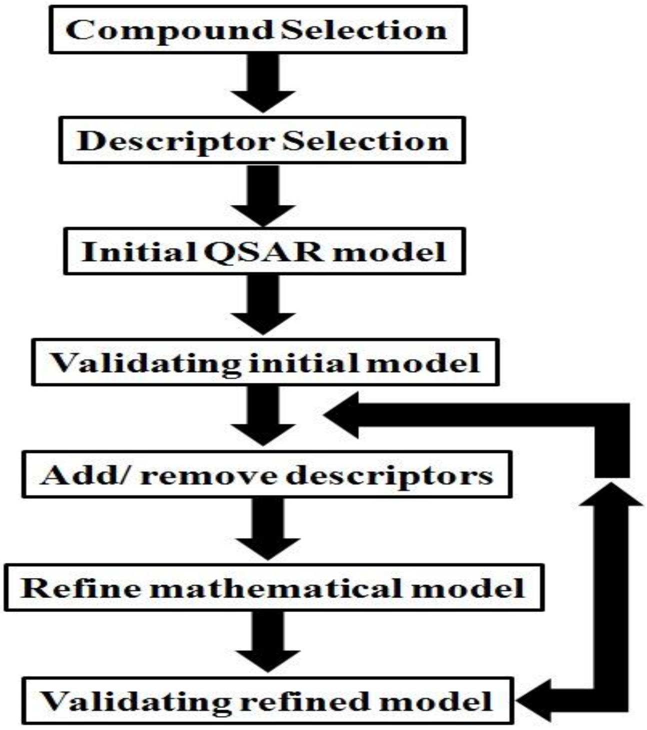

Figure 2: Flowchart of traditional QSAR

(a) Compounds may belong to congeneric series or may have structural diversity even within a chemical class.

(b) Compounds collected as dataset should have same mechanism of action.

(c) Compounds should bind to the same target.

(d) Initially outliers should not be considered in the dataset.

(e) Compounds should encompass a large range of descriptor values which are relevant to biological activity.

Descriptor selection is an important and vital step. One needs to identify descriptors or physicochemical properties which directly influence the biological property under study. Once the selection of descriptors is done then the next step is calculation of values of descriptors for all the collected compounds. The values of descriptors may either be collected from experimental outcomes or theoretical approaches. For theoretical calculation of descriptors, the dataset is subjected to variety of descriptor/feature calculating tools to generate as many as possible theoretical descriptors for each and every compound in the data set. Various softwares are known which calculates wide range of theoretical descriptors like OASIS [11], CODESSA [12], DRAGON [13], etc.

After the calculation of descriptors, compounds are randomly divided into training and test set. For large dataset the division is done by 3:1 rule (i.e. if total 200 compounds are present in data set, out of these 150 random compounds will be separated as training set and rest 50 compounds in test set). For smaller dataset then division is done on 1:1 basis (i.e. half randomly selected compounds will be separated to constitute training and rest half will make the test set) [14]. Scaling of descriptors is again a vital step. It makes the QSAR computationally less expensive at the same time lower values are not overshadowed by higher values. In order to scale the descriptors min-max normalization may be used. Another method for scaling of descriptor is using Standard deviation. Calculate standard deviation and mean for each collected descriptor. Scaling is done using following equation:

Next step is to select the most relevant descriptors. There are two types of methods available for this process. One, manual and second automated. Manual selection requires complete domain knowledge. In short thorough understanding of the structure-activity relationship is required which is exploited to generate analyses. Automated method uses computational algorithms such as forward selection, backward elimination, Stepwise regression etc [15-16].

The starting point for deriving the equation is the study table. It consists of a spread sheet of the molecules with values of biological activity and descriptors down the column. Generally the first column contains the molecular identification (e.g. compound name, 2D structure). The second column contains activity value and the subsequent columns contain corresponding values of the descriptors.

Study table leads to the graphical analysis. This step is of extreme importance and leaves space for “hunches” and primary interpretations. The most obvious trends in system are identified and correlation process starts. The initial analysis guides to the right mathematical equation which contains information about the behavior of the system and allows its interpretation.

Once the equation is established, the validation of the generated equation is done. There is large number of methods available for validating developed QSAR model [17]. Standard deviation is the easiest way to validate the developed QSAR model. Another method of validation is Correlation Index (r2). It measures the degree of correlation between the activity values calculated by model and those measured experimentally coefficient i.e. it tries to find out the trend between experimental and calculated values. Next method to validate a generated QSAR model is T-test for single descriptors and Significance of r2. Correlation index (r2) alone is not sufficient to determine whether the relationship has occurred by chance; its significance of r2 can be calculated using t-statistic for single descriptor. Another validation method for QSAR model is F-test. It is an extension of t-test. The only change is that the quantity depicting number of descriptors is added to formula [18].

3. CONCLUSION

De novo drug design and QSAR are the two most extensively used computational drug designing methods. These two methods have significantly added the rationalization in the traditional hit and trial method of drug designing. With rapid computational advancements and continuously improving algorithms, it seems to assist the drug designing more efficiently in near future.

Conflict of interest: None

REFERENCES

[1] Schneider, G., Fechner, U. (2005). Computer-based de novo design of drug-like molecules. Nat Rev Drug Discov. 4(8), 649-63.

[2] Halperin, I., Ma, B., Wolfson, H., Nussinov, R. (2002). Principles of Docking: An Overviewof Search Algorithms and a Guide to Scoring Functions. Proteins: Structure, Function, and Genetics 47:409–443.

[3] Mihasan, M. (2012). What in silico molecular docking can do for the 'bench-working biologists'. J. Biosci. 37, 1089–1095.

[4] Baxter, C. A., Murry, C. W., Clark, D. E, Westhead, D. R., Eldridge, M. D. (1998). Flexible docking using tabu search and an empirical estimate of binding affinity. Proteins: Structure, Function, and Bioinformatics. 33 (3), 367-382.

[5] Bouvier, G., Evrard-Todeschi, N., Girault, J. P., Bertho, G. (2010). Automatic clustering of docking poses in virtual screening process using self-organizing map. Bioinformatics 26 (1), 53-60.

[6] Gardiner, E. J., Willett, P., Artymiuk, P.J. (2001). Protein docking using a genetic algorithm. Proteins. 44 (1), 44-56.

[7] Lorenzen, S., Zhang, Y. (2007). Monte Carlo refinement of rigid-body protein docking structures with backbone displacement and side-chain optimization. Protein Sci. 16 (12), 2716–2725.

[8] Rotstein, S. H., Murcko, M. A. (1993). GroupBuild: A Fragment-Based Method for De Novo Drug Design. J. Med. Chem. 36, 1700-1710.

[9] Hansch, C., Fujita, T. (1964). Port Analysis. A method for the correlation of biological activity and chemical structure. J. Am. Chem. Soc., 86, 1616-1626.

[10] Leo, A., Hansch, C., Elkins, D. (1971). Partition Coefficients and Their Uses. Chem. Rev., 71, 525-616.

[11] MEkenyan, O., Bonchev, D. (1986). OASIS method for predicting biological activity of chemical compounds. Acta Pharm Jugosl 36, 225

[12] Katritzky, A.R., Lobanov, V.S. (1994) CODESSA: Version 5.3, University of Florid, Gainesville.

[13] Todschini, R., Consonni, V., Mauri, A., Pavan, M. (2006). DRAGON-Software for the calculation of molecular descriptors. Ver 5.4 for Windows, Talete srl, Milan, Italy.

[14] Kumar R., Sharma A., Tiwari R.K. (2014) Introduction to drug designing and development, Nova Science Publishers, USA pp. 1-292

[15] Karelson, M. (2000). Molecular Descriptors in QSAR/ QSPR. Wiley-InterScience, New York.

[16] Tideschini, R. Consonni, V. (2000). Handbook of Molecular Descriptors, Wiley-VCH, Weinhein, Germany.

[17] Hecker, E., Kubinyi, H., Bresch Eine neue Gruppe von , H (1995). Burger´s Medicinal Chemstry and Drug Discovery, 5th Edition, Vol. I, Principles and Practice, pp. 497-571, Wolff, M. E., Ed., John Wiley & Sons, New York.

[18] Kubinyi, H. (1997) QSAR and 3D QSAR in drug design. Part 1. Methodology, Drug Discovery Today. 2 (11), 457-467