ISSN 2249-3875

International Journal of Pharmaceutical Erudition

Review Article

QSAR : An Approach To Develop New Drug Molecule In The Field Of Medicinal Chemistry

Sanchika Nayak *, S. K. Shyama, P. K. Goyal, Bhupendra Vyas, C. S. Chauhan

B. N. Institute of Pharmaceutical Sciences, Udaipur Rajasthan 313001

QSAR (Quantitative Structure Activity Relationships) have been applied for decades in the development of relationships between physicochemical properties of chemical substances and their biological activities to obtain a reliable mathematical and statistical model for prediction of the activities of new chemical entities. (QSAR) have helped the scientists in the development of mathematical relationships linking chemical structures and pharmacological activity in quantitative manner of series of compound. The fundamental principle underlying the QSAR is that the difference in structural properties is responsible for the variations in biological activities of the compounds. In the classical QSAR studies, affinities of ligands to their binding sites, inhibition constants, rate constants, and other biological end points, with atomic, group or molecular properties such as lipophilicity, polarizability, electronic and steric properties (Hansch analysis) or with certain structural features (Free-Wilson analysis) have been correlated. QSAR certainly decreases the number of compounds to be synthesized by facilitating the selection of the most promising candidates. This review seeks to provide a view of the different QSARapproaches employed within the current drug discovery process to construct predictive structure– activity relationships and also discusses the limitations that are fundamental to these approaches, as well as those that might be overcome with the improved strategies.

Key words: Quantitative Structure Activity relationship, Hansch analysis, QSAR, Drug Design, applications of QSAR.

INTRODUCTION

QSARs (Quantitative Structure–Activity relationships) are based on the assumption that the structure of a molecule (i.e. its geometric, steric and electronic properties) must contain the features responsible for its physical, chemical, and biological properties, and on the ability to represent the chemical by one, or more, numerical descriptor(s). The QSPR (Quantitative Structure–Property relationship) acronymous is used when a property is modeled. Simply it means that "The structure of chemical compound influences its properties and bioactivity.[1] QSAR in simplest terms, is a method for building computational or mathematical models which attempts to find a statistically significant correlation between structure and function using a chemometric technique. In terms of drug design, structure here refers to

***Address for Correspondence**
sanchikanayak@gmail.com

the properties or descriptors of the molecules, their substituents or interaction energy fields, function corresponds to an experimental biological/biochemical end point like binding affinity, activity, toxicity or rate constants. Various QSAR approaches have been developed gradually over a time span of more than a hundred years and served as a valuable predictive tool, particularly in the design of pharmaceuticals and agrochemicals. All one and two dimensional and related methods are commonly referred to as 'classical' QSAR methodologies. It is sometime used in more sense as a hansch analysis[2].

Objectives of QSAR[3]:

Mostly all the QSAR methods focus on the following goals:

1. Quantitative relationship between the structure and physiochemical properties of substances and their biological activity are being used as the foundation stone in search of new medicines. The mathematical and statistical analysis helps us to predict the drug activity.

2. QSAR makes it easy now to reach the conclusion for any of the congener that still not in process, in way that whether it will optimal and profitable or not.

3. To quantitatively correlate and recapitulate the relationships between trends in chemical structure alterations and respective changes in biological endpoint for comprehending which chemical properties are most likely determinants for their biological activities.

4. To optimize the existing leads so as to improve their biological activities.

5. To predict the biological activities of untested and sometimes yet unavailable compounds.

Techniques and Tools of QSAR[4]:

1. Compound Selection: In setting up to run a QSAR analysis, compound selection is an important angle that needs to be addressed. One of the earliest manual methods was an approach devised by Craig, which involves two-dimensional plots of important physicochemical properties. Care is taken to select substituents from all four quadrants of the plot. The Topliss operational scheme allows one to start with two compounds and construct a potency tree that grows branches as the substituent set is expanded in a stepwise fashion. Topliss later proposed a batchwise scheme including certain substituents such as the 3,4-Cl_2, 4-Cl, 4-CH_3, 4-OCH_3, and 4-H analogs.

2. Biological Parameters[5]: In QSAR analysis, it is vital important that the

biological data be both accurate and precise to develop a meaningful model. The equilibrium constants and rate constants that are used extensively in physical organic chemistry and medicinal chemistry are related to free energy values Δ*G*. Thus for use in QSAR, standard biological equilibrium constants such as Ki or Km should be used in QSAR studies.

Source of Activity Biological Parameters:

- Isolated receptor
- Cellular systems
- In vivo systems

3. Statistical Methods[6]: Linear Regression Analysis: The most widely used mathematical technique in QSAR analysis is multiple regression (MRA). Regression analysis is a powerful means for establishing a correlation between independent variables and a dependent variable such as biological activity**.** Yi = *b* + a Xi + Ei Certain assumptions are made with regard to this procedure:

1. The independent variables, which in this case usually include the physicochemical parameters, are measured without error. Unfortunately, this is not always the case, although the error in these variables is small compared to that in the dependent variable.

2. For any given value of *X*, the *Y* values are independent and follow a normal distribution. The error term Ei possesses a normal distribution with a mean of zero.

3. The expected mean value for the variable *Y*, for all values of X, lies on a straight line.

4. The variance around the regression line is constant. The best straight line for model Yi = b + a

Zi + E is drawn through the data points, such that the sum of the squares of the vertical distances from the points to the line is minimized.

Parameters Used in QSAR[7]

1. Lipophilic Parameters : Partition Coefficient, Hydrophobicity parameters.

2. Electronic Parameters : Hammet Constant, Dipole Moment.

3. Polarizability Parameters : Molar refractivity, Parachor.

4. Steric Parameters : Taft's Constant.

5. Miscellaneous Parameters : Molecular Weight, Geometric Parameters

QSAR METHOD[8]:

The QSAR method involves recognition that a molecule (organic,peptide, protein, etc.) is really a three-dimensional distribution of properties. The most importantof these properties are steric (eg shape and volume), electronic (eg electric charge andelectrostatic potential) and lipophilic properties(how polar or non-polar the sections of molecular are usually exemplified

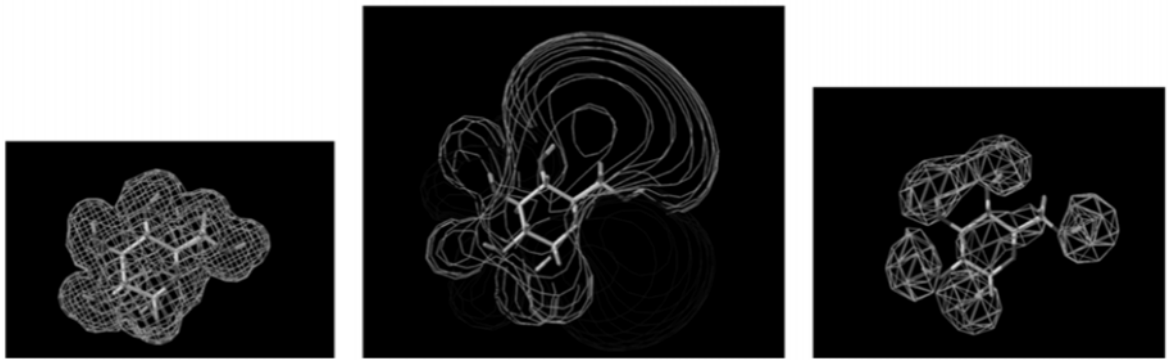

Fig. 1: A small organic molecule (glucopyranole) viewed in steric (left), electrostatic (centre), lipophilic (right) space.

by the log of the octanol-water partition coefficient, log P). Scientists are used to visualizing mainly steric properties of molecules. However, molecules look different when viewed in electrostatic or lipophilic space.

The QSAR method (and analogously QSTR and QSPR) involves a number of key steps[9]:

1. Converting molecular structures into mathematical descriptors that encapsulate the key properties of the molecules relevant to the activity or property being modelled.

2. Selecting the best descriptors from a larger set of accessible, relevant descriptors.

3. Mapping the molecular descriptors into the properties, preferably using a model-free mapping system in which no assumptions are needed as to the functional form of the structure–activity relationship. These relationships are often complex, unknown and non-linear.

4. Validating the model to determine how predictive it is, and how well it will generalise to new molecules not in the date set used to generate the model (the training set).

Various Descriptors Used in QSAR[10]:

- **Molecular descriptors**
- **Fragment descriptors**
- **Whole molecule descriptors**

1. Constitutional Descriptor: Molecular weight, no. of atoms, no. of non-H atoms, no. of bonds, no. of heteroatoms, no. of multiple bonds (nBM), no. of aromatic bonds, no. of functional groups (hydroxyl, amine, aldehyde, carbonyl, nitro, nitroso, etc.), no. of rings, no. of circuits, no of H-bond donors, no of H-bond acceptors, no. of Nitrogen atoms (nN), chemical composition, sum of Kier-Hall electrotopological states (Ss), mean atomic polarizability (Mp),

ISSN 2249-3875

International Journal of Pharmaceutical Erudition

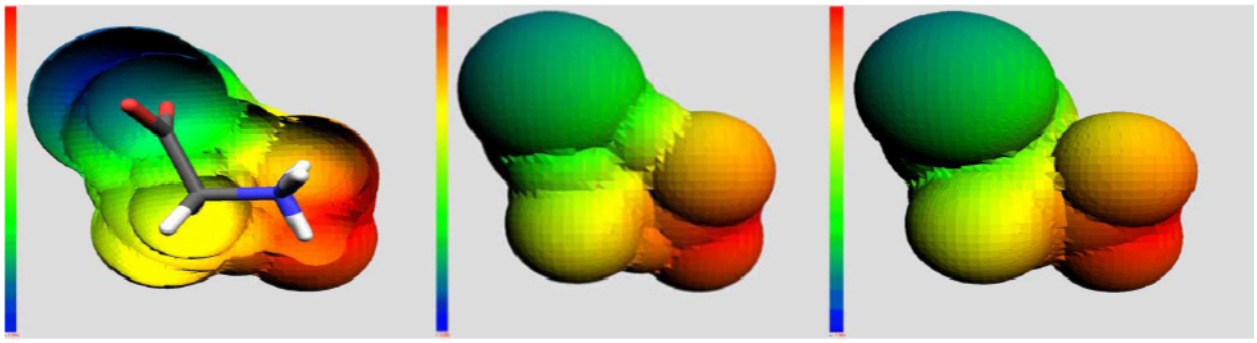

Fig. 2 : Constitutional Descriptor

number of rotable bonds (RBN), mean atomic Sanderson electronegativity (Me), etc. Total number of atoms in the molecule[11].

2. Geometrical Descriptor:

Descriptors using the atomic coordinates (x,y,z) of a molecules are therefore called 3D descriptors. Examples: vander Waals volume, molecular surface, polar surface, etc As a consequence they usually depend on the conformation. 3D petijean shape index (PJI3), Gravitational index, Balaban index, Wiener index, etc.

3. Quantum Mechanical Descriptor:

Highest occupied Molecular Orbital Energy (HOMO) , Lowest Unoccupied Molecular Orbital Energy (LUMO), Most positive charge (MPC), Least negative charge (LNC), Sum of squares of charges (SSC), Sum of square of positive charges (SSPC), Sum of square of negative charges (SSNC), Sum of positive charges (SUMPC), Sum of negative charges (SUMNC), Sum of absolute of charges (SAC), Total dipole moment (DMt), Molecular dipole moment at X-direction (DMX), Molecular dipole moment at Y-direction (DMY), Molecular dipole moment at Z direction (DMZ), Electronegativity (χ= -0.5 (HOMOLUMO)), Electrophilicity ($\omega = \chi 2/2\ \eta$), Hardness (η = 0.5 (HOMO+ LUMO)), Softness (S=1/ η).

4. Functional Group Descriptor:

Number of total tertiary carbons (nCt), Number of H-bond acceptor atoms (nHAcc), number of total hydroxyl groups (nOH), number of unsubstituted aromatic C(nCaH), number of ethers (aromatic) (nRORPh), etc.

5. Chemical Descriptor:

LogP (Octanol-water partition coefficient), Hydration Energy (HE), Polarizability (Pol), Molar refractivity (MR), Molecular volume (V), Molecular surface area(SA). **6. Substituent Electronic Descriptors:**

RMSQ (Root mean square error of charges), SPQ (Sum of positive charges), SNQ (Sum of negative charges), RMSDM (Root mean square of dipole moments at any Cartesian coordinate direction), TDM (Total dipole moment), FRMS (Root mean square force that any atom in constituent molecule see right before the optimization), FMAX (Maximum force on molecule), HOMO (Highest occupied molecular orbital), LUMO (Lowest unoccupied molecular orbital), HD (Hardness), SOF (Softness), EPH (Electrophilicity), EN (Electronegativity)[12].

Descriptor Selection:

To build a good QSAR model, a minimal set of information-rich descriptors is required. The large number of possible indices creates several problems for the modeller.

1. Many descriptors do not contain molecular information relevant to the problem.

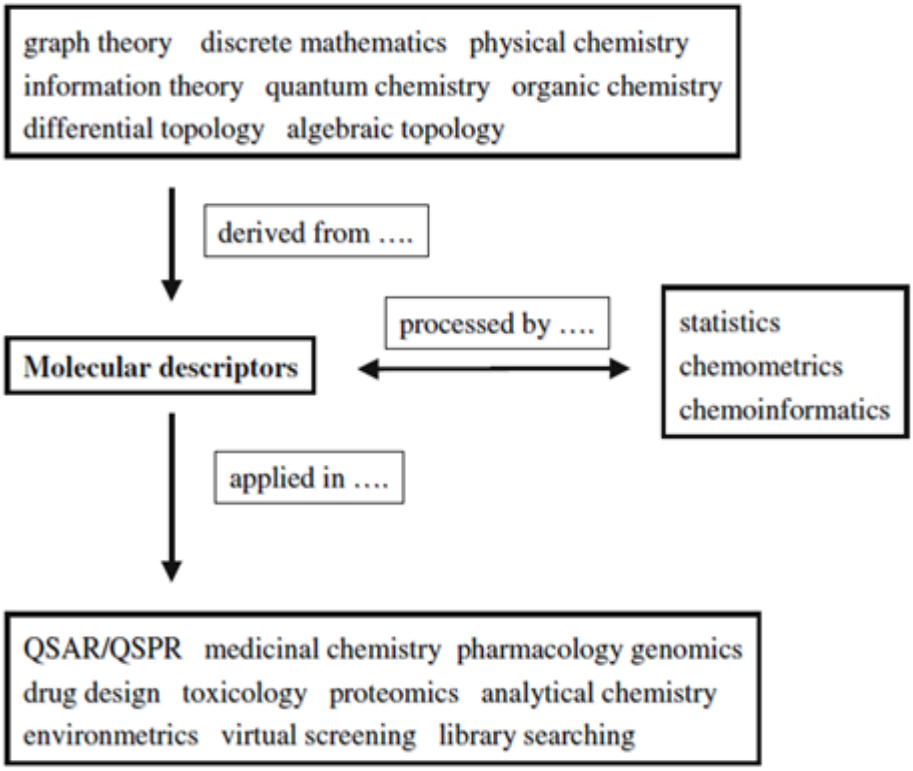

Fig. 3: general scheme of the relationships among molecular structure , molecular descriptors , chemoinformatics and QSAR/ QSPR modeling.

2. Many descriptors are linearly dependent (contain essentially the same information).

3. Use of poor descriptors in QSAR yields poor and misleading models.

4. Including too many descriptors in the model, even if they contain relevant information, can result in overfitting of the model, and loss of ability of the model to generalize to unseen molecules.

5. Many methods of screening this large pool of potential descriptors for relevant ones can lead to chance correlations (correlations that arise by chance because so many descriptors have been tried in models). In other words, if a large number of random numbers are generated as potential descriptors (which clearly do not contain any useful molecular information), and various subsets of these are used to build models, apparently significant models can arise by chance[13].

Methods used in QSAR analysis[14]:

Hansch Analysis:

Corwin Hansch Said **"Similar compounds behave similarly"** Corwin Hansch (born October 6, 1918, Kenmare, North Dakota) is Professor of Chemistry at Po+++mona College inCalifornia. Hansch taught Organic Chemistry for many years at Pomona Colleg+e . His course inPhysical Bio-Organic Medicinal Chemistry was ground-breaking at an undergraduate level.Hansch may be best known as the father of the concept of Quantitative Structure-Activity Relationship (Q.S.A.R.), the quantitative correlation of the physicochemical properties of molecules with their biological activities. He is also noted for the Hansch equation, which is used in

(1) Multivariate Statistics

(2) Hansch Analysis

(3) Hansch-Fujita π constant

Importance of Lipophilicty: Hansch visualized that diffusion into cell is slow process so as it is also important one to determine .It is highly dependent on molecular structure of the drug. Drug must pass out two barriers to put out their effect at site of action ,lipophilliic barrier(cell membrane) and aqueous barrier(cytoplasm) as we know that cytoplasm is made of fatty acids and membrane is made of glycolipids and phospholipids, they have two ends -

(i)Lipophilicity or Hydrophobic—steroids and hydrocarbons

(ii)Hydrophillic end— hydroxyl group in cholesterol, sugar in glycolipids ammonia moiety in phospholipids

1.Linear Hansch model:

The correlation of biological activity with

physicochemical properties is often termed an Extrathermodynamic relationship. Because it follows in the line of Hammett and Taft equations that correlate thermo-dynamic and related parameters, it is appropriately labeled. The Hammett equation represents relationships between the logarithms of rate or equilibrium constants and substituent constants. The linearity of many of these relationships led to their designation as linear free energy relationships. The Hansch approach represents an extension of the Hammett equation from physical organic systems to a biological milieu. It should be noted that the simplicity of the approach belies the tremendous complexity of the intermolecular interactions at play in the overall biological response. It has given rise to the underlying linear Hansch equation also called extra-thermodynamic approach.

log 1/C = b π + c p Ka +d Es + a (Eq.1)

log 1/C = a log P + b σ + c ES + d (Eq. 2)

2.Nonlinear Hansch models:-

(i) Increase in log P value from log Po does not linearly cause increase in biological activity some time its decreases.

(ii) If P values spread over a large range Thus, Hansch et al suggested that the compounds could be involved in a random-walk process: low hydrophobic molecules had a tendency to remain in the first aqueous compartment, whereas highly hydrophobic analogs sequestered in the first lipoidal phase that they encountered. This led to the formulation of a parabolic equation, relating biological activity and hydrophobicity.

log 1/C =-a(log P)2 + b logP+ constant(k) (eq.3)

In the random-walk process, the compounds log 1/C partition in and out of various compartments and interact with myriad biological components in the process. To deal with this conundrum, Hansch proposed a general, comprehensive equation for QSAR.

log 1/*C* = -a (log P)2 + b. log P + ρ σ + δ ES + k (eq. 4)

Where, P = n-octanol/ water partition coefficient, σ = Hammett electronic parameter, a,b,c = regression coefficients, ES = taft's steric factor, k = constant term, ρ (rho) and δ are Proportionality constant

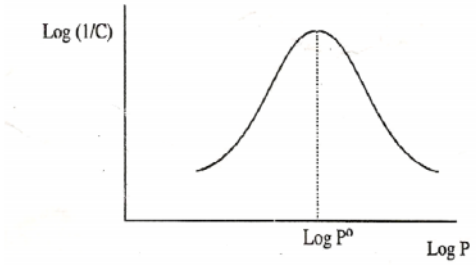

Fig. 4: Graph plotted between Log 1/C vs. log P (Linear Hansch model).

designating the sensitivity of the reaction to electron density. The optimum value of log P for a given system is log Po and it is highly influenced by the number of hydrophobic barriers a drug encounters in its walk to its site of action. The coefficients (a, b, c, d, e) are determined by multi-regression analysis.

Applications of QSAR[14,15]:

• The rational identification of new leads with pharmacological, biocidal or pesticidal activity.

• The optimization of pharmacological, biocidal or pesticidal activity.

• The rational design of numerous other products such as surface-active agents, perfumes, dyes, and fine chemicals.

• The identification of hazardous compounds at early stages of product development or the screening of inventories of existing compounds.

• The designing out of toxicity and side-effects in new compounds.

• The prediction of toxicity to humans through deliberate, occasional and occupational exposure.

• The prediction of toxicity to environmental species.

• The selection of compounds with optimal pharmacokinetic properties, whether it be stability or availability in biological systems.

Conclusion: It involves the mathematical and statistical analysis of SAR-data which helps to reduce the number of educated guesses in molecular modification. QSAR is thus a scientific achievement and an economic necessity to reduce an empiricism in drug design to ensure that every drug synthesized and pharmacologically tested should be as meaningful.

REFERNCE

1. C. Hansch and A. Leo. Substituent Constants for Correlation Analysis in Chemistry and Biology,John Wiley & Sons, New York, 1979.

2. A Crum-Brown, TR Fraser, Trans R Soc Edinburgh; 25; 151: 1868–69.

3. Richardson BJ. Physiological research on alcohols. Med Times Gaz 1868; 2: 703-6.

4. Mills EJ. On melting point and boiling point as related to composition. Philos Mag 1884; 173 87.

5. Richet C. On the relationship between the toxicity and the physical properties of substances. Compt. Rendus Seances Soc Biol 1893; 9: 775-76.

6. Overton E. Osmotic properties of cells in the bearing on toxicology and pharmacology. Z Physik Chem 1897; 22: 189-209.

7. Meyer H. On the theory of alcohol

narcosis I. Which property of anesthetics gives them their narcotic activity? Arch Exp Pathol Pharmakol 1899; 42: 109-18.

8. Hammett LP. Some relations between reaction rates and equilibrium constants. Chem Rev 1935;17: 125-36.

9. Hammett LP. The effect of structure upon the reactions of organic compounds. Benzene derivatives. J Am Chem Soc 1937; 59: 96-103.

10. Ferguson J. The Use of Chemical Potentials as Indices of Toxicity. Proc R Soc Lond B 1939; 127: 387-404.

11. Bell PH, Roblin RO. Studies in chemotherapy. vii. a theory of the relation of structure to activity of sulfanilamide type compounds1. J Am Chem Soc 1942; 64: 2905-17.

12. Albert A, Goldacre R, Phillips J. The strength of heterocyclic bases. J Chem Soc 1948: 2240- 49.

13. Taft RW. Polar and steric substituent constants for aliphatic and o- Benzoate groups from rates of esterification and hydrolysis of esters1. J Am Chem Soc 1952; 74: 3120-28.

14. Hansch C, Maloney PP, Fujita T, Muir RM. Correlation of biological activity of phenoxyacetic acids with hammett substituent constants and partition coefficients. Nature 1962; 194: 178-80.

15. G Klopman, JK Li, S Wang, M Dimayuga. J Chem Inf Comput Sci 1994; 34: 752.