Quantitative Structure–Activity Relationship (QSAR): A Review for Beginners to Intermediate Researchers

1.Mrs.MOUNIKA K, 2.Dr. R. THIRUMURTHY, 3.Mrs. ANNALAKSHMI S, 4.Mrs.PORSELVI R

How to Cite this Article:

K, M., S, A. & R, P. (2026). Quantitative Structure–Activity Relationship (QSAR): A Review for Beginners to Intermediate Researchers. International Journal of Creative and Open Research in Engineering and Management, 02(03). https://doi.org/10.55041/ijcope.v2i3.049

https://doi.org/10.55041/ijcope.v2i3.049

Abstract

Quantitative Structure–Activity Relationship (QSAR) is an important computational technique used in modern drug discovery to predict the biological activity of chemical compounds based on their molecular structures. QSAR models establish mathematical relationships between chemical structure descriptors and biological activities. This method significantly reduces the cost, time, and resources required for experimental screening of large numbers of compounds. QSAR plays a major role in medicinal chemistry, toxicology, and environmental chemistry by helping researchers understand how structural features influence biological activity. The development of QSAR involves dataset preparation, calculation of molecular descriptors, model development, validation, and prediction. Modern QSAR approaches also integrate machine learning and artificial intelligence to improve prediction accuracy. This review article provides a clear and systematic overview of QSAR concepts, types of QSAR models, molecular descriptors, model development processes, validation techniques, applications, advantages, and limitations. The article is intended to provide fundamental knowledge of QSAR for learners and moderate-level researchers in the fields of medicinal chemistry and computational drug design.

Keywords: QSAR, drug discovery, molecular descriptors, computational chemistry, structure–activity relationship

1. Introduction

Drug discovery is a complex and time-consuming process that requires extensive experimental testing of chemical compounds. Computational methods have greatly improved the efficiency of this process by predicting biological activity before laboratory experiments are conducted. One of the most widely used computational techniques is **Quantitative Structure–Activity Relationship (QSAR)**.

QSAR is based on the principle that the **biological activity of a compound is related to its chemical structure**. By analyzing structural features of molecules, QSAR models can predict the activity of new compounds even before synthesis or biological testing.

The concept of QSAR evolved from **Structure–Activity Relationship (SAR)** studies. SAR focuses on qualitative relationships between chemical structures and biological activity, whereas QSAR provides **quantitative mathematical relationships**.

The modern QSAR approach was developed by pioneers such as **Corwin Hansch** and **Toshio Fujita**, who introduced mathematical models correlating physicochemical properties with biological activities.

QSAR methods are widely used in **medicinal chemistry, environmental toxicology, agrochemical research, and drug design**. With the advancement of computational tools and machine learning techniques, QSAR has become an essential component of modern drug discovery pipelines.

2. Basic Principles of QSAR

The fundamental principle of QSAR is that **molecular structure determines biological activity**. Chemical compounds contain structural features such as functional groups, electronic distribution, and molecular size that influence their interaction with biological targets.QSAR models attempt to establish a mathematical relationship between these structural features and biological activity.

The general QSAR workflow can be summarized as:

Chemical Structure → Molecular Descriptors → Mathematical Model → Biological Activity Prediction

Molecular descriptors represent numerical values that describe physicochemical properties of molecules. These descriptors are used to construct mathematical models that correlate with biological activity.

One of the earliest QSAR models is the **Hansch equation**, which correlates biological activity with hydrophobicity, electronic effects, and stearic factors. The general form of the equation is:

$$\text{Activity} = a(\log P) + b(\sigma) + c(Es) + \text{constant}$$

Where:

- **logP** represents hydrophobicity
- **σ** represents electronic parameters
- **Es** represents stearic parameters

These parameters help determine how molecules interact with biological targets such as enzymes or receptors.

3. Types of QSAR Models

QSAR models can be classified into several categories depending on the type of descriptors and dimensionality used.

3.1 One-Dimensional QSAR (1D-QSAR)

1D-QSAR models use simple physicochemical properties of molecules. These include parameters such as:

- Molecular weight
- LogP (lipophilicity)
- pKa
- Polar surface area

These descriptors provide basic information about the compound and are useful for simple predictive models.

3.2 Two-Dimensional QSAR (2D-QSAR)

2D-QSAR models use structural information derived from molecular connectivity. These descriptors include:

- Topological descriptors
- Fragment descriptors
- Connectivity indices

2D-QSAR models are widely used because they do not require three-dimensional molecular structures and are relatively easy to compute.

3.3 Three-Dimensional QSAR (3D-QSAR)

3D-QSAR models incorporate the three-dimensional structure of molecules. These models analyze steric and electrostatic interactions between ligands and biological targets.

Two widely used 3D-QSAR techniques include:

- Comparative Molecular Field Analysis (CoMFA)

- Comparative Molecular Similarity Indices Analysis (CoMSIA)

These techniques help identify regions of molecules that contribute positively or negatively to biological activity.

4. Molecular Descriptors

Molecular descriptors are numerical values representing chemical information encoded within molecular structures. They play a crucial role in QSAR modeling.

Descriptors can be classified into several categories:

4.1 Physicochemical Descriptors

These describe physical and chemical properties of molecules.

Examples include:

- Molecular weight
- LogP
- Solubility
- Polar surface area

4.2 Topological Descriptors

Topological descriptors are derived from molecular connectivity and represent the arrangement of atoms within a molecule.

Examples include:

- Connectivity indices
- Wiener index
- Molecular branching

4.3 Geometrical Descriptors

These descriptors describe the three-dimensional shape and size of molecules.

Examples include:

- Molecular volume
- Surface area
- Shape indices

4.4 Electronic Descriptors

Electronic descriptors represent electronic properties of molecules.

Examples include:

- Dipole moment
- HOMO–LUMO energy
- Partial atomic charges

Several software tools are available for calculating molecular descriptors, including:

- PaDEL-Descriptor
- Molinspiration
- Dragon

These tools allow researchers to calculate hundreds or even thousands of descriptors for a single compound.

5. QSAR Model Development

Developing a reliable QSAR model involves several important steps.

5.1 Dataset Collection

The first step is collecting a dataset of chemical compounds with known biological activity values. These activities may include IC50, EC50, or inhibition constants.

5.2 Molecular Structure Optimization

The chemical structures of compounds are optimized using computational chemistry software to obtain stable conformations.

5.3 Descriptor Calculation

Next, molecular descriptors are calculated using cheminformatics software.

5.4 Descriptor Selection

Since many descriptors may be generated, statistical methods are used to select the most relevant descriptors. Common techniques include:

- Stepwise regression
- Principal component analysis
- Genetic algorithms

5.5 Model Building

Statistical methods are used to build QSAR models. Common techniques include:

- Multiple Linear Regression (MLR)
- Partial Least Squares (PLS)
- Machine learning algorithms

6. Validation of QSAR Models

Validation is essential to ensure that QSAR models provide reliable predictions.

Two main types of validation are used:

6.1 Internal Validation

Internal validation assesses the predictive ability of the model using the same dataset. Cross-validation methods such as leave-one-out (LOO) are commonly used.

Important parameters include:

- R^2 (coefficient of determination)
- Q^2 (cross-validated correlation coefficient)

6.2 External Validation

External validation uses an independent dataset not involved in model development. This step provides a more accurate measure of predictive performance.

Other statistical parameters used in validation include:

- Root Mean Square Error (RMSE)
- Mean Absolute Error (MAE)

7. Applications of QSAR

QSAR has numerous applications in pharmaceutical and chemical research.

7.1 Drug Discovery

QSAR helps identify potential drug candidates by predicting biological activity before synthesis.

7.2 Toxicity Prediction

QSAR models are widely used to predict the toxicity of chemicals and pharmaceuticals.

7.3 Environmental Chemistry

QSAR helps predict environmental behavior of chemicals, including biodegradation and ecological toxicity.

7.4 Lead Optimization

QSAR assists medicinal chemists in modifying chemical structures to improve potency, selectivity, and pharmacokinetic properties.

For example, QSAR studies are often conducted to predict inhibitors of enzymes such as **Acetylcholinesterase**, which is an important target in neurodegenerative disease research.

8. Advantages and Limitations of QSAR

Advantages

QSAR offers several benefits in drug discovery and chemical research:

- Reduces cost and time required for experimental screening
- Predicts biological activity of new compounds
- Helps understand structure–activity relationships
- Supports rational drug design

Limitations

Despite its advantages, QSAR also has limitations:

- Requires high-quality experimental data
- Models may not perform well outside their applicability domain
- Complex biological systems may not be fully captured by simple models

9. Future Perspectives

Recent advancements in artificial intelligence and machine learning have significantly enhanced QSAR modeling. Modern approaches integrate deep learning algorithms with large chemical datasets to improve predictive accuracy.

Machine learning techniques such as random forests, support vector machines, and neural networks are increasingly used in QSAR modeling. These methods allow researchers to analyze complex nonlinear relationships between molecular descriptors and biological activity.

In the future, QSAR models will likely become more accurate and widely integrated with other computational methods such as molecular docking, molecular dynamics simulations, and virtual screening.

10. Conclusion

Quantitative Structure–Activity Relationship (QSAR) is a powerful computational tool widely used in modern drug discovery and chemical research. By establishing mathematical relationships between molecular structure and biological activity, QSAR enables researchers to predict the activity of new compounds efficiently.

The QSAR modeling process involves dataset preparation, descriptor calculation, model development, and validation. Advances in computational techniques and machine learning have significantly improved the predictive capabilities of QSAR models.

Overall, QSAR remains an essential approach in medicinal chemistry, providing valuable insights into structure–activity relationships and supporting the development of safer and more effective drugs.

References

1. Corwin Hansch, A. Leo, and D. Hoekman. (1995). *Exploring QSAR: Fundamentals and Applications in Chemistry and Biology*. American Chemical Society.
2. Toshio Fujita. (1964). The analysis of structure-activity relationships. **Journal of the American Chemical Society**, 86, 1616-1626.
3. Johann Gasteiger and T. Engel. (2003). *Chemoinformatics: A Textbook*. Wiley-VCH.
4. Cherkasov, A., et al. (2014). QSAR modeling: Where have you been? Where are you going to? **Journal of Medicinal Chemistry**, 57, 4977-5010.
5. Todeschini, R., and V. Consonni. (2009). *Molecular Descriptors for Chemoinformatics*. Wiley-VCH.
6. Roy, K., Kar, S., and R. Das. (2015). *A Primer on QSAR/QSPR Modeling*. Springer.
7. Tropsha, A. (2010). Best practices for QSAR model development and validation. **Molecular Informatics**, 29, 476-488.
8. Gramatica, P. (2007). Principles of QSAR models validation. **QSAR & Combinatorial Science**, 26, 694-701.
9. Kubinyi, H. (1993). QSAR: Hansch analysis and related approaches. **Drug Discovery Today**.
10. Hugo Kubinyi. (1998). QSAR and 3D-QSAR in drug design. **Drug Discovery Today**, 3, 538-546.
11. Verma, J., et al. (2010). QSAR modeling: Concepts and applications. **Chemical Biology & Drug Design**, 76, 186-196.
12. Cherkasov, A. (2005). QSAR methods in drug discovery. **Journal of Chemical Information and Modeling**.

13. Leach, A. R., and V. J. Gillet. (2007). *An Introduction to Chemoinformatics*. Springer.
14. Katritzky, A. R., et al. (2010). QSAR modeling methods. **Chemical Reviews**, 110, 5714-5789.
15. Eriksson, L., et al. (2003). *Multi- and Megavariate Data Analysis*. Umetrics.
16. Golbraikh, A., and A. Tropsha. (2002). Beware of q^2. **Journal of Molecular Graphics and Modelling**, 20, 269-276.
17. Winkler, D. (2004). Neural networks in QSAR modeling. **Journal of Chemical Information and Computer Sciences**.
18. Liu, H., et al. (2017). Machine learning in QSAR modeling. **Current Pharmaceutical Design**.
19. Baskin, I. (2019). Machine learning methods in QSAR. **Molecular Informatics**.
20. Vamathevan, J., et al. (2019). AI in drug discovery. **Nature Reviews Drug Discovery**.
21. Varnek, A., and I. Baskin. (2012). Machine learning in chemoinformatics. **Journal of Chemical Information and Modeling**.
22. Yap, C. W. (2011). PaDEL-Descriptor software. **Journal of Computational Chemistry**.
23. Todeschini, R. (2000). Dragon descriptor software. **Talete SRL**.
24. Todeschini, R., and V. Consonni. (2000). Handbook of molecular descriptors. **Wiley**.
25. Tetko, I. V., et al. (2005). Virtual computational chemistry laboratory. **Journal of Computer-Aided Molecular Design**.
26. Sheridan, R. (2013). Interpretability in QSAR models. **Journal of Chemical Information and Modeling**.
27. Gasteiger, J. (2006). Chemoinformatics in drug discovery. **Wiley**.
28. Todeschini, R., et al. (2004). Topological descriptors in QSAR. **Chemometrics**.
29. Gramatica, P. (2013). On validation of QSAR models. **International Journal of Quantitative Structure-Property Relationships**.
30. Cherkasov, A. (2008). QSAR modeling techniques and applications. **Current Topics in Medicinal Chemistry**.