Quantitative Structure-Activity Relationship And its Application in Drug Discovery

Miss. Pranali Jadhav[1], Miss. Kajal Ingavale[2], Miss. Kusum Hanber[3], Dr. Bhagyesh Janugade[4], Ms. Sushama Garud

[1,2,3]*Student of Krishna foundation's Jaywant institute of pharmacy wathar*

[4]*Principal Jaywant institute of pharmacy wathar*

[5]*Assistant Professor of KF's Jaywant institute of pharmacy wathar*

Abstract- Quantitative Structure-Activity Relationship (QSAR) is a mathematical method that finds a quantitative link between the chemical structure of a group of compounds or molecules and their pharmacological activity. Quantitative Structure-Activity Relationship analysis is a ligand-based drug design technique used in drug design and medicinal chemistry. This review gives a full picture of the QSAR method, covering the most important steps: preparing the data, calculating molecular descriptors, building the model, and testing it. It has changed from a Quantitative Structure-Activity Relationship to a 3D-Quantitative Structure-Activity Relationship over time. Quantitative Structure-Activity Relationship (QSAR) retroversion models connect a group of "predictor" variables (X) to the strength of the response variable (Y). In Quantitative Structure-Activity Relationship modeling, the predictors are physicochemical properties. Quantitative Structure-Activity Relationship (QSAR) models initially encapsulate the correlation between chemical structure and biological activity within a dataset of chemicals. Second, Quantitative Structure-Activity relationship models use quantitative structure–property relationships to guess what new chemicals will do. In these models, the physicochemical property is the response variable. This method helps find important molecular features that affect activity and improve lead compounds. QSAR helps figure out how new compounds will act, which is useful for drug design, figuring out how toxic a substance is, and making chemicals with certain properties. Quantitative structure- activity relationship is a useful tool in drug discovery and computational chemistry. It helps scientists make molecules that are safer and work better.

Keywords: QSAR, Biological activity, Drug Design, Predictor, Physicochemical properties.

I.INTRODUCTION

Early in the 1960s, Corwin Hansch expanded on the idea of Linear Free Energy Relationships (LFER) to explain how effective a biologically active molecule is. The resulting equations were dubbed Quantitative Structure-Activity Relationships (QSAR) or Quantitative Structure-Property Relationships (QSPR), and this method quantitatively connected a compound's structure to its activity. Establishing a relationship between a set of molecules physiochemical characteristics and forms of activity (biological activity) is the goal of QSAR. Similar relationships between structure and chemical reactivity were the foundation of QSAR [1]. The use of computational methods, especially Quantitative Structure-Activity Relationship (QSAR) and molecular docking studies, has greatly advanced the field of drug discovery and design [2].

❖ QSAR'S PHYSICOCHEMICAL PARAMETERS

- Groups of related compounds are studied using QSAR.
- QSAR studies on sets of compounds with different structural makeup are more prevalent, though.
- A large number of parameters should be taken into account in both situations.

❖ Physical-chemical characteristics

- The hydrophobic nature of the compound
- The ability of substitutes to repel water
- Electronic characteristics of substitutes
- The steric characteristics of substitutes

❖ The parameters

In QSAR studies, a number of parameters are:

1. Lipophilic characteristics: partition coefficient, n-constant of substitution.
2. Electronic parameters: dipole moment and Hammett constant.
3. Steric parameters Verloop steric parameter, molar refractivity, and Taft's constant.

❖ PARTITION COFFICIENT

- A drug's hydrophobic nature is essential to its ability to pass through cell membranes and may also play a significant role in receptor interactions.
- A drug's hydrophobicity is experimentally assessed by measuring its relative distribution in an octanol water mixture.
- The partition coefficient is the name given to this relative distribution.[3]

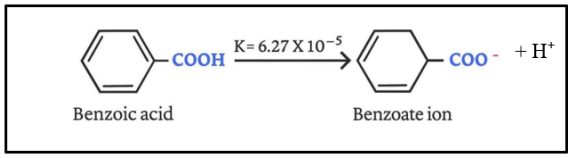
$$\text{Partition coefficient } P = \frac{[\text{Conc. in octanol}]}{[\text{Conc. in water}]}$$

Often, drug activity is connected to P The biological activity

$$\text{Log}(1/c) = K_1 \log P + K_2$$

HAMMETT CONSTANT (σ);

- The Hammett substituent constant (σ) quantifies the electron-withdrawing or electron- donating nature of a substituent group attached to an aromatic system. It reflects how a functional group influences the electronic environment of a molecule, particularly in relation to its reactivity and equilibrium behavior. This constant is typically derived by comparing the dissociation behavior of substituted benzoic acid derivatives to that of unsubstituted benzoic acid, under controlled conditions. The substituent's effect is evaluated based on its position (meta or para) relative to the functional group, allowing for a standardized measure of electronic influence.
- Louis Hammett (1894–1987) pioneered this approach, establishing a linear free- energy relationship that correlates the electronic characteristics of organic acids and bases with their equilibrium constants and reaction rates. His work laid the foundation for understanding how molecular structure governs chemical behavior, enabling predictive models for reaction outcomes across a wide range of organic transformations.



Two distinct scenarios could arise at this point because of the following:

1. A moiety that withdraws electrons, and
2. The aromatic ring may contain an electron-releasing (donating) moiety, which would result in an entirely different electronic status from the "Aryl Nucleus.".

1. ELECTRON- WITHDRAWING GROUPS (EWGs)

- These groups pull electron density away from the aromatic ring and the carboxyl group, stabilizing the negative charge on the carboxylate anion after deprotonation.

➢ Examples of EWGs:

- Nitro (NO_2)
- Cyano (CN)
- Carboxylic acid (COOH)
- Ester (COOR)
- Amide ($CONH_2$, CONHR, $CONR_2$)
- Aldehyde (CHO)
- Ketone (COR)
- Sulfonyl (SO_2R, SO_2OR)
- Nitroso (NO)

➢ Effect:

- Stabilizes the conjugate base (carboxylate anion)
- Shifts equilibrium toward ionization
- Increases acidity (lower pKa)
- Results in a higher KX value (greater dissociation constant)

2. ELECTRON- DONATING GROUPS (EDGs)

- These groups push electron density into the ring, destabilizing the carboxylate anion and making the acid weaker.

- Examples of EDGs:
  - Alkyl groups (e.g., CH_3)
  - Hydroxy (OH)
  - Alkoxy (OR)
  - Amino (NH_2, NHR, NR_2)

- Effect:
  - Destabilizes the conjugate base
  - Shifts equilibrium toward the non-ionized form
  - Decreases acidity (higher pKa)
  - Results in a lower KX value

II.QSAR MODELS

QSAR Development Process:

The first stages of QSAR research involve the input of molecular structures and the creation of 3D models. A three-dimensional molecule model is required in order to compute the geometric description. The development of molecular structure descriptors is the second crucial phase in QSAR research. Selecting descriptors is the third phase, and feature assortment techniques are used for this. The fourth key phase in QSAR investigations is creating a model using a descriptor set; the fifth and last step is validating the model by predicting the molecule's activities using an external forecasting set. The estimation results are compared with those from the training and cross-validation sets in order to quickly identify the best fitting model. (4,5)

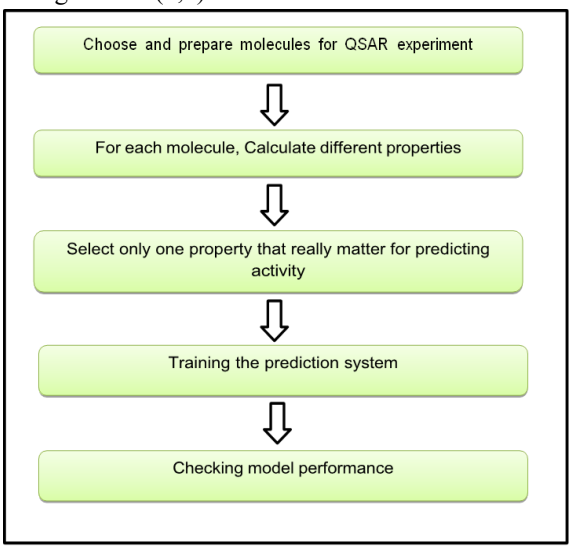


FIG.2.1: QSAR DEVELOPMENT PROCESS

MODELS OF QSAR:

Following the creation of QSAR method of introducing different models in QSAR:

Hansch Analysis:

Two kinds of linear Free-related energy methods:

1. Models That Are Linear Corwin:

According to Hansch, the basic lipophilicity, which is generally understood to be the Octanol-water partition coefficient. (P), on activity with a biological origin in 1969. This property determines the compound's bioavailability, which compound makes it to the intended location. The formula is:

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

The molar level of 'C` in those equations is concentration of the substance responsible for producing a overall reaction, such as LD50, ED50, IC50, EC50, etc. The correlation improved with the help of Hammett's electronic parameter combined with Hansch's lipophilicity degree with the help of utilizing the equation as follows:

$= k1\pi + k2\sigma + k3 = \log (1/C)$

Hence, π is comparable to σ since σ = Hammett substitution factor[6].

2. Non-Linear Models:

The development of the Hansch parabolic equations, which incorporate (log P) elements in the QSAR equations, was prompted by the failure of linear equations in the broad range of hydrophobic ties. The fact that several membranes must be crossed in order for chemicals to pass through together at the appropriate location can be explained in one of two ways. The compounds that have the highest hydrophobicity of the membranes will first become localized after crossing [6,7].

According to Hansh's method, variations in chemical structure are associated with variations in properties, such as electronic, lipophilic, and potentially steric substituents in biological reactions.It is represented mathematically as:

$\log (1/C) = \Delta Gh + \Delta Ge + \Delta Gs + \text{constant}$

$\log (1/C) = a \log P - b (\log P)\, 2+ c\sigma + dEs + \text{constant}$

The Taft steric constant is Es, the Hammett electronic constant is σ, and the logarithm in the partition coefficient is Log P. Multiple regression analysis

yields coefficients that are appropriate for biological data a, b, c, and d.[6].

➢ Benefits:

1. Small organic molecule descriptors (σ, π, Es, etc.) that are used to describe biological systems.
2. Predictions can be measured and quantified statistically.
3. It's easy and quick.
4. Potential for extrapolation.

➢ Drawbacks:

1. A lot of the compounds are needed.
2. There is comparatively little application of small molecule descriptors to biological systems.
3. The use of steric factors in biological systems is limited.
4. Partial protonation of drugs in physiological settings[6].

❖ Free Wilson Analysis/De Novo Approach:

Wilson analysis is a straightforward procedure. In the early phases of lead structure optimization, this tool is helpful. The smallest quantity of substances required for One kind of regression equation analysis that examines each of the parameters and is called free Wilson analysis[8,9,10].

A set of formulated linear equations is solved using linear regression analysis. A fact-based SAR model was the free-wilson approach. An indicator variable is created for each structural trait that is different from a set of randomly chosen compounds[6].

$$\Sigma ajXij + \mu = \log (1/C)$$

"A standard QSAR, which presumes that substituent effects are additive and constant, is similar to this de novo technique."[6]. Log (1/C) has shown signs of physiological activity. If a particular substituent or structural feature, the third substituent Xj, is present, its value is 1, and if not, it is 0. This represents a total average action and shows how crucial the Jth substituent is to physiological activity. Each level's total action counts add up to zero.[6]

Advantages:

1. Making a table for regression analysis is easy.
2. Compound additions and deletions are simple processes that barely affect the values of other regression coefficients.
3. Any compound can be used as a reference compound.
4. A pseudo substituent is made up of two substituents and always appears together in two different places within the molecule.
5. Singularity issues are typically avoided.

Limitations:

1. First and foremost, two distinct substitution positions require structure variation. Otherwise, each component would result in a pointless collective contribution.
2. Its inability to offer a strong basis for assessing outcomes in the form of drug-receptor interactions is a drawback.

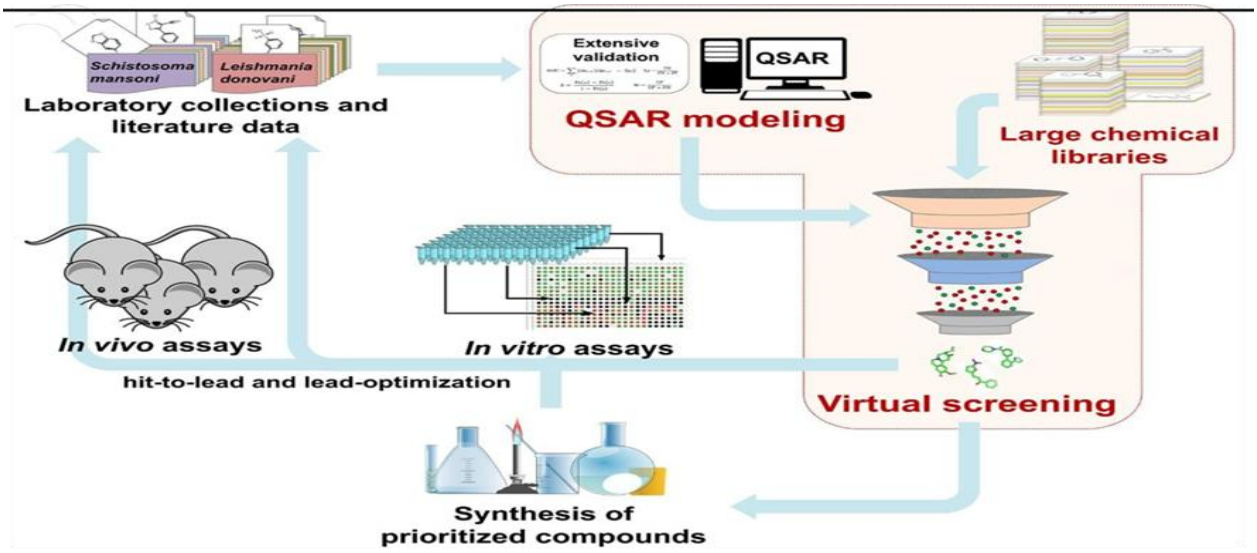


FIG2.1: FLOW CHAER OF QSAR MODELING [11].

III. APPLICATIONS OF QSAR

Predicting Drug Activity is one of the many applications of QSAR models in drug design:

- Prediction of Drug Activity: QSAR models aid in forecasting the novel compounds' biological activity according to their structural characteristics. This can help choose which compounds to test experimentally.
- Drug Candidate Optimization: By altering the chemical structure of lead compounds, QSAR enables the optimization of these compounds to improve desired properties (e.g., increasing potency or reducing toxicity).
- Virtual Screening: Before physical synthesis, QSAR can be used to screen huge libraries of compounds for possible drug candidates, saving time and money.
- Toxicological Predictions: QSAR plays a crucial role in risk assessment and medication safety by being used to predict the toxicity of chemicals.
- Pharmacokinetic Property Prediction (ADMET): Predicts Absorption, Distribution, Metabolism, Excretion, and Toxicity properties.
- Receptor–Ligand Interaction Studies: Helps understand binding affinity and interaction mechanisms.
- Optimization of Lead Compounds: Guides structural modification to improve potency, selectivity, or safety.
- Environmental Chemistry: Predicts environmental fate (biodegradability, bioaccumulation, etc.) of pollutants.
- Cosmetic and Agrochemical Research: Used for designing safe cosmetic ingredients and effective pesticides/herbicides.
- Mechanism of Action Studies: Correlates structural features with specific biological mechanisms.
- Formulation Development: Helps predict solubility, permeability, and stability of compounds in formulations.
- Metabolic Stability Prediction: Assesses how structural changes affect metabolic degradation in the body.
- Nanomaterial Safety Evaluation: Used to model the toxicity and reactivity of nanomaterials based on their structure.
- Protein–Ligand Binding Affinity Prediction: Estimates how strongly a ligand binds to its target protein without experimental data.
- Chemical Risk Assessment: Supports regulatory agencies in evaluating chemical safety without animal testing.
- Biochemical Pathway Modelling: Helps predict interactions of molecules in enzymatic and signaling pathways.
- Green Chemistry: Assists in designing environmentally friendly chemicals with minimal toxicity.
- Food Chemistry: Used to predict antioxidant, antimicrobial, or preservative activity of food additives[12].

Limitations of QSAR:

- Data Quality Dependence: QSAR accuracy depends on reliable experimental data; poor data lead to poor models.
- Limited Applicability Domain: Works only within the range of compounds similar to the training set.
- Overfitting Risk: Too many descriptors can make the model fit noise instead of real trends.
- Descriptor Selection Issues: Choosing inappropriate descriptors can reduce prediction accuracy.
- Lack of Mechanistic Insight: QSAR may not explain why activity occurs, only predicts it.
- Biological Variability: Does not account for metabolic, absorption, or in vivo biological factors.
- Structural Diversity Constraint: QSAR performs poorly when compounds have highly diverse structures.
- Insufficient Sample Size: Small datasets reduce model reliability and statistical significance.
- Assumption of Additivity: QSAR often assumes each descriptor contributes independently to activity, which may not be true[13].

Challenges of QSAR

QSAR's challenges Despite providing valuable insights into drug design, QSAR has a number of drawbacks.

- Data Availability: A sizable and varied dataset of substances with established biological activities is necessary for QSAR models. Predictions that are not accurate can result from incomplete or biased datasets.
- Model Overfitting: An overly complicated model may perform well on the training set but not be able to generalize to new compounds, which is known as overfitting.
- Descriptor Selection: The model's performance is greatly impacted by the molecular descriptors chosen. For QSAR to be successful, choosing the appropriate descriptors is essential.

IV.CONCLUSION

Quantitative Structure–Activity Relationship (QSAR) studies provide a powerful computational approach to predict the biological activity of compounds based on their chemical structure. Through descriptor calculation, model development, and validation, QSAR enables efficient screening of large compound libraries, reducing time and cost in drug discovery. When applied to antimicrobial or antidiabetic agents, QSAR helps in identifying promising lead molecules before synthesis and experimental testing. In conclusion, QSAR serves as a reliable, cost-effective, and time-saving tool in modern drug design, aiding in rational molecule selection and optimization for therapeutic applications.

REFERENCE

[1] Patel, H. (2023). Medicinal Chemistry 3: Unit 5 Notes (Semester 6). PharmaDBM. https://pharmdbm.com/medicinal-chemistry-3-unit-5-notes-semester-6

[2] Tropsha A, Gramatica P, Gombar VJ, 2003. The Importance of Being Earnest: Validation is the Absolute Essential for Successful Application and Interpretation of QSPR Models. QSAR&CombSci.22, Pages69-77. DoI:10.1021/cr950066q (http://dx.doi.org/10.1021%2Fcr950066q).

[3] Chirico N, Gramatica P, 2012. Real external predictivity of QSAR models. Part 2. New intercomparable thresholds for different validation criteria and the need for scatter plot inspection. J Chem Inf Model. 52(8), Pages 2044-2058. Doi: 10.1021/ci300084j (http://dx.doi.org/10.1021%2Fci300084j).

[4] Illah AL, Veljovic E, Gurbeta L and Badnjevic A: Application of QSAR Study in Drug Design. International Journal of Engineering Research and Technology (IJERT)2017; 6(06): ISSN: 2278-0181

[5] Sahigara F, Mansouri K, Ballabio D, Mauri A, Consonni V and Todeschini R: Comparison of Different Approaches to Define the Applicability Domain of QSAR Models. Molecules 2012; 17: 4791-4810.

[6] Vaidya A and Jain S: Quantitative structure activity relationship: A novel approach of drug design and discovery. Journal of Pharmaceutical Sciences and Pharmacology, American Scientific Publishers 2014; 1:219-32.

[7] Penniston JT, Beckett L, Bentley DL and Hansch C: Passive permeation of organic compounds through biological tissue: a nonsteady-state theory. Molecular Pharmacology 1969;5:333-36

[8] Fujita T and Ban T: Structure-activity study of phenylethylamines as substrates of biosynthetic enzymes of sympathetic transmitters. Journal of Medicinal Chemistry 1971; 14: 148–52.

[9] Free J and Wilson SM: Mathematical contribution to structure activity studies. Journal of Medicinal Chemistry 1964; 7: 395–99.

[10] Kubinyi H: Free Wilson Analysis. Theory, Applications and its Relationship to Hansch Analysis. Quantitative Structure-Activity Relationships 1988; 7(3): 121–33.

[11] Zuiker E, Serpa G, De M, Liu Y, Wozniak DJ, Gowdy KM, Charron J, Birket SE, Kiedrowski MR, Hemann EA, Long ME. Targeting the MEK1/2 pathway to combat Staphylococcus aureus infection and inflammation in cystic fibrosis. mBio. 2025 Jul 9;16(7):e0077525. doi: 10.1128/mbio.00775-25. Epub 2025 May 27. PMID: 40422262; PMCID: PMC12239583.

[12] Ajmani S, Jadhav K, Kulkarni SA, 2008. Group-Based QSAR (G-QSAR): Mitigating Interpretation Challenges in QSAR. QSAR & Combinatorial Science. 28(1), Pages 36–51. Doi:10.1002/qsar.200810063.

[13] Ojha L.K., Sharma R., and Bhawsar M.R.,

(2013) Modern drug design with advancement in QSAR: A review, 'International Journal of Research in BioSciences' Vol. 2

[14] Manoharan P, Vijayan RSK, Ghoshal N, 2010. Rationalizing fragment based drug discovery for BACE1: insights from FB-QSAR, FB-QSSR, multi objective (MO-QSPR) and MIF studies. Journal of Computer-Aided Molecular Design. 24(10), Pages 843–864. Doi: http://dx.doi.org/10.1007%2Fs10822-010-9378-9).