



**INTERNATIONAL JOURNAL OF  
PHARMACEUTICAL SCIENCES**  
[ISSN: 0975-4725; CODEN(USA): IJPS00]  
Journal Homepage: <https://www.ijpsjournal.com>



## Review Article

# Quantitative Structure–Activity Relationship (QSAR) in Modern Drug Design: Concepts, Applications, and Future Prospects

Ubale Asmita\*, Shingote Vishakha

Vidya Niketan Institute of Pharmacy and Research Center, Bota, Sangamner, Ahilyanagar 422602

### ARTICLE INFO

Published: 08 July 2026

**Keywords:**

Quantitative Structure–Activity Relationship (QSAR); Drug Design; Molecular Descriptors; Computational Chemistry; Pharmacokinetics and Toxicity Prediction

**DOI:**

10.5281/zenodo.21267104

### ABSTRACT

Quantitative Structure–Activity Relationship (QSAR) has emerged as one of the most powerful computational approaches in rational drug design, bridging the gap between chemical structure and biological activity. By utilizing molecular descriptors such as hydrophobicity, electronic distribution, and steric parameters, QSAR enables researchers to build predictive models that correlate structural features with pharmacological responses. The significance of QSAR lies in its ability to reduce experimental costs, shorten timelines, and minimize animal testing by identifying and optimizing lead compounds in silico before synthesis. Over the decades, QSAR has evolved from simple linear regression models into sophisticated multidimensional approaches, including 2D, 3D, and even 6D-QSAR, which integrate receptor interactions, conformational flexibility, and machine learning algorithms. In drug discovery, QSAR is widely applied in lead identification, optimization, and toxicity prediction. It has been successfully used in the development of cardiovascular drugs, anticancer agents, antivirals, and antibiotics.

### INTRODUCTION

Drug discovery is a complex, expensive, and time-consuming process that has remained one of the greatest challenges in modern pharmaceutical sciences. The traditional pipeline, starting from the identification of bioactive compounds through clinical evaluation, can take more than a decade and require investments exceeding billions of dollars. Despite such massive investments, the

success rate of new drug candidates remains discouragingly low. Studies suggest that only one out of 5,000–10,000 screened compounds eventually reaches the market, with many candidates failing due to poor efficacy, toxicity, or unfavorable pharmacokinetic profiles. This high rate of attrition not only increases costs but also delays the delivery of novel therapies to patients who are in urgent need.[1]

\*Corresponding Author: Ubale Asmita

Address: Vidya Niketan Institute of Pharmacy and Research Center, Bota, Sangamner, Ahilyanagar 422602

Email ✉: [ubaleasmita446@gmail.com](mailto:ubaleasmita446@gmail.com)

**Relevant conflicts of interest/financial disclosures:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.



To address these challenges, the pharmaceutical industry has progressively turned toward computational approaches to streamline drug discovery. Over the last few decades, *in silico* methodologies have evolved from basic molecular modeling techniques to highly sophisticated computer-aided drug design (CADD) tools. These strategies enable researchers to simulate drug–target interactions, predict biological activities, and evaluate pharmacokinetic and toxicological properties prior to laboratory synthesis. Among these computational strategies, Quantitative Structure–Activity Relationship (QSAR) modeling stands out as one of the most widely used and reliable tools in rational drug design.[2]

QSAR is based on the fundamental principle that the biological activity of a compound is directly related to its chemical structure. By mathematically correlating molecular descriptors—such as hydrophobicity, electronic properties, steric parameters, and hydrogen-bonding capacity—with observed biological responses, QSAR provides predictive insights into the activity of novel compounds. This approach allows medicinal chemists to identify lead molecules, optimize drug-like properties, and minimize toxicity risks without extensive experimental screening. The integration of QSAR into the early stages of drug design significantly

reduces both cost and time, making the process more efficient and sustainable.[3]

The importance of QSAR extends beyond predicting activity. It plays a pivotal role in addressing several critical aspects of drug discovery, including:

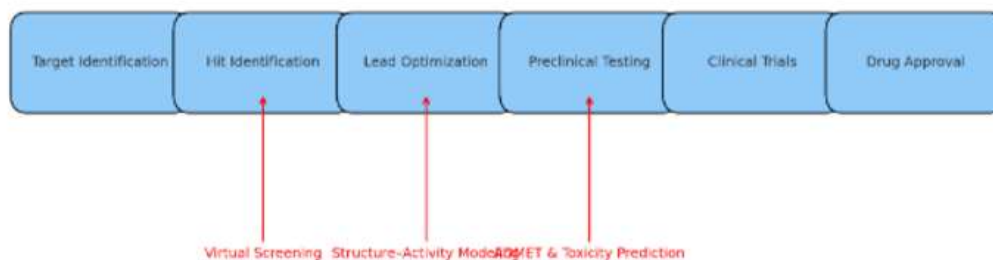
**Lead Identification and Optimization:** Assisting in the selection of promising compounds from large chemical libraries.

**ADMET Prediction:** Evaluating absorption, distribution, metabolism, excretion, and toxicity profiles before clinical testing.

**Drug Repurposing:** Identifying new therapeutic indications for existing drugs.

**Environmental and Regulatory Sciences:** Assessing the toxicity of chemicals and aiding in regulatory decision-making.[5]

Given its wide applications, QSAR has undergone significant evolution since its inception in the 1960s with Hansch and Fujita’s linear models. The field has progressed from simple regression-based 2D models to advanced multidimensional approaches, including 3D-QSAR, 4D-QSAR, and even 6D-QSAR, integrating machine learning and artificial intelligence for enhanced accuracy.



**Figure 1 : Role of QSAR in the Drug Discovery Pipeline**

### Objective of this Review

The purpose of this review is to provide a comprehensive overview of QSAR and its role in modern drug design. Specifically, this paper will:



1. Discuss the fundamental principles of QSAR and its methodologies.
2. Highlight the different types of QSAR models and molecular descriptors.
3. Explore the applications of QSAR in various therapeutic areas.
4. Examine the challenges and limitations of QSAR approaches.
5. Present future perspectives, including the integration of artificial intelligence and big data.

By systematically exploring these aspects, this review aims to highlight the transformative role of QSAR in reducing the time, cost, and risks associated with drug discovery, ultimately supporting the development of safer and more effective therapeutic agents.

### Historical Background of QSAR

The concept of relating chemical structure to biological activity has been a cornerstone of medicinal chemistry for more than a century. However, the systematic and quantitative framework that underpins modern QSAR (Quantitative Structure–Activity Relationship) emerged in the 1960s through the pioneering work of Corwin Hansch. His contributions laid the foundation for computational drug design and still form the basis of many QSAR studies today.

### The Work of Corwin Hansch and the Birth of QSAR

In the early 1960s, Corwin Hansch, a professor of chemistry at Pomona College, recognized that the biological activity of a compound was influenced by multiple physicochemical properties, such as lipophilicity, electronic distribution, and steric

effects. He proposed a systematic method to mathematically correlate these properties with biological responses. This idea became the foundation of the QSAR field.

The classical Hansch–Fujita equation (1964) expressed biological activity ( $\log 1/C$ , where  $C$  = molar concentration producing activity) as a linear function of molecular descriptors:

$$\log(1/C) = a\pi + b\sigma + cE_s + k$$

Where:

- $\pi$  (**pi**) represents hydrophobicity or lipophilicity (partition coefficient).
- $\sigma$  (**sigma**) represents electronic effects derived from Hammett substituent constants.
- $E_s$  denotes steric parameters based on substituent size.
- **a, b, c, k** are regression coefficients obtained statistically.

This equation provided the first formal mathematical model linking structure to activity, enabling researchers to predict biological effects of untested compounds. The Hansch–Fujita approach was revolutionary because it offered not only mechanistic insights but also predictive power.

### Evolution and Milestones in QSAR Development

Since Hansch’s work, QSAR has undergone continuous refinement and expansion, leading to multiple “generations” of models:

1. **1D-QSAR** (1960s): Based on simple physicochemical parameters such as  $\log P$ ,  $pK_a$ , and Hammett constants. Provided linear regression models for small datasets.



2. **2D-QSAR** (1970s–80s): Incorporated topological, constitutional, and fragment-based descriptors derived from 2D molecular structures. Widely used in early computational drug design.
3. **3D-QSAR** (1980s–90s): Introduction of Comparative Molecular Field Analysis (CoMFA) and Comparative Molecular Similarity Indices Analysis (CoMSIA), allowing spatial visualization of steric and electrostatic influences on activity.
4. **4D-QSAR** (1990s): Incorporated conformational flexibility and dynamic aspects of molecules by analyzing multiple conformers.
5. **5D- and 6D-QSAR** (2000s onward): Integrated solvation effects, receptor flexibility, and time-dependent dynamics, providing a more realistic representation of ligand–target interactions.

These milestones illustrate how QSAR evolved from simple linear models to multidimensional frameworks that incorporate advanced computational and biophysical concepts.[7]

### Impact of Computational Advancements

The evolution of QSAR has been closely tied to progress in computational technologies. Early QSAR models were restricted to small datasets due to limited computing power. With the advent of high-performance computing, molecular modeling software, and machine learning algorithms, the scope of QSAR has dramatically expanded. Today, researchers can analyze large chemical libraries containing millions of compounds and employ sophisticated nonlinear methods such as:

### Artificial Neural Networks (ANNs)

### Support Vector Machines (SVMs)

### Random Forests (RF)

### Deep Learning (DL)

These advancements have enhanced predictive accuracy and applicability, transforming QSAR into a central tool of modern drug discovery. Additionally, integration with databases (e.g., PubChem, ChEMBL) and big data approaches has allowed QSAR to reach beyond drug design, impacting toxicology, regulatory sciences, and environmental chemistry.[11]

### Fundamentals of QSAR

#### Definition and Concept

Quantitative Structure–Activity Relationship (QSAR) is a computational modeling technique that establishes a mathematical correlation between the structural properties of chemical compounds and their biological activity or physicochemical behavior. The fundamental assumption of QSAR is the similarity principle: “structurally similar molecules exhibit similar biological activities.”

In drug discovery, QSAR is used to predict the activity of untested compounds, optimize lead molecules, and reduce the need for expensive experimental screening. QSAR integrates chemistry, biology, mathematics, and statistics to provide predictive insights that accelerate rational drug design.[13]

#### Stepwise QSAR Approach

##### Data Selection

- The foundation of any QSAR model is a well-curated dataset.



- Data is collected from experimental assays, literature, or chemical databases.
- Key considerations:
  - Structural diversity (avoid redundancy).
  - Reliable and reproducible bioassay results.
  - Clear activity endpoint (e.g., IC<sub>50</sub>, K<sub>i</sub>, LD<sub>50</sub>).

### Descriptor Calculation

- Molecular descriptors are numerical values that encode structural and physicochemical information about molecules.
- Types of descriptors:
  - **1D descriptors:** Molecular weight, number of atoms, logP.
  - **2D descriptors:** Topological indices, connectivity indices, fingerprints.
  - **3D descriptors:** Molecular fields, steric/electrostatic potentials.
  - **Higher dimensions (4D–6D):** Conformational flexibility, receptor binding, dynamic simulations.[17]

### Model Building

- Mathematical models are built by correlating descriptors with biological activity.
- Common statistical/machine learning techniques:
  - **Regression methods:** Multiple Linear Regression (MLR), Partial Least Squares (PLS).
  - **Classification models:** Decision Trees, Support Vector Machines (SVM).

- **Machine Learning/AI:** Random Forests, Neural Networks, Deep Learning.

### Validation

- Validation ensures the reliability and predictive power of a QSAR model.
- Techniques:
  - **Internal validation:** Cross-validation, bootstrapping, y-randomization.
  - **External validation:** Testing on an independent dataset not used in training.
  - **Statistical parameters:** R<sup>2</sup> (goodness of fit), Q<sup>2</sup> (predictive ability), RMSE (error).[19]

### Prediction

- Once validated, the QSAR model can be used to predict activity/toxicity of new molecules.
- Applications include:
  - Virtual screening of large chemical libraries.
  - Lead optimization in medicinal chemistry.
  - Early ADMET profiling.
  - Environmental risk assessment.

### General QSAR Equation

The classical Hansch–Fujita type QSAR equation can be represented as:

$$\text{Biological Activity} = a(\text{Hydrophobic}) + b(\text{Electronic}) + c(\text{Steric}) + d$$

$$\log(1/C) = a\pi + b\sigma + cE_s + k$$

Where:



- **a, b, c** = regression coefficients (weights of contribution).
  - **Hydrophobic term ( $\pi$ )**:  $\log P$ , partition coefficient (lipophilicity).
  - **Electronic term ( $\sigma$ )**: Hammett constant, electronic substituent effects.
  - **Steric term ( $E_s$ )**: Taft constant, spatial hindrance.
  - **d** = constant (intercept).[23]
- $\log P > 0 \rightarrow$  lipophilic compound;  $\log P < 0 \rightarrow$  hydrophilic compound.
  - Example: Many CNS drugs require higher  $\log P$  values to cross the blood–brain barrier.

## Molecular Descriptors in QSAR

Molecular descriptors form the foundation of QSAR modeling, as they provide a quantitative representation of chemical structures. They transform complex molecular information into numerical values that can be correlated with biological activity. These descriptors capture essential physicochemical, structural, and electronic features of compounds, thereby helping in the prediction of pharmacological profiles. Based on their nature, molecular descriptors can be broadly categorized into hydrophobic, electronic, steric, and advanced descriptors.[29]

## Hydrophobicity Descriptors

Hydrophobic interactions play a central role in determining drug–receptor binding affinity, membrane permeability, and bioavailability. Hydrophobic descriptors measure the lipophilic character of molecules.

## Key Parameters

### 1. $\log P$ (Partition Coefficient):

- Defined as the logarithm of the ratio of solute concentration in octanol to that in water.

### 2. $\log D$ (Distribution Coefficient):

- Unlike  $\log P$ ,  $\log D$  considers pH-dependent ionization.
- Useful for ionizable drugs since biological environments have variable pH (e.g., stomach vs. blood).[30]

### 3. Hydrophobic Substituent Constant ( $\pi$ ):

- Introduced by Hansch,  $\pi$  values quantify the effect of substituents on lipophilicity.
- Positive  $\pi \rightarrow$  lipophilic substituents (e.g.,  $-\text{CH}_3$ ,  $-\text{Cl}$ ).
- Negative  $\pi \rightarrow$  hydrophilic substituents (e.g.,  $-\text{OH}$ ,  $-\text{NH}_2$ ).

## Relevance to Lipinski's Rule of Five

- Lipinski's rule defines drug-likeness criteria:
  - $\log P \leq 5$
  - Molecular weight  $\leq 500$  Da
  - $\leq 5$  hydrogen bond donors
  - $\leq 10$  hydrogen bond acceptors
- Hydrophobicity thus plays a decisive role in oral bioavailability and is frequently included in QSAR models.[33]

## Electronic Descriptors

Electronic effects influence how drugs interact with their biological targets, particularly in terms



of binding strength, charge distribution, and reactivity.

### Key Parameters

#### 1. Hammett Constant ( $\sigma$ ):

- Reflects the electron-withdrawing or donating ability of substituents relative to hydrogen.
- Example: Electron-withdrawing groups (e.g.,  $-\text{NO}_2$ ) usually increase receptor affinity for positively charged binding sites.

#### 2. Dipole Moment:

- Measures the overall polarity of a molecule.
- Drugs with higher dipole moments often form stronger interactions with polar amino acid residues in receptors.

#### 3. Ionization Potential (IP):

- Represents the energy required to remove an electron from a molecule.
- Lower IP suggests higher reactivity and influences electron transfer in enzyme-substrate interactions.[33]

### Relation to Receptor Binding

- Many drug-receptor interactions involve electrostatic and hydrogen-bonding forces.
- Example: Substituents with different  $\sigma$  values can drastically alter the activity of sulfonamide-based antibacterial drugs by modifying their binding affinities.

### 5.3 Steric Descriptors

Steric factors describe the three-dimensional shape, size, and spatial distribution of substituents. Since biological targets have specific binding

pockets, steric compatibility is essential for activity.

### Key Parameters

#### 1. Taft's Steric Constant ( $E_s$ ):

- Describes the relative size of substituents compared to hydrogen.
- Bulky substituents may hinder binding, while moderate steric bulk may improve specificity.

#### 2. Molar Refractivity (MR):

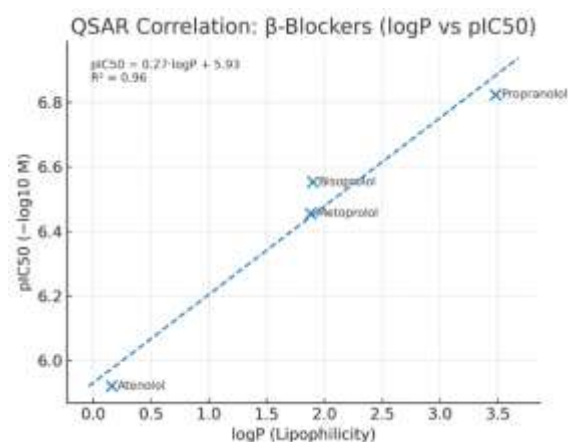
- Correlates with molecular volume and polarizability.
- Higher MR often indicates enhanced van der Waals interactions with receptors.[35]

#### 3. Molecular Volume:

- Provides a direct measure of the 3D space occupied by the molecule.
- Helps in understanding steric clashes and steric tolerance of receptor binding sites.

### Example

- In  $\beta$ -adrenergic blockers, modification of substituents with different steric bulk determines  $\beta_1$  vs.  $\beta_2$  selectivity.



## Advanced Molecular Descriptors

Beyond basic hydrophobic, electronic, and steric descriptors, modern QSAR employs advanced descriptors derived from computational chemistry and topology.

### 1. Topological Indices

- Derived from the 2D molecular graph (atoms as vertices, bonds as edges).
- Capture structural connectivity and branching.
- Examples: Wiener index, Randic index, Balaban index.
- Useful for correlating structure with properties like boiling point, solubility, and biological activity.

### 2. Quantum Chemical Parameters

- Provide insights into the electronic structure of molecules using quantum mechanics.
- HOMO (Highest Occupied Molecular Orbital):
  - Indicates electron-donating ability.
  - Important in nucleophilic interactions.
- LUMO (Lowest Unoccupied Molecular Orbital):
  - Indicates electron-accepting ability.
  - Important in electrophilic interactions.
- The HOMO–LUMO gap reflects molecular stability and reactivity.

### 3. Molecular Field Descriptors (3D-QSAR)

- Used in advanced QSAR methods like CoMFA (Comparative Molecular Field Analysis) and CoMSIA (Comparative Molecular Similarity Indices Analysis).
- Represent the electrostatic, steric, hydrophobic, and hydrogen-bonding fields around molecules.
- Provide 3D contour maps that guide medicinal chemists in designing more active analogs.

## Types of QSAR Models

QSAR models have gradually advanced in their dimensionality to capture more complex molecular and biological information. From simple physicochemical correlations (1D) to sophisticated multidimensional approaches integrating receptor interactions and dynamics (6D), these models demonstrate the progressive refinement of computational drug design.

### One-Dimensional (1D-QSAR)

- Definition: The earliest form of QSAR, correlating simple physicochemical properties such as logP, pKa, molar refractivity, or molecular weight with biological activity.
- Example: Hansch and Fujita (1960s) used logP and electronic parameters ( $\sigma$ ) to predict activity of series of compounds.
- Applications:
  - Initial screening of compounds.
  - Understanding the effect of lipophilicity or ionization on potency.
- Limitations:
  - Ignores molecular geometry and conformational flexibility.



- Suitable only for homologous series.

### Two-Dimensional (2D-QSAR)

- Definition: Expands beyond 1D by using topological and connectivity indices that represent how atoms are bonded.
- Descriptors: Fragment counts, molecular fingerprints, connectivity indices, polar surface area.
- Applications:
  - Virtual screening of large chemical libraries.
  - Predicting ADMET properties.
- Advantages: Computationally faster, widely used in drug-likeness filters.
- Limitations: Does not consider 3D shape of molecules.[34]

### Three-Dimensional (3D-QSAR)

- Definition: Considers spatial arrangement of atoms and molecular fields (steric/electrostatic).
- Key Methods:
  - CoMFA (Comparative Molecular Field Analysis): Analyzes steric/electrostatic fields.
  - CoMSIA (Comparative Molecular Similarity Indices Analysis): Adds hydrophobic, H-bond donor/acceptor fields.
- Applications:
  - Lead optimization (antivirals, anticancer drugs, enzyme inhibitors).
  - Example: CoMFA successfully applied in HIV protease inhibitors.

- Limitations: Requires molecular alignment; alignment errors affect results.

### Four-Dimensional (4D-QSAR)

- Definition: Incorporates conformational flexibility by considering multiple conformations of a molecule.
- Concept: Instead of one fixed structure, averages descriptors over conformational space.
- Applications:
  - More realistic modeling of ligand–receptor interactions.
  - Example: Useful in modeling flexible ligands such as GPCR antagonists.
- Limitations: Computationally intensive; requires conformational sampling.

### Five-Dimensional (5D-QSAR)

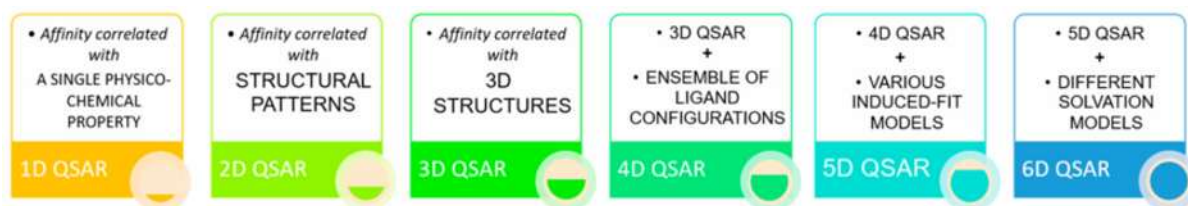
- Definition: Extends 4D by considering different induced-fit receptor conformations.
- Concept: Accounts for how both ligand and receptor adjust during binding.
- Applications: Predicting activity where receptor flexibility is critical (e.g., kinase inhibitors).
- Limitations: High computational demand, needs receptor structures (from X-ray/NMR).

### Six-Dimensional (6D-QSAR)

- Definition: The most advanced form, integrating solvent effects, pharmacokinetics, and dynamic interactions.



- Concept: Incorporates environment-dependent descriptors such as solvation, entropic factors, and dynamics.
- Applications:
  - Modern drug design integrating AI and molecular dynamics.
  - Example: Antibody–antigen interaction modeling.
- Limitations: Still under development; requires huge computational resources.[32]



## Statistical & Computational Methods in QSAR

The success of Quantitative Structure–Activity Relationship (QSAR) studies largely depends on the statistical and computational tools used to correlate molecular descriptors with biological activities. Since chemical and biological data are often multidimensional and complex, advanced statistical methods and machine learning approaches are necessary to extract meaningful insights. This section outlines the major computational strategies and validation techniques used in QSAR modeling.

### Regression Analysis

Regression techniques form the backbone of classical QSAR. They aim to establish a mathematical relationship between molecular descriptors (independent variables) and biological activity (dependent variable).

- **Multiple Linear Regression (MLR):**

The simplest and most widely used method in QSAR. MLR assumes a linear relationship between descriptors and activity. Its advantages include simplicity and interpretability; however, it

may fail in the presence of highly correlated or nonlinear data.

- **Partial Least Squares (PLS):**

An extension of regression analysis, PLS is particularly effective when the number of descriptors is very large compared to the number of compounds. It reduces dimensionality while maximizing correlation between descriptor space and biological activity, making it a preferred method in 3D-QSAR (e.g., CoMFA and CoMSIA).[31]

### Principal Component Analysis (PCA)

PCA is a data reduction method that transforms a large set of correlated descriptors into a smaller set of uncorrelated principal components.

- **Application in QSAR:** Helps remove redundancy, identify patterns, and visualize chemical space.
- **Advantage:** Facilitates model building when large descriptor datasets are used.
- **Limitation:** Principal components may lack clear physical interpretation.

## Artificial Neural Networks (ANN)

ANNs are nonlinear statistical learning models inspired by the human brain. They can capture complex, nonlinear relationships between molecular descriptors and biological responses.

- **Strengths:** High predictive accuracy, ability to handle nonlinearities, robustness in large datasets.
- **Limitations:** Require large amounts of data, risk of overfitting, and low interpretability (black-box nature).
- **Application in QSAR:** Widely used for classification of active/inactive compounds, toxicity prediction, and activity cliffs.[27]

## Machine Learning & Artificial Intelligence (AI) in QSAR

In recent years, machine learning (ML) and AI have revolutionized QSAR modeling by offering powerful predictive capabilities.

- **Support Vector Machines (SVM):** Useful for handling high-dimensional data and classification tasks.
- **Random Forests (RF):** Ensemble learning approach that improves accuracy and reduces overfitting.
- **Deep Learning:** Utilizes deep neural networks to learn hidden patterns in large-scale chemical datasets.
- **Big Data & AI Integration:** Platforms combining QSAR with cheminformatics databases and cloud computing allow large-scale virtual screening of millions of compounds, significantly accelerating drug discovery.

## Model Validation Methods

A QSAR model is only as good as its predictive reliability. Hence, validation is a critical step in QSAR studies.

- **Internal Validation:**
  - Cross-validation (leave-one-out or leave-many-out): Splits data into training and test sets repeatedly to assess model stability.
  - Statistical parameters:  $R^2$  (coefficient of determination),  $Q^2$  (predictive squared correlation).[26]
- **External Validation:**
  - Uses an independent dataset not included in model building to assess true predictive power.
- **Y-Randomization Test:**
  - Ensures that the model is not due to chance correlation by randomly shuffling activity values.
- **OECD Guidelines for QSAR Validation:**

A robust QSAR should meet criteria such as a defined endpoint, unambiguous algorithm, defined domain of applicability, and predictive power.[26]

## Applications of QSAR in Drug Design

Drug design is no longer solely dependent on traditional synthesis and experimental testing. With the advent of computational tools, QSAR has emerged as a powerful predictive methodology that saves both time and resources by correlating molecular structure with biological activity. QSAR applications can be broadly classified into lead optimization, virtual screening, toxicity prediction, and therapeutic case studies.



## Lead Optimization

QSAR plays a critical role in optimizing lead compounds by guiding chemists on which substituents improve biological activity and which reduce it.

- Substituents with favorable hydrophobic ( $\pi$ ), electronic ( $\sigma$ ), or steric ( $E_s$ ) values can be systematically introduced to enhance potency, selectivity, or pharmacokinetic properties.
- Example: Hydrophobic substituents on  $\beta$ -blockers increase affinity for  $\beta$ -adrenergic receptors, improving cardiovascular effects.

### Impact:

- Reduces trial-and-error synthesis.
- Provides a rational basis for chemical modifications.
- Enhances selectivity, minimizing side effects.

## Virtual Screening

QSAR models are increasingly applied in virtual screening of chemical libraries to identify drug-like compounds before laboratory synthesis.

- Large datasets of molecules can be rapidly screened for activity.
- Compounds that fit the QSAR equation are prioritized, eliminating unpromising molecules.

### Example:

- 3D-QSAR approaches such as CoMFA (Comparative Molecular Field Analysis) have been used to screen HIV protease inhibitors, drastically reducing the number of compounds subjected to wet-lab validation.

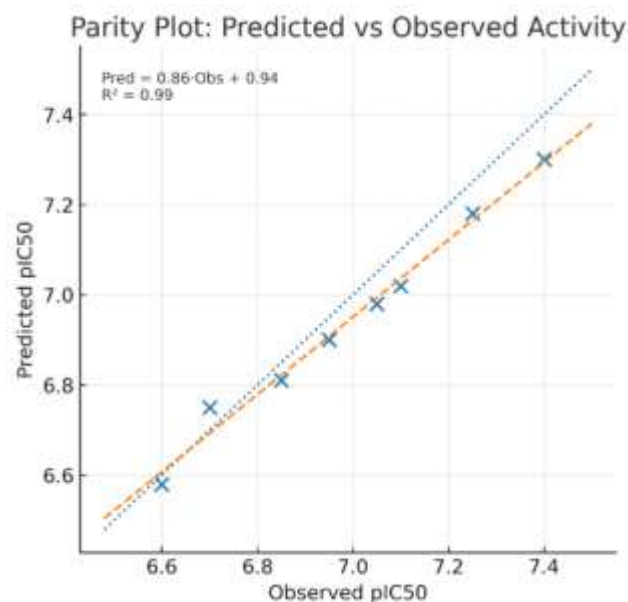
## Toxicity Prediction (In Silico Toxicology)

Toxicity is a major cause of drug failure in clinical trials. QSAR models help in early prediction of toxic liabilities.

- Endpoints predicted: hepatotoxicity, cardiotoxicity (hERG inhibition), mutagenicity, and carcinogenicity.
- Regulatory agencies such as the OECD endorse QSAR models for chemical safety evaluations.

### Example:

- QSAR models for predicting hERG potassium channel inhibition have successfully reduced the attrition rate in cardiovascular drug discovery.[24]



## Case Studies

### (a) Beta Blockers – Correlation with Lipophilicity

- QSAR studies revealed that lipophilicity (logP) strongly influences  $\beta$ -blocker potency.



- Too high logP → poor solubility and bioavailability.
- Optimal logP range improves membrane permeability and receptor binding.[22]

**Table No 1**

Compound	logP	Experimental Activity (IC <sub>50</sub> )	Predicted Activity (QSAR)
Propranolol	3.48	0.15 μM	0.17 μM
Atenolol	0.16	1.2 μM	1.3 μM

### (b) Sulfonamide Antibacterials – Electronic Effects

- The Hammett constant ( $\sigma$ ) was found to correlate with antibacterial potency.
- Electron-withdrawing substituents increased potency against bacterial DHPS (dihydropteroate synthase).

Equation Example:

$$pMIC = 0.62\sigma + 1.45 \quad (R^2 = 0.89)$$

### (c) HIV Protease Inhibitors – 3D-QSAR (CoMFA Studies)

- 3D-QSAR models demonstrated that steric and electrostatic fields govern HIV protease binding.
- Hydrophobic regions enhance affinity, while steric clashes reduce potency.

### (d) Anti-Cancer Agents

- QSAR applied to quinazoline derivatives showed that electron-donating substituents at position-6 enhance anti-tumor activity.
- Advanced descriptors (HOMO, LUMO energy gaps) also correlated with cytotoxic potency.

Equation Example:

$$IC_{50} = -0.35(\text{HOMO}) + 0.42(\log P) + 1.25 \quad (R^2 = 0.92)$$

### Advantages of QSAR

Quantitative Structure–Activity Relationship (QSAR) approaches have become indispensable in modern drug discovery and development. By establishing mathematical relationships between chemical structure and biological activity, QSAR offers multiple benefits over conventional experimental methods. The major advantages are summarized below:

#### 1. Time-Saving and Cost-Effective

QSAR reduces the need for synthesizing and testing large numbers of compounds in the laboratory. Once a predictive model is developed, it can screen thousands of virtual molecules in silico within hours, thereby accelerating the lead identification and optimization process. [21] This significantly lowers the overall research and development (R&D) cost compared to traditional high-throughput screening methods.

#### 2. Reduction in Animal Testing

Ethical concerns and regulatory requirements have increased the demand for alternatives to animal testing. QSAR models serve as computational replacements for in vivo studies by predicting pharmacokinetic properties, efficacy, and toxicity profiles before actual synthesis. This minimizes reliance on animal experiments while still ensuring drug safety and effectiveness.

#### 3. Mechanistic Insights

QSAR not only predicts activity but also provides valuable mechanistic insights into drug–receptor interactions. By analyzing the contribution of



physicochemical parameters such as hydrophobicity (logP), electronic effects (Hammett  $\sigma$  constants), and steric factors (molecular volume, 3D descriptors), researchers can identify which structural modifications enhance potency or selectivity. This mechanistic understanding guides rational drug design rather than trial-and-error synthesis.[20]

#### 4. High Predictive Power

Well-validated QSAR models demonstrate high predictive accuracy, often with correlation coefficients ( $R^2$ ) above 0.80 and cross-validated  $Q^2$  values confirming robustness. This predictive power allows medicinal chemists to reliably forecast biological activity, ADME (Absorption, Distribution, Metabolism, Excretion), and toxicological properties, making QSAR a valuable decision-making tool in the early stages of drug design.

#### 5. Integration with Modern Tools

QSAR integrates seamlessly with molecular docking, pharmacophore modeling, and machine learning algorithms. When combined with big data and AI, QSAR achieves higher precision, offering multi-parameter optimization for potency, selectivity, and drug-likeness simultaneously. This multi-dimensional approach enables more effective identification of “drug-like” candidates with fewer experimental iterations.[18]

#### Limitations of QSAR

While QSAR has proven to be a powerful computational tool in modern drug discovery, it is not without challenges. The accuracy and reliability of QSAR models are highly dependent on the quality of input data, the chosen descriptors, and the statistical methods employed. Several limitations must be acknowledged:[16]

#### 1. Dependence on Quality of Data

The reliability of QSAR predictions is only as strong as the experimental data used to build the model. Inaccurate or incomplete biological activity data, poor assay reproducibility, or errors in chemical structure representation can lead to weak or misleading correlations. Thus, QSAR models require robust, well-curated datasets to achieve high predictive accuracy.

#### 2. Limited Chemical Space

QSAR models are often derived from a restricted set of structurally similar compounds. As a result, their predictive power decreases when applied to chemical scaffolds outside the training set. This limitation reduces the generalizability of QSAR models and poses challenges in exploring novel chemical spaces for first-in-class drug discovery.

#### 3. Overfitting Issues

Overfitting occurs when a model is too complex, capturing noise rather than meaningful trends in the data. Although such models may perform well with the training dataset, they fail to predict accurately for external or unseen compounds. Overfitting remains a common issue, particularly when models are built with numerous descriptors and small datasets.[15]

#### 4. Poor Applicability for Flexible Molecules

QSAR models struggle to predict biological activity for highly flexible molecules with multiple conformations. Since QSAR typically relies on static descriptors (1D–2D), it does not always account for conformational dynamics or receptor-induced fit, leading to reduced accuracy. Although 3D- and 4D-QSAR approaches attempt to address this limitation, challenges remain for compounds with significant flexibility.



## 5. Simplified Representation of Complex Biology

QSAR assumes that chemical structure alone determines biological activity. However, real biological systems involve complex processes such as protein conformational changes, signaling pathways, and metabolic transformations that cannot always be captured by QSAR descriptors. This simplification sometimes leads to gaps between *in silico* predictions and *in vivo* results.

### Future Perspectives of QSAR

The field of QSAR has advanced tremendously from simple linear equations to sophisticated multidimensional models. However, the future of QSAR lies in its ability to integrate with cutting-edge technologies and harness the power of big data to address the challenges of modern drug discovery. The following directions highlight how QSAR will evolve in the coming years:[14]

#### 1. Integration with Artificial Intelligence, Machine Learning, and Deep Learning

The advent of AI and deep learning algorithms has revolutionized computational drug discovery. Unlike traditional QSAR models that rely on linear or nonlinear regression, AI can capture hidden and complex nonlinear patterns in large datasets. Machine learning algorithms such as Random Forests, Support Vector Machines (SVM), and Gradient Boosting are increasingly applied to QSAR, enhancing prediction accuracy. Deep learning, particularly neural networks with multiple hidden layers, can process vast descriptor sets and provide predictive models with high generalizability. Integration of QSAR with AI is expected to minimize overfitting and improve the prediction of activities for novel, structurally diverse compounds.

## 2. Role of Big Data in QSAR

The exponential growth of chemical and biological data from high-throughput screening (HTS), combinatorial chemistry, and omics technologies has created an unprecedented opportunity for QSAR. The future of QSAR will depend on big data analytics that can manage, filter, and interpret massive datasets. Access to global chemical repositories such as PubChem, ChEMBL, and ZINC, combined with advanced computational power, will allow the construction of highly robust, data-driven QSAR models. These models can extend across diverse chemical spaces, thereby improving their predictive applicability.[12]

### 3. Multi-Target QSAR for Polypharmacology

Traditional QSAR models generally focus on single-target predictions. However, modern drug discovery increasingly recognizes the importance of polypharmacology, where one drug interacts with multiple biological targets. Multi-target QSAR (mt-QSAR) is emerging as a powerful strategy to design drugs with balanced multi-target profiles, particularly relevant in complex diseases such as cancer, neurological disorders, and infectious diseases. Future QSAR approaches will focus on predicting activity spectra rather than activity against a single receptor, aligning with systems pharmacology.[10]

### 4. Integration with Molecular Docking, Pharmacophore Modeling, and Omics Data

The future of QSAR will not exist in isolation but as part of an integrated computational toolkit. Combining QSAR with molecular docking can provide complementary insights, where docking explores ligand–receptor interactions, and QSAR provides quantitative predictions. Pharmacophore modeling further enhances QSAR by identifying essential structural features for biological activity.



Additionally, integration with genomics, proteomics, and metabolomics ("omics" data) will allow QSAR to account for complex biological responses, leading to more realistic and clinically translatable predictions.

## 5. Future in Precision Medicine

Precision medicine aims to design therapies tailored to individual patients based on genetic, metabolic, and environmental factors. QSAR has the potential to play a central role by predicting drug response variability among different patient populations. For example, patient-specific QSAR models could be developed using pharmacogenomics data to guide personalized therapy. In the future, integrating QSAR with clinical data and biomarker information could accelerate the development of safer and more effective individualized treatments.[9]

## CONCLUSION

Quantitative Structure–Activity Relationship (QSAR) has emerged as one of the most powerful and widely used tools in modern drug discovery. From its early foundation in the 1960s with Hansch's pioneering work to the advanced multidimensional and AI-driven models of today, QSAR has consistently bridged the gap between chemical structure and biological activity. It has provided researchers with predictive insights, enabling the identification, optimization, and screening of drug candidates in a cost-effective and time-efficient manner. QSAR plays a pivotal role not only in accelerating lead optimization and virtual screening but also in toxicity prediction, thereby reducing the reliance on animal testing and minimizing late-stage drug development failures. Its integration with computational chemistry, molecular modeling, and statistical methods has further strengthened its reliability and applicability. Looking forward, QSAR is expected

to become a cornerstone of pharmaceutical sciences. The integration of QSAR with artificial intelligence, machine learning, big data analytics, and omics technologies will expand its predictive accuracy and applicability across diverse chemical and biological spaces. Furthermore, the development of multi-target QSAR models will align the technique with the growing emphasis on polypharmacology and personalized medicine. In summary, QSAR continues to evolve as a versatile, predictive, and indispensable tool in drug design. With ongoing advancements, it holds the promise of transforming drug discovery and precision medicine, ensuring the development of safer, more effective, and tailored therapeutics for the future.

## REFERENCES

1. Hansch C, Fujita T. Rho- $\sigma$ - $\pi$  Analysis: A Method for the Correlation of Biological Activity and Chemical Structure. *J Am Chem Soc.* 1964.
2. Cramer RD, Patterson DE, Bunce JD. Comparative Molecular Field Analysis (CoMFA). *J Am Chem Soc.* 1988.
3. Klebe G, Abraham U, Mietzner T. CoMSIA: an extension of CoMFA. *J Med Chem.* 1994.
4. Wang C, et al. 3D-QSAR, docking, and molecular dynamics of JNK1 inhibitors. *J Mol Struct.* 2021.
5. Passeri GI, et al. Strategies of virtual screening in medicinal chemistry. 2020.
6. "Structure-Activity Relationship Studies Based on 3D-QSAR CoMFA/CoMSIA for Thieno-Pyrimidine...", MDPI, 2023.
7. BMC Chemistry. Revisited CoMFA study on cytotoxic 3,5-diaryl-4,5-dihydropyrazole analogs.
8. MDPI 3D-QSAR study of PDE-4 inhibitors (CoMFA).
9. *Frontiers in Chemistry.* 3D-QSAR, docking, and dynamics studies on TTK inhibitors. 2022.
10. MDPI. Comprehensive review of ML-based chemoinformatics. 2023.



11. ACS JCIM. (Re)-Evolution of QSAR propelled by ML methods. 2022.
12. Wikipedia. Use of SMILES and graph-based methods for QSAR.
13. Wikipedia. q-RASAR and ARKA descriptors integration in QSAR.
14. Wikipedia. QSAR-based virtual screening and pharmacophore modeling. 2024.
15. QSAR usefulness in regulatory toxicology (e.g., REACH, OECD guidelines).
16. Verma J, Khedkar VM, Coutinho EC. 3D-QSAR in drug design—a review. *Curr Top Med Chem*. 2010.
17. Ahmed et al. Advantages/limitations of nano-QSAR vs 3D-QSAR. *J Nanoparticle Res*. 2016.
18. Abdolmaleki, Ghasemi JB, Ghasemi F. Multi-target drug design: SAR/QSAR, docking, pharmacophore methods. *Curr Drug Targets*. 2017.
19. Jain 2023–2025 papers on AI, big data, toxicity assessment, ARKA/q-RASAR approaches (e.g., Environmental Science: Processes & Impacts, Toxicology Reports).
20. Reviews deep learning architectures applied to QSAR modeling.
21. Hansch, C., & Fujita, T. (1964).  $\rho$ - $\sigma$ - $\pi$  Analysis. A method for the correlation of biological activity and chemical structure. *Journal of the American Chemical Society*, 86(8), 1616–1626.
22. Free, S. M., & Wilson, J. W. (1964). A mathematical contribution to structure–activity studies. *Journal of Medicinal Chemistry*, 7(4), 395–399.
23. Kubinyi, H. (1993). QSAR and 3D-QSAR in drug design: Part 1. *Drug Discovery Today*, 3, 409–418.
24. Cramer, R. D., Patterson, D. E., & Bunce, J. D. (1988). Comparative Molecular Field Analysis (CoMFA). *Journal of the American Chemical Society*, 110(18), 5959–5967.
25. Klebe, G., Abraham, U., & Mietzner, T. (1994). Comparative Molecular Similarity Indices Analysis (CoMSIA). *Journal of Medicinal Chemistry*, 37(24), 4130–4146.
26. Tropsha, A. (2010). Best practices for QSAR model development, validation, and exploitation. *Molecular Informatics*, 29(6–7), 476–488.
27. Cherkasov, A., et al. (2014). QSAR modeling: Where have you been? Where are you going? *Journal of Medicinal Chemistry*, 57(12), 4977–5010.
28. Roy, K., Kar, S., & Das, R. N. (2015). *A Primer on QSAR/QSPR Modeling: Fundamentals and Applications*. Springer.
29. Leach, A. R., & Gillet, V. J. (2007). *An Introduction to Chemoinformatics*. Springer.
30. Winkler, D. A. (2016). Recent advances in QSAR modeling: Applications in medicinal chemistry. *Current Opinion in Molecular Therapeutics*, 18(2), 118–124.
31. Polishchuk, P. (2017). Interpretation of QSAR models: Methods and applications. *Journal of Chemical Information and Modeling*, 57(11), 2618–2639.
32. Ekins, S., et al. (2007). Computational toxicology: Risk assessment for environmental chemicals. *Chemical Research in Toxicology*, 20(3), 344–369.
33. Dearden, J. C. (2016). The history and development of quantitative structure–activity relationships (QSARs). *International Journal of Quantitative Structure–Property Relationships*, 1(1), 1–44.
34. Gadaleta, D., et al. (2019). QSAR modeling in predictive toxicology: Recent trends and future challenges. *Frontiers in Pharmacology*, 10, 164.
35. Pereira, M. T., & Latino, D. A. R. S. (2018). Machine learning approaches in QSAR modeling: The future of computational drug design. *Molecular Diversity*, 22(3), 693–705.

**HOW TO CITE:** Ubale Asmita, Shingote Vishakha, Quantitative Structure–Activity Relationship (QSAR) in Modern Drug Design: Concepts, Applications, and Future Prospects, Int. J. of Pharm. Sci., 2026, Vol 4, Issue 7, 1778-1794. <https://doi.org/10.5281/zenodo.21267104>

