Genome Publications
https://doi.org/10.61096/978-81-990998-7-6_4

Chapter 4

Classical QSAR and Statistical Models: Correlating Descriptors with Activity and ADMET

Dr. Amareswarapu V Surendra
Assistant Professor, KL college of Pharmacy, Koneru Lakshmaiah Educational Foundation Greenfields, Vaddeswaram, Guntur, Andhra Pradesh-Pin:522302

Dr. Sushma.N
Assistant Professor, KL college of Pharmacy, Koneru Lakshmaiah Educational Foundation Greenfields, Vaddeswaram, Guntur, Andhra Pradesh-Pin:522302

Mandava Mahima Swaroopa
KL college of Pharmacy, Koneru Lakshmaiah Educational Foundation Greenfields, Vaddeswaram, Guntur, Andhra Pradesh-Pin:522302

Abstract: Quantitative Structure Activity Relationship (QSAR) modelling has long served as a cornerstone of computer-aided drug design, linking molecular descriptors to biological activity through statistical and mathematical frameworks. This chapter explores the evolution, theory, and practice of classical QSAR, emphasizing its role in correlating physicochemical and topological parameters with pharmacological responses and ADMET (absorption, distribution, metabolism, excretion, and toxicity) profiles. Beginning with Hansch and Free Wilson models, the discussion extends through regression-based and multivariate methods, focusing on descriptor selection, model building, and validation strategies. Special attention is given to statistical models such as multiple linear regression (MLR), partial least squares (PLS), and principal component analysis (PCA), along with their integration into predictive ADMET modeling. Case studies demonstrate successful QSAR applications in enzyme inhibition, receptor binding, and toxicology screening. The chapter concludes by examining challenges such as overfitting, descriptor redundancy, and interpretability, while projecting how hybrid statistical and machine-learning approaches can enhance predictive reliability and mechanistic insight. By integrating historical rigor with modern computational paradigms, this chapter provides a methodological bridge between early QSAR frameworks and advanced AI-driven drug discovery.

Keywords: QSAR, multiple linear regression, molecular descriptors, ADMET prediction, statistical modelling.

Citation: Amareswarapu V Surendra, Sushma.N, Mandava Mahima Swaroopa. Classical QSAR and Statistical Models: Correlating Descriptors with Activity and ADMET. *Comprehensive Approaches in Computer-Aided Drug Design: QSAR, Docking, Screening, Homology, Pharmacophore and AI-Driven Insights.* Genome Publication. 2025; Pp39-48**.** https://doi.org/10.61096/978-81-990998-7-6_4

4.0 INTRODUCTION

The concept of Quantitative Structure Activity Relationship (QSAR) originated from the fundamental observation that molecular structure governs biological activity. This principle, articulated by Corwin Hansch in the early 1960s, catalyzed a paradigm shift from qualitative medicinal chemistry intuition to quantitative prediction of bioactivity through mathematical models. The Hansch analysis, which related hydrophobicity (logP), electronic, and steric parameters to biological potency, represented the first systematic attempt to encode chemical intuition into regression-based equations capable of predicting pharmacological response. The Free–Wilson model, emerging shortly thereafter, expanded this framework by incorporating indicator variables representing substituent presence or absence at defined molecular positions, allowing additive contributions of functional groups to be quantified. Classical QSAR models served as a foundation for understanding ligand–receptor interactions before the advent of high-resolution crystallography and computational docking. In the absence of detailed structural data, medicinal chemists used QSAR-derived equations to rationalize why structural analogues exhibited variations in potency or selectivity. The earliest models focused on simple physicochemical properties such as lipophilicity, ionization constants (pKa), molar refractivity, and steric bulk, reflecting the available experimental data of the era. These descriptors, though rudimentary by modern standards, captured key features influencing membrane permeability, receptor binding, and metabolic stability.

The historical progression of QSAR mirrors the evolution of statistical methodologies and computational capabilities. Early models were limited to linear correlations due to computational constraints, but subsequent decades witnessed the incorporation of multivariate regression, pattern recognition, and dimensionality reduction techniques. The 1980s saw the rise of Partial Least Squares (PLS) regression and Principal Component Analysis (PCA), which allowed researchers to manage correlated descriptors and interpret latent variable contributions to activity. By the early 2000s, QSAR had become integral to regulatory submissions and toxicity screening under programs such as REACH and OECD guidelines, underscoring its continued relevance in preclinical decision-making. Classical QSAR thus represents both a methodological and philosophical bridge anchored in empirical data yet driven by the aspiration to model biological complexity. Despite the advent of machine learning and deep neural networks, the interpretability, transparency, and mechanistic alignment of classical statistical models continue to offer irreplaceable value in rational drug design and ADMET evaluation.

4.1 Theoretical Foundations of Classical QSAR

At its core, QSAR operates under the assumption that molecular properties can be mathematically correlated with biological activity. The general functional form of a QSAR model can be expressed as:

$$\mathbf{Activity=f(Descriptors)=f(X1,X2,...,Xn)}\ \boldsymbol{Activity=f(Descriptors)=f(X1,X2,...,Xn)}$$

where Xi*Xi* represents a molecular descriptor (e.g., hydrophobic constant, Hammett σ, Taft steric parameter), and *Activity* denotes a measurable biological endpoint such as enzyme inhibition constant (Ki), half-maximal inhibitory concentration (IC50), or receptor binding affinity (pKi). Classical QSAR posits that biological response is a linear or nonlinear combination of structural features encoded numerically. The two foundational formulations Hansch Analysis and Free–Wilson Analysis employ distinct theoretical perspectives. Hansch Analysis treats biological activity as a continuous function of physicochemical parameters, yielding regression equations of the form:

$$\mathbf{log(1/C)=alogP+b(logP)2+c\sigma+dEs+klog(1/}\boldsymbol{C}\mathbf{)=}\boldsymbol{a}\mathbf{log}\boldsymbol{P}\mathbf{+}\boldsymbol{b}\mathbf{(log}\boldsymbol{P}\mathbf{)2+}\boldsymbol{c\sigma}\mathbf{+}\boldsymbol{dEs}\mathbf{+}\boldsymbol{k}$$

where C*C* represents concentration required for biological effect, logPlog*P* denotes lipophilicity, σ*σ* is the Hammett constant (electronic effect), and Es*Es* is the Taft steric parameter. The quadratic term allows for the parabolic relationship often observed between lipophilicity and activity reflecting an optimal balance between membrane permeability and aqueous solubility. In contrast, the Free–Wilson Model assumes additivity of structural contributions without explicitly invoking physicochemical parameters:

Activity=∑aiXi+k*Activity*=∑*aiXi*+*k*

where Xi*Xi* are binary indicators of substituent presence and ai*ai* their corresponding contributions to activity. This approach allows medicinal chemists to infer the significance of specific substituents but lacks mechanistic depth regarding underlying physicochemical mechanisms. The conceptual fusion of these models led to mixed QSAR approaches, wherein both physicochemical and structural indicators were combined to capture synergistic effects. Over time, additional theoretical refinements emerged, such as the introduction of interaction terms, normalization of descriptors, and correction for collinearity, all of which improved the robustness of predictions. QSAR theory also incorporates the assumption of similarity, namely that structurally similar molecules exhibit similar biological activity. This assumption underpins modern chemoinformatics methods such as nearest-neighbor and cluster-based screening. However, it is important to note that the relationship between structure and activity is not always monotonic activity cliffs, tautomerism, and conformational flexibility can disrupt linear correlations, emphasizing the need for careful descriptor selection and model validation.

4.2 Descriptor Selection and Statistical Parameterization

The predictive capacity of a QSAR model is inherently dependent on the choice of descriptors. Descriptors serve as mathematical representations of molecular features that influence pharmacokinetic and pharmacodynamic behavior. Classical QSAR typically employs physicochemical, topological, and electronic descriptors derived from empirical measurements or quantum chemical calculations. Commonly used physicochemical descriptors include hydrophobicity (logP), molecular weight, polar surface area (PSA), molar refractivity, and hydrogen bond donor/acceptor counts. Electronic descriptors such as dipole moment, HOMO–LUMO gap, and charge distribution reflect the electron-withdrawing or -donating tendencies of substituents. Steric descriptors (e.g., Taft's Es*Es*, Verloop's sterimol parameters) quantify three-dimensional volume and shape effects relevant to receptor binding.

A central challenge in descriptor selection lies in balancing comprehensiveness with parsimony. Including too many descriptors can lead to overfitting, where the model captures noise rather than true signal. Conversely, too few descriptors may yield an under-specified model incapable of generalization. Statistical feature selection techniques such as stepwise regression, forward selection, backward elimination, and genetic algorithms have historically been employed to identify the most informative subset of descriptors. To ensure interpretability and orthogonality, intercorrelation matrices and variance inflation factors (VIF) are often computed to eliminate redundant descriptors. The Hansch constant correlation matrix, for instance, provides insights into descriptor dependencies. Additionally, Principal Component Analysis (PCA) is commonly used to transform correlated descriptors into orthogonal components, enabling the construction of simplified models that retain maximal variance.

Modern computational environments such as MOE, KNIME, and QSARINS facilitate descriptor calculation and selection workflows. QSARINS, in particular, provides built-in statistical validation (R^2,

Q^2, RMSE) and applicability domain visualization, allowing researchers to identify chemical space boundaries where predictions are reliable.

4.3 Multiple Linear Regression (MLR) and Model Building

Among classical statistical methods, Multiple Linear Regression (MLR) remains the most widely used for QSAR analysis due to its simplicity, interpretability, and analytical transparency. MLR assumes a linear relationship between dependent (biological activity) and independent (descriptor) variables:

$$Y=b0+b1X1+b2X2+...+bnXn+\epsilon Y=b0+b1X1+b2X2+...+bnXn+\epsilon$$

where bi*bi* are regression coefficients and ϵϵ represents random error. MLR coefficients convey both the direction and magnitude of descriptor contributions to activity positive coefficients indicate direct correlation, whereas negative coefficients signify inverse relationships.

A typical QSAR modeling workflow involves

1. Dataset preparation and standardization of activity data (e.g., converting IC50 to pIC50).
2. Descriptor calculation and normalization.
3. Correlation analysis to remove highly collinear variables.
4. Model generation using least-squares fitting.
5. Validation using internal (cross-validation) and external (test set) procedures.

Model adequacy is evaluated using metrics such as coefficient of determination (R^2), cross-validated coefficient (Q^2), standard error of estimate (SEE), and F-statistic. Acceptable QSAR models typically exhibit R2>0.6*R*2>0.6 and Q2>0.5*Q*2>0.5, although these thresholds vary depending on dataset complexity. However, MLR assumes homoscedasticity, linearity, and independence of residuals conditions that may not hold for nonlinear biological phenomena. To address this, researchers often introduce polynomial or interaction terms or employ data transformations (e.g., logarithmic scaling) to linearize relationships. Residual analysis and outlier diagnostics (Cook's distance, leverage plots) are essential for identifying compounds that disproportionately influence regression parameters. Despite its limitations, MLR's strength lies in interpretability it enables mechanistic hypotheses about how specific physicochemical properties influence biological activity. For example, a positive coefficient for logP may suggest enhanced receptor penetration with increasing hydrophobicity, while a negative coefficient for molecular weight may reflect steric hindrance in ligand–target interaction.

4.4 Partial Least Squares (PLS) and Principal Component Analysis (PCA) in QSAR

The progression of QSAR from simple linear correlations to multivariate analysis marked a major methodological advancement in computational drug design. When descriptor intercorrelation becomes significant, Partial Least Squares (PLS) and Principal Component Analysis (PCA) offer robust statistical alternatives to traditional MLR by projecting high-dimensional data into orthogonal latent variables that maximize covariance between descriptors and activity. Principal Component Analysis (PCA) is an unsupervised dimensionality reduction technique used to summarize variance within descriptor matrices. In PCA, new orthogonal variables (principal components, PCs) are constructed as linear combinations of original descriptors such that the first few components capture the majority of variance. In QSAR, PCA facilitates data visualization, cluster identification, and the removal of redundant descriptors. Molecules projected onto a principal component space often reveal structure activity trends, allowing chemists to discern which physicochemical attributes most influence potency. However, PCA alone does not incorporate biological activity during component construction. This limitation is overcome by Partial Least Squares (PLS), a supervised method that finds latent variables

(LVs) maximizing covariance between descriptor (X) and response (Y) matrices. PLS is particularly powerful when descriptors are numerous and collinear conditions common in chemoinformatics datasets. In contrast to MLR, PLS allows more descriptors than compounds while still preventing overfitting through dimensionality reduction. The PLS algorithm iteratively extracts components by projecting X and Y onto new axes that capture maximum covariance, optimizing predictive performance.

Mathematically, PLS can be represented as:

$$X = TPT + E \text{ and } Y = UQT + F \quad X = TPT + E \text{ and } Y = UQT + F$$

where T*T* and U*U* are score matrices, P*P* and Q*Q* are loadings, and E*E*, F*F* represent residuals. The relationship between T*T* and U*U* reflects the linear dependence between descriptor and activity subspaces. PLS regression coefficients are then used to predict biological activity for new compounds. PLS has become a cornerstone of 3D-QSAR methods such as CoMFA (Comparative Molecular Field Analysis) and CoMSIA (Comparative Molecular Similarity Indices Analysis), where thousands of field descriptors (steric, electrostatic) are reduced to a manageable number of latent components. Validation metrics such as cross-validated Q^2, predictive R^2 (R^2_pred), and root mean square error (RMSE) are used to assess performance. A well-calibrated PLS model typically displays $Q^2 > 0.5$ and low RMSE for external test sets.

Both PCA and PLS allow graphical representation through score plots and loading plots, providing visual interpretability of chemical space. Score plots cluster compounds by activity, while loading plots identify descriptors or molecular regions responsible for differences. Consequently, PLS and PCA remain indispensable in classical QSAR, offering mechanistic transparency and statistical rigor essential for regulatory acceptance.

4.5 Validation of QSAR Models: Internal, External, and Applicability Domain

Validation is the defining criterion separating credible QSAR models from spurious correlations. The OECD principles for QSAR validation stipulate five criteria: (1) defined endpoint, (2) unambiguous algorithm, (3) defined applicability domain (AD), (4) appropriate measures of goodness-of-fit, robustness, and predictivity, and (5) mechanistic interpretation, if possible [1]. Each aspect ensures that the model is scientifically sound, transparent, and reproducible.

Internal validation assesses model robustness using subsets of the training data. Common techniques include:

- Leave-One-Out (LOO) Cross-Validation, where each compound is sequentially omitted and predicted using the remaining dataset to compute Q^2(LOO).
- Leave-Many-Out (LMO) or k-Fold Cross-Validation, which improves reliability by averaging predictions over multiple partitions.
- Bootstrapping, where multiple random samples are drawn to evaluate coefficient stability.

External validation evaluates model generalization using an independent test set excluded from training. Predictive performance is quantified by external R^2 (R^2_pred), RMSE_pred, and concordance correlation coefficient (CCC). An ideal QSAR model exhibits both internal consistency (high Q^2) and external predictivity (high R^2_pred) without significant discrepancy, indicating generalizable trends rather than overfitting. The applicability domain (AD) defines the chemical space within which the model can make reliable predictions. Methods for AD estimation include the leverage approach (Williams plot) and distance-based methods such as Euclidean or Mahalanobis distance. Compounds lying beyond the AD are considered extrapolations and require caution in interpretation.

QSARINS and KNIME implement AD visualization tools enabling researchers to identify safe prediction boundaries.

Statistical criteria are complemented by mechanistic interpretability, ensuring that model coefficients align with known pharmacological principles. For example, a positive correlation between lipophilicity and activity should be chemically reasonable for hydrophobic binding sites but not for aqueous transporters. Such mechanistic congruence is crucial for confidence in QSAR-derived hypotheses. In toxicology and pharmacokinetics, validated QSAR models are recognized by regulatory agencies including the European Chemicals Agency (ECHA) and the U.S. Environmental Protection Agency (EPA) for risk assessment under OECD guidelines. Consequently, rigorous validation not only ensures scientific credibility but also facilitates regulatory acceptance of in silico predictions as non-animal alternatives in ADMET evaluation.

4.6 Integration of Classical QSAR with ADMET Prediction

While early QSAR studies focused primarily on pharmacodynamic endpoints such as receptor binding or enzyme inhibition, modern drug discovery requires simultaneous optimization of pharmacokinetic and toxicological properties. The extension of QSAR principles to ADMET prediction represents one of the most transformative shifts in computational pharmacology. In Absorption, QSAR models correlate molecular descriptors with permeability data (e.g., Caco-2 cell permeability, PAMPA assays). Descriptors such as molecular weight, topological polar surface area (tPSA), and logP are crucial determinants of membrane transport. For instance, Lipinski's Rule of Five parameters are empirically grounded in QSAR-derived correlations that delineate orally bioavailable compounds [2].
For Distribution, classical QSAR models predict plasma protein binding, blood–brain barrier (BBB) penetration, and volume of distribution (Vd). Correlations between logP, molecular volume, and polarizability often provide reliable BBB predictions. Notably, Hansch-type models have been used to distinguish CNS-active from peripherally restricted drugs based on lipophilicity thresholds.
In Metabolism, QSAR has been applied to predict cytochrome P450 inhibition and metabolic stability. Electronic descriptors such as frontier orbital energies (EHOMO, ELUMO) and Mulliken charges correlate with metabolic susceptibility to oxidation or hydrolysis. Linear regression and PLS models trained on experimental clearance data enable early identification of metabolic liabilities.

Excretion modeling relies on polar descriptors (e.g., hydrogen bond donors/acceptors) influencing renal filtration and biliary elimination. In Toxicity, classical QSAR remains central to the prediction of mutagenicity (Ames test), carcinogenicity, and hepatotoxicity. Regulatory databases such as VEGA, admetSAR, and OECD QSAR Toolbox employ classical statistical models to estimate toxicological endpoints based on curated descriptors. An illustrative case involves predicting hepatotoxicity using PLS regression trained on hydrophobicity, molecular volume, and aromaticity indices. Compounds exhibiting high hydrophobic surface area and aromatic density often show positive regression coefficients correlating with hepatotoxic risk due to bioactivation and accumulation mechanisms. Thus, classical QSAR provides the theoretical underpinning for modern in silico ADMET filters serving as the first computational checkpoint in early drug discovery pipelines. The interpretability of linear models allows medicinal chemists to rationally modify structures to improve pharmacokinetic profiles while retaining potency.

4.7 Case Studies Illustrating Classical QSAR Applications

4.7.1 β-Adrenergic Antagonists (Hansch Analysis)

One of the earliest demonstrations of QSAR efficacy was the correlation between lipophilicity and β-blocking activity of aryloxypropanolamines. Hansch and Fujita established that biological activity (log(1/C)) exhibited a parabolic relationship with logP, identifying an optimal hydrophobicity window balancing receptor affinity and solubility [3]. The model guided synthesis of analogues with improved selectivity and reduced side effects, exemplifying rational drug optimization.

4.7.2 Benzodiazepine Derivatives (Free–Wilson Approach)

The Free Wilson model successfully quantified substituent contributions to anxiolytic activity in benzodiazepines. Binary indicator variables representing electron-withdrawing or donating substituents at ortho-, meta-, and para-positions allowed additive modeling of potency. The model revealed key structural motifs essential for receptor binding and sedative properties, validating the additive assumption in classical QSAR.

4.7.3 Toxicity Prediction in Aromatic Amines

QSAR models employing Taft steric constants and Hammett σ parameters accurately predicted mutagenic potential in aromatic amines. Regression analysis showed that electron-donating substituents enhanced mutagenicity via increased formation of electrophilic intermediates, aligning with mechanistic biochemical evidence. Such interpretability remains a key advantage of classical QSAR over opaque AI models.

4.7.4 ADMET Modeling of NSAIDs

In an applied pharmacokinetic context, linear regression models correlated molecular size, logP, and hydrogen-bonding capacity with gastrointestinal toxicity among nonsteroidal anti-inflammatory drugs (NSAIDs). Higher lipophilicity and acidic pKa were associated with mucosal irritation, offering insights for designing safer analogues. This study demonstrated the translational relevance of classical QSAR for risk minimization. Collectively, these examples underscore the adaptability of classical QSAR principles across therapeutic classes and mechanistic domains from potency optimization to safety profiling.

4.8 Critical Evaluation: Strengths, Limitations, and Evolving Role

The enduring relevance of classical QSAR arises from its balance between simplicity, interpretability, and computational efficiency. Linear models provide explicit relationships between structure and function, enabling hypothesis-driven medicinal chemistry. They require minimal computational resources and offer transparency necessary for regulatory compliance. Moreover, the statistical rigor of regression analysis provides quantifiable uncertainty estimates and confidence intervals, fostering reproducibility. However, limitations are inherent. The linear assumption may oversimplify complex biological phenomena governed by nonlinear interactions, allosteric effects, and conformational dynamics. Descriptor redundancy and multicollinearity can distort coefficient estimation, while small datasets risk overfitting. Classical QSAR also assumes that biological activity arises primarily from equilibrium interactions, neglecting kinetic or temporal dimensions of pharmacology. Another challenge involves the applicability domain: models trained on limited chemical space often fail to generalize to novel scaffolds. Additionally, classical QSAR lacks the ability to capture non-additive effects and synergistic interactions between molecular features. These

shortcomings have prompted the adoption of nonlinear methods such as support vector machines, random forests, and neural networks discussed in later chapters.

Nevertheless, classical QSAR continues to serve as an interpretive and regulatory benchmark. Its outputs are mechanistically interpretable, aligning with medicinal chemistry intuition. In practice, hybrid strategies combining linear descriptors with nonlinear learners often yield optimal balance between accuracy and explainability. As computational power expands and descriptor libraries grow richer, classical QSAR remains foundational not as an outdated relic but as a transparent scaffold upon which modern AI-based methods are built. Its statistical discipline, validation rigor, and interpretive clarity ensure that even in the age of deep learning, the principles of Hansch and Free Wilson endure at the core of rational drug design.

Table 4.1 Comparison of Classical Statistical Methods Used in QSAR Modelling

Statistical Method	Mathematical Basis	Typical Application in QSAR	Advantages	Limitations
Simple Linear Regression (SLR)	One-variable least-squares fit	Early hydrophobicity–activity (Hansch) correlations	Easy interpretation; minimal data need	Ignores multivariable interplay; poor generalization
Multiple Linear Regression (MLR)	Multivariate least-squares optimization	Structure–activity correlations with physicochemical descriptors	Quantitative coefficients; mechanistic insight	Sensitive to collinearity; assumes linearity
Principal Component Analysis (PCA)	Eigenvector decomposition of variance–covariance matrix	Data reduction and visualization of descriptor space	Identifies hidden trends; mitigates redundancy	Unsupervised; no direct activity linkage
Partial Least Squares (PLS)	Latent-variable projection maximizing X–Y covariance	3D-QSAR (CoMFA/CoMSIA) and collinear datasets	Handles > descriptors than samples; robust to multicollinearity	Requires interpretation of latent components
Stepwise Regression / Genetic Algorithms	Iterative descriptor selection by statistical criteria	Feature subset optimization	Automated; improves model simplicity	Risk of overfitting; dataset-dependent
Principal Component Regression (PCR)	Regression on PCA components	When descriptors are highly correlated	Reduces dimensionality; computationally efficient	Loss of direct descriptor meaning
Ridge / Lasso Regression	Penalized least squares (L2 / L1 norms)	Regularization of QSAR models	Controls overfitting; improves stability	Parameter tuning required; less intuitive coefficients

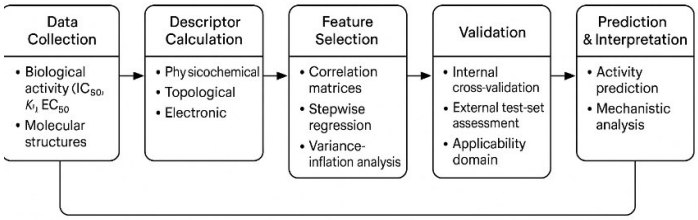

Figure 1 Workflow of Classical QSAR Model Development

4.9 Future Perspectives: Bridging Classical and AI-Driven QSAR

The trajectory of QSAR methodology is increasingly convergent with artificial intelligence, yet classical statistical foundations remain indispensable. Future research is expected to integrate hybrid models where MLR- or PLS-derived interpretable coefficients inform deep-learning architectures, preserving explainability while enhancing predictive performance. Transfer learning and multi-task regression could extend classical QSAR principles to multitarget pharmacology, simultaneously optimizing efficacy and ADMET endpoints. Emerging quantum chemical descriptors, topological indices from graph theory, and molecular interaction fingerprints will expand descriptor diversity, allowing more comprehensive mapping of structure activity landscapes. Coupling classical QSAR with molecular dynamics simulations can incorporate conformational flexibility into regression models, bridging static descriptors with dynamic behavior.

In regulatory science, the movement toward transparent AI aligns with QSAR's long-standing emphasis on interpretability. OECD-compliant hybrid models may soon become the standard for submission-ready computational toxicology. As data availability increases through FAIR-compliant repositories, statistically grounded QSAR frameworks will play a central role in model reproducibility and open-science validation. Ultimately, the integration of statistical QSAR, ADMET modeling, and AI represents not a replacement but an evolution of classical principles extending the quantitative bridge between molecular structure and pharmacological function toward a fully data-centric paradigm of predictive drug design.

4.10 CONCLUSION

Classical QSAR represents one of the most enduring and scientifically grounded paradigms in computational drug design. Originating from the pioneering work of Hansch and Free Wilson, it established the quantitative link between molecular structure and biological activity that remains central to modern pharmacological modeling. Through linear regression, multivariate analysis, and rigorous statistical validation, classical QSAR models provide interpretable relationships that enable rational optimization of potency, selectivity, and pharmacokinetic behavior.

Despite the rapid evolution of artificial intelligence and deep learning, classical QSAR endures as the conceptual and regulatory foundation upon which modern predictive systems are built. Its advantages lie in transparency, simplicity, and mechanistic clarity qualities critical for hypothesis generation and decision-making in medicinal chemistry. The continued relevance of multiple linear regression (MLR), partial least squares (PLS), and principal component analysis (PCA) in ADMET prediction demonstrates that statistically interpretable methods remain indispensable tools for understanding the molecular determinants of efficacy and safety.

REFERENCES

1. OECD. Principles for the Validation, for Regulatory Purposes, of (Quantitative) Structure–Activity Relationship Models. OECD Series on Testing and Assessment No. 69. Paris: OECD; 2004.
2. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev.* 2012;64:4–17.
3. Hansch C, Fujita T. p–σ–π Analysis. A method for the correlation of biological activity and chemical structure. *J Am Chem Soc.* 1964;86(8):1616–1626.
4. Todeschini R, Consonni V. *Handbook of Molecular Descriptors.* Wiley-VCH; 2009.
5. Gramatica P. Principles of QSAR modeling: Validation, internal and external. *QSAR Comb Sci.* 2007;26(5):694–701.
6. Roy K, Kar S, Das RN. *A Primer on QSAR Modeling.* Springer; 2015.
7. Dearden JC. The history and development of quantitative structure–activity relationships (QSARs). *Int J Quant Struct-Prop Relat.* 2019;4(1):1–44.
8. Toropov AA, Benfenati E, Leszczynski J. QSAR modeling of drug-induced liver injury. *Chem Res Toxicol.* 2020;33(6):1489–1500.
9. Eriksson L, Johansson E, Kettaneh-Wold N, Wold S. *Multi- and Megavariate Data Analysis.* MKS Umetrics; 2006.
10. ECHA. *QSAR Toolbox User Manual v4.6.* European Chemicals Agency; 2023.
11. Puzyn T, Leszczynska D, Leszczynski J. Toward the development of “intelligent” QSARs: Advances and challenges. *Curr Comput Aided Drug Des.* 2020;16(2):80–92.
12. Veith GD et al. QSAR models for estimating plasma protein binding. *Chemosphere.* 2018;208:886–892.
13. Honorio KM, da Silva ABF. Quantitative structure–activity relationships for the modeling of drug metabolism. *Curr Drug Metab.* 2017;18(2):126–137.
14. Tropsha A. Best practices for QSAR model development, validation, and exploitation. *Mol Inform.* 2022;41(6):2100131.
15. Gramatica P, Cassani S, Chirico N. QSARINS: A software for developing, validating, and analyzing QSAR models. *Chemom Intell Lab Syst.* 2014;127:123–139.