Available online on 15.05.2019 at http://jddtonline.info

Journal of Drug Delivery and Therapeutics

Open Access to Pharmaceutical and Medical Research

Open Access Review Article

Quantitative Structure Activity Relationships in Computer Aided Drug Design: A Review

Divya Niranjan[1]*, Simran Gill[1], Deepali Shrivastava[1], Aakash Jain[1], Om Prakash Agrawal[2]

[1]School of Pharmacy, ITM University, Jhansi Rd, Gwalior, Madhya Pradesh-475001

[2]School of Pharmacy, Madhyanchal Professional University, Ratibad, Bhopal-462044

ABSTRACT

The process of processing a drug and reaching to the market consists of various steps. In the early times of Drug Discovery, researchers faced with little or no Structure Activity Relationships information regarding any chemical moiety. Computer Aided Drug Designing (CADD) is a discipline allowing various aspects of research to merge together and stimulate each other. CADD acts as a tunnel in Drug Discovery and accelerates finding new lead compounds. The theoretical basis of CADD involves quantum mechanics and molecular modelling studies like Structure-based design, Ligand-based design, database searching and binding affinity. QSAR is structural descriptors of chemical compound to its biological activity. It is very important to find out relationships between molecular structure and useful properties and so Drug Discovery and Development get more complex. But automation of chemical synthesis and pharmacological screening provides a vast amount of experimental data.

Keywords: QSAR, Drug Design, CADD, Descriptor.

Article Info: Received 27 March 2019; Review Completed 03 May 2019; Accepted 08 May 2019; Available online 15 May 2019

Cite this article as:

Niranjan D, Gill S, Shrivastava D, Jain A, Agrawal OP, Quantitative Structure Activity Relationships in Computer Aided Drug Design: A Review, Journal of Drug Delivery and Therapeutics. 2019; 9(3):645-649

http://dx.doi.org/10.22270/jddt.v9i3.2662

***Address for Correspondence:**

Divya Niranjan, School of Pharmacy, ITM University, Jhansi Rd, Gwalior, Madhya Pradesh-475001

INTRODUCTION

The pipeline of drug discovery from idea to market consists of seven steps*viz.* disease selection, target selection, lead compound identification and leadoptimization, preclinical trial testing, clinical trial testing and pharmacological optimization. Compounds for testing are naturally obtained from plants, animals and microorganisms. These compounds can be rejected if found absence or low activity, existence of toxicity or carcinogenicity, complexity of synthesis, insufficient efficiency etc[1]. As a result, only about 1/100000 investigated compounds are introduced in market.

The process of Drug discovery and developing a new medicine is long, complex, risky and highly risky process. This is why, Computer Aided Drug Design (CADD) being widely used in Pharmaceutical Industry to accelerate this process. On an average, it takes 10-15 years and US$ 500-800 million for introducing a new drug into market in which synthesis and testing of lead analogues are large contributors[2].Computational tools are much easier to applyin hit to lead optimization to cover a wider range and whereby reducing the number of compounds must be synthesised and tested *in vitro*.The computational optimization involves structural based analysis of docking poses and energy profiles for hit analogues;ligand based screening, prediction of favourable affinity or optimizes drug metabolism, excretion and potential of toxicity[3]. The lower cost of CADD with respect to chemical synthesis and biological characteristics of compounds make more attractive to focus, reduce and diversify the chemical space[4]. However, most of the molecular discoveries are the results of an iterative in three – phase cycle of design, synthesis and test. Fig. 1 shows the steps in Drug discovery.

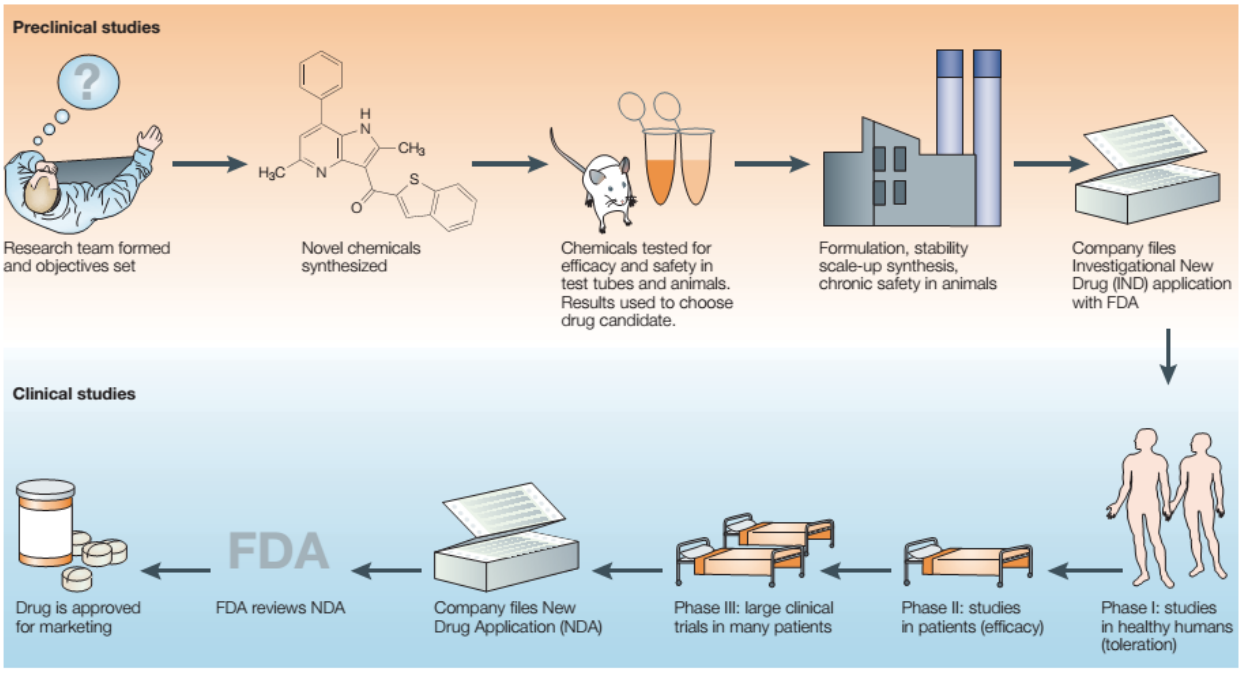

Fig. 1 Stages of Drug Discovery

After analysing, the results from one interaction provide information and knowledge that gives initiation to the next cycle of discovery. The analysis stage has the common feature of construction of some models enabling the observed activity to be related to molecular structure[5]. These models are called as Quantitative Structure Activity Relationships (QSAR). The conventional process of Drug Discovery was all about blind screening approach which was time taking and laborious. This advantage of Conventional Drug Discovery led to the concept of Rational Drug Discovery in 1960's. The knowledge of QSAR ushered in the beginning of CADD[6-8].

Software Used: - Some of the features used software for drug design and their salient features are as follows:-

1. Affinity: - It is automated software, which uses energy of the ligand receptor complex to automatically find the best binding modes of the ligand to the receptor.
2. AutoDock: - It consist of three separate programs-

AutoDock

AutoGrid

AutoTors

It provides an automated procedure for predicting the interaction of ligands with bio-molecular targets and help to narrow the conformational possibilities and in identification of the most suitable structure.

3. Comb build: - It is a structure based drug design program which created to aid the design of combinational libraries.
4. Dock Vision: -DockVision is a docking package created by scientists for scientists by including Monte Carlo, Genetic Algorithm and database screening docking algorithms.
5. FRED:- It is an accurate and extremely fast, multi-conformer docking programme which examines all possible poses within a protein active site, filtering for shape complementarity.
6. FlexDack: - It is a simple, flexible docking of ligands into binding sites on proteins.
7. FlexX: - It is a fast computer program for predicting protein ligand interaction. Its two main applications:-

✓ Complex prediction.

✓ Virtual screening.

8. Glide: - This is a high throughput ligand-receptor docking for fast library screening. It is also a fast and accurate docking programme.
9. Gold: - It calculates docking modes of small molecules into protein binding site which is based on generic algorithm for protein-ligand docking.
10. Dock: - Dock generates many possible orientation of a putative ligand within a user selected region of a receptor structure. It also searches databases for DNA binding compounds.
11. Hint: - It is a hydropathic interaction which translates the well-developed medicinal chemistry and QSAR formation of logP and hydrophobicity into a free energy interaction model for all bio-molecular systems based on the experimental data form solvent partitioning.
12. Ligplot: - It is a program for automatically plotting protein-ligand interaction which generates schematic diagrams of protein-ligand interactions for a given PDB file.
13. Situs: - It is a program package for modelling of atomic resolution structures into low resolution density maps.
14. Vegs: - It calculates ligand-receptor interaction energy.
15. Icm-Dock: - It provides access to the chemical information and a unique set of tools for accurate ligand-protein docking, peptide-protein docking and protein-peptide docking.
16. GRAMM (Global range molecular matching):- This is an empirical approach to smoothing the intermolecular

energy functions by changing the range of the atom-atom potentials.

17. Bielefeld Protein Docking: - It detects geometrical and chemical complementarities between surfaces of proteins and estimates docking positions.
18. Bigger: - It helps in bio-molecular complex generation with global evaluation and ranking.
19. ClusPro: - It integrated approach to protein-protein docking.
20. Ludi: - It fits molecules into the active site of a receptor by matching complementary polar and hydrophobic groups.
21. Ludi/CAP: - It ensures synthetic feasibility of compounds proposed by ludi.
22. DoT: - It is the daughter of TURNIP. Which is used for computation of the electrostatic potential energy between two proteins or other charged molecules?
23. Haddock: - It is a high – ambiguity driven protein-protein docking.
24. Hex: - It is a protein docking and molecular superposition program.
25. Racheal :-It is a real time automated combinatorial Heuristic enhancement of lead compounds.[3]

The concept of QSAR is to convert the new compound into mathematically quantified and computerized. There are two assumptions made with respect to relationship between chemical structure and biological potency of compound[9-10]. First is that, quantitative measure can be derived from structural properties (physicochemical properties like partition coefficient in sub-structural as presence or absence of certain chemical features) significant to biological activity. Other is the relationship between biological activity and molecular property can be derived mathematically.Before designing any drug, it is important to know the feature an 'Ideal Drug 'should have-

- It must be safe and effective.
- It should be absorbed orally.
- It should have high bioavailability.
- It should be metabolically stable.
- It should has long half-life.
- It should be non-toxic.
- It should have minimum or no side-effect.
- It should be selectively distributed to target tissues[11-13].

Computer methods of Drug design works on the postulate that pharmacologically active compounds interact with macromolecules (mainly proteins) through-

- Electrostatic forces
- Hydrophobic Interaction
- H-bond formation

Which are mainly considered during analysis and prediction of interaction. For modelling of interaction between ligand and macromolecules, various methods of calculations are required. These studies are done by multiprocessor computers under UNIX management.[10] In developing new drug, it starts with designing of "ligands" based on how these are recognised by the target proteins to bind it. There is a powerful tool called "LIGBUILD" makes this in Brookheaven format. Performing experiment to know protein dynamics is expensive and time taking[14-16]. That's why, computational tools of dynamics of molecule becomes important. Evaluation is done by approaching 'score' which is a tool to evaluate the binding affinity of protein-ligand complex with known 3D structure. There are many other criteria as well to screen the candidate molecules[17]. Permeability across the bio-membrane is very important.' XLOGP' is able to calculate logP (logarithm of partition coefficient of solute between Octanol and Water) of common Organic compounds. It can provide detailed hydrophobicity distribution information of molecule. 'PLOP' is another tool used to find $\log_P$ values of peptides along with Molecular LipophilicityPotential (MLP) profile with known structure. Rational programs in Drug Design fall in one of the three categories[18-19]-

1. Scanners
2. Builders
3. Hybrids

- Scanners: These programs are used in screening of lead compounds.
- Builders and Hybrids: These mainly used for *de novo* generation of lead compounds. The database contains fragments of chemical building blocks instead of complete compound which creates population of derivatives with compound receptor complementarity[20-22].

It has been 40 years since QSAR found its way practicing in Pharmaceutical Industry. QSAR involves recognising the molecule. The important properties are steric, electronic, lipophilic properties.

QSAR models are necessary because:-

- They are very fast
- They reduce the number of animals used in experiment[23].

QSAR involves:-

1. Conversion of molecular descriptors into mathematical descriptors which encapsulate the key properties of molecules relevant to activity or property being modelled.
2. From a large number of descriptors, selecting the relevant descriptors.
3. Molecular descriptors are mapped into properties.
4. Models are validated to determine how predictive it is[24].

Considerations in QSAR:-

1. The compounds should belong to congeneric series.
2. They should have same mechanism of action.
3. Biological activity should be same[25].

History of QSAR

The first formulation of QSAR was published in 1868 by Crum-Brown and Feaserstating, the physiological activity (Ø) is expressed as function of the chemical structure(c)

$\Phi = f(C)$

After a few decades, Richet, Meyer and Overton found independently the linear relationship between lipophilicity (expressed as solubility or oil-water partition coefficient) and biological effects (like toxicity and narcotic activity. L. Hammett in1930's correlated electronic properties of Organic acids and bases with their equilibrium constants and reactivity. The first steric parameters along with the way of separating polar, steric and resonance effects were introduced by Taft. Hammett and Taft together raised the mechanistic basis for developed QSAR paradigm by Hansch and Fujita to yield the linear Hansch equation and its many extended forms[26-27].

$$\text{Log } 1/C = a\sigma + b\pi + ck \quad \text{..... Linear form}$$

$$\text{Log } 1/C = a \log P - b (\log P)2 + c\sigma + k \quad \text{..... Nonlinear form}$$

Where,

C - Concentration required producing a standard response

Log P - partition coefficient between 1-octanol and water

σ - Hammet substituent parameter

π - Relative hydrophobicity of substituents

a, b, c, k - Model co-efficient

Other methods were also developed to tackle this structure activity questions. Free-Wilson equation is described as

$$BA = \Sigma a i x i + u$$

Where BA is the biological activity, u is the average contribution of the parentmolecule, and a i is the contribution of each structural feature; xi denotes thepresence x i = 1 or absence x i = 0 of a particular structural fragment.

Limitations of this equation led the Fujita Ban equation to be more sophisticated. The equation as follows

$$\text{Log } BA = \Sigma G i X i + u$$

u is defined as the calculated biological activity value of the un-substituted parentcompound of a particular series. G i represents the biological activity contribution ofthe substituents, whereas X i is ascribed with a value of one when the substituent ispresent or zero when it is absent. The chemical information about molecular structure encoded by mathematical procedure is explained in terms of numerical representation called Molecular descriptors. The information content depends on two major factors[28]-

- The molecular representation of compounds.
- The algorithm which is used for calculation of the descriptors.

There are three major types of parameters initially suggested-

- Hydrophobic
- Electronic
- Steric

Software's used in Molecular Descriptors:-

- Dragon
- GAUSSSIAN
- Hyperchem
- CODESSA
- MOE[14]

QSAR methods are classified as:-

1. **Based on dimensionality**

- 1D QSAR: It correlates activity with molecular properties like pka, log P etc.
- 2D QSAR: It correlates activity with structural patterns without taking 3D representation.
- 3D QSAR: It correlates activity with non – covalent interaction fields surrounding molecule.
- 4D QSAR: It includes ensemble of ligand configuration in 3D QSAR.
- 5D QSAR: It represents different Induced-fit models in 4D QSAR.
- 6D QSAR: It further incorporates different solvation methods in 5D QSAR.
- Based on the type of chemometric methods.

2. **Depending upon type of correlation technique employed, QSAR methods are classified as followed:-**

- Linear methods : Include Linear Regression (LR), Multiple Linear Regression(MLR), Partial Least Squares (PLS) and Principal Component Analysis/Regression(PCA/PCR)
- Non-Linear methods: It consists of artificial neural networks (ANN), K-Nearest Neighbours (KNN) and Baysian Neural Nets[29].

Limitations of QSAR

1. Sometimes, the activity measured is inaccurate.

2. 3D components in terms of physicochemical properties are difficult to express.

3. It is difficult to study chiral compounds[30].

CONCLUSION

Approaches used in CADD cannot replace the experimental tests. The purpose of CADD is to generate the hypothesis of probable new compounds and their interaction with targets. These methods can reduce the number of new compounds needed to be synthesised so capable to decrease time-consuming and financial expenses in developing new drug. It is been accepted globally that QSAR based on well-established principles of statistics is valuable medical tool whose application range from explain Structure Activity Relationships quantitatively and retrospectively to endowing synthetic guidance leading to logical and experimentally testable hypothesis.

REFERENCES

1. Albert A, Selective toxicity, Chapman and Hall London, 7th edition, 1985, 324-354.
2. Augen, J. Drug Discovery Today, 2002; 7:315.
3. Baldi A, Computational Approaches for Drug and Discovery. An Overview, College of Pharmacy, Dr.Shri R.M.S Institute of Science and Technology, Bhanpura, Dis. Mandsaur (M.P).
4. Basak S.C., Chemobioinformatics : the advancing frontier of Computer Aided Drug Design in the post genomic era current Computer Aided Drug Design 2012; 8:1-2.
5. B. Hamman, Recurrent Networks for Structural Data. A Unifying Approach and its Properties, Cognitive Systems Research, 2002; 3:145 -185.
6. Borchardt , JK, The Alchemist 6 December 2001.
7. Enyedy IJ, Egan WJ, can we use docking and scoring for hit-to-lead optimization? J. Computer Aided Molecuar Design 2008; 22:161-168.

8. Free SM and Wilson, J.W, J.Med.Chem., 1964; 7:395.
9. Fujita T, Junkichi I, Hansch C, J. Amchan, Soc. 1964; 86:5175.
10. Guida, WC Current Structure Biology, 1994, 4, 777.
11. Hammett L.P,Chem Rev. 1935; 17:125.
12. Hansch C, LeoA, Substituent Constants for Correlation Analysis in Chemistry and Biology join Wiley, 1979, New York.
13. Hansch C, Muer RM, Fujita T, Maloney P, Geiger E, Streich M, J. Amchem. Soc. 1963; 85:2817.
14. http://www.biophy.mpg.de/en/theoretical.biophysics/computaational-drug-design.html
15. Ivanov AS, Veselovsky AV, Strategy of Computer Aided Drug Design "Current Drug Targets Infectious Disorders, 2003, Page No.33-40.
16. Kubinyi. H,Prog. Drug Res, 1979; 23:97.
17. LeoAJ, Weigner D, CROGP, Pomona College Medicinal Chemistry Project, Claremont, 1987, California.
18. Lohse, M.J Trends Pharmacology Science 1998; 19:198.
19. Malik Jitender K, SoniHimesh, Singhai AK, PandeyHansh QSAR. Application in Drug Design. International Journal of Pharmaceutical Research and Allied Sciences.
20. Mccall J. J. Chromatography, 1975; 1:113.
21. Meyer H. Arch Exp. Pathol, Pharmacel, 1899; 42:109.
22. NysBG, Rekker RF,Chim. Therap, 1963; 8:521.
23. OkselCeyda, Wang Xuez, Quantitative Structure Activity Relationship models, University of Leeds (2013).
24. Overton E, Studies Uber diesNarkox, Fischer, Jena, 1901, 45.
25. Swain CG, CuptonEC, JAmchemSec, 1968, 90, 4328.
26. Takagi T, Kimera A, Da YZ, Nakai H, Flujiwara H, Anal. Sci. 1992, 8,761.
27. Tanaka A, Nakamura K, Nakanishi I, Fujwara H, J. Med Chem, 37,1994, 4563.
28. Tomlinson E Chem. Rav.1935; 17:125
29. VermaJitender, Coutinho Evans, Khedka Vijay M, 3D QSAR in Drug Design Current Topics in Medicinal Chemistry, November 2009,95-115.
30. Winkler David A, The role of Quantitative Structure Activity Relationships in Bio-molecular Discovery, 22nd November 2001.