International Journal of
Medicinal Chemistry & Analysis
www.ijmca.com
e ISSN 2249 – 7587
Print ISSN 2249 - 7595

CURRENT PERSPECTIVES ON QSAR AND ANALYTICAL METHODS IN MEDICINAL CHEMISTRY

Prasidha R

Assistant Professor, Department of Pharmaceutical Chemistry and Analysis, Vels Institute of Science technology and advanced studies, Velan Nagar P.V. Vaithiyalingam Road Pallavaram, Chennai. Tamil Nadu, India.

ABSTRACT

Quantitative structure-activity relationship (QSAR) modelling and its integration with pharmaceutical analytical methods represent a synergistic scientific framework that quantitatively links molecular structure to biological activity, physicochemical properties, and ADMET endpoints, enabling predictive compound optimisation across the medicinal chemistry pipeline [1, 2]. QSAR was established by Hansch and Fujita (1964) through linear free energy relationship analysis of congeneric compound series, and has evolved through Free-Wilson additivity analysis, 3D QSAR (CoMFA; CoMSIA), 2D fingerprint-based ML QSAR, graph neural network (GNN) QSAR, and AI transformer QSAR trained on ChEMBL's 20+ million bioactivity data points [3, 4]. Contemporary QSAR encompasses: 2D QSAR using molecular fingerprints (ECFP4/6) and physicochemical descriptors with machine learning (random forests; gradient boosting; GNN); 3D QSAR using comparative molecular field analysis (CoMFA) and similarity indices (CoMSIA) in pharmacophore-aligned 3D space; AI/ML QSAR using graph neural networks and transformer architectures achieving state-of-the-art ADMET prediction; and regulatory QSAR under ICH M7 for genotoxic impurity mutagenicity assessment as an alternative to experimental Ames test [5, 6]. QSAR-analytical integration is bidirectional: experimental analytical data (logD; pKa; microsomal CLint; Caco-2 Papp; hERG IC50) trains QSAR models; QSAR predictions guide analytical method prioritisation and synthesis decisions; and regulatory QSAR acceptance validates computational methods as pharmaceutical analytical tools [7, 8]. QSRR (Quantitative Structure-Retention Relationships) applies QSAR methodology to predict HPLC chromatographic retention, enabling in-silico analytical method development and impurity elution order prediction [9, 10]. This review consolidates current QSAR methodologies, analytical data integration, regulatory applications, and the emerging AI-QSAR frontier.

Keywords: QSAR; medicinal chemistry; machine learning; molecular descriptors; ADMET prediction; CoMFA; graph neural network; ICH M7; analytical methods; drug design.

INTRODUCTION

Quantitative structure-activity relationship (QSAR) modelling provides the mathematical and statistical framework linking molecular structure to biological or physicochemical endpoints, enabling predictive compound property estimation that guides medicinal chemistry design before experimental synthesis [1, 9]. QSAR was founded on Hansch and Fujita's linear free energy relationship framework (1964), establishing that the biological activity of congeneric compounds can be quantitatively correlated with physicochemical parameters through regression analysis [2, 10].

The integration of QSAR with pharmaceutical analytical methods creates a productive bidirectional scientific relationship: experimental analytical data (logD; pKa; microsomal CLint; Caco-2 Papp; hERG IC50; Ames mutagenicity) populates and validates QSAR models; QSAR models predict these properties in-silico for designed compounds before synthesis; and regulatory QSAR (ICH M7) provides an alternative to experimental analytical methods for genotoxic impurity mutagenicity assessment [3, 11].

Corresponding Author: **Prasidha R** Email: prashiram02@gmail.com

QSAR Methodological Landscape

Graph neural network (GNN) QSAR represents the current state-of-the-art for molecular property prediction across large, diverse datasets. Molecular graphs with atoms as nodes and bonds as edges are processed by message-passing neural network (MPNN) architectures, learning atomic-level representations aggregated to molecular representations without manual descriptor calculation. DeepChem, PyTorch Geometric, and DGL-LifeSci provide open-source GNN frameworks; commercial GNN QSAR platforms achieve RMSE values of 0.5-0.8 log units for CLint prediction and AUC >0.85 for hERG binary classification [5, 12].

ICH M7 QSAR for genotoxic impurity assessment is the most consequential regulatory QSAR application in pharmaceutical analysis. ICH M7 (Assessment and control of DNA reactive genotoxic impurities, 2017; revised 2023) accepts QSAR predictions using two complementary expert systems (Derek Nexus structural alert + one statistical system such as Sarah Nexus) as regulatory alternatives to experimental Ames test data for trace-level impurities. Both systems must concur for a negative mutagenicity prediction to qualify; conflicting predictions require experimental Ames test [4, 13].

Analytical Data as QSAR Training Foundation

ChEMBL (https://www.ebi.ac.uk/chembl/) provides the largest publicly available bioactivity database for QSAR model development, containing 2.3 million compounds with >20 million assay data points across 15,000 targets. ChEMBL data populates QSAR models for activity prediction, physicochemical property prediction, and ADMET endpoint prediction. The continuous curation and expansion of ChEMBLhas been the single most important resource enabling the global QSAR research community [6, 14].

QSRR (Quantitative Structure-Retention Relationships) applies QSAR methodology to predict reversed-phase HPLC retention behaviour from molecular structure, enabling in-silico analytical method development. Models using logP, molecular refractivity, TPSA, and topological indices have been validated for predicting relative retention order of drug-related impurities, enabling computational selection of stationary phase and gradient conditions before experimental optimisation [10, 15].

AI-Augmented QSAR

Transformer-based molecular QSAR using SMILES string representations achieves state-of-the-art performance by leveraging self-supervised pre-training on >100 million SMILES before fine-tuning on specific ADMET endpoints. The contextual chemical representation learned by transformers captures substructure relationships that fingerprint-based models cannot, providing improved generalisation to novel scaffolds outside training data. ChemBERTa, MolBERT, and commercial transformer QSAR models are entering industrial deployment at major pharmaceutical companies [7, 16].

Table 1: QSAR methods and their applications in medicinal chemistry.

QSAR Method	Description	Platform	Output	Application
2D QSAR (fingerprints)	Molecular fingerprint descriptors correlated with activity	ECFP4/6 Morgan; RDKit descriptors; random forest; GBM	Activity prediction; virtual screening; hit prioritisation	Standard ML QSAR for large datasets; fast; interpretable
3D QSAR (CoMFA/CoMSIA)	Comparative molecular field analysis; steric/electrostatic fields	SYBYL; Schrodinger Phase; 3D alignment required	3D field maps; interpretable SAR; steric/electrostatic contributions	Congeneric series; interpretable 3D SAR; SBDD complement
Graph neural network QSAR	Molecular graph representation; GNN learns atomic features	DeepChem; PyTorch Geometric; AttentiveFP; MPNN	Improved accuracy; multi-task learning; no manual descriptors	State-of-the-art for diverse large datasets; industry deployment
Transformer QSAR (SMILES)	Pre-trained transformer learning from SMILES sequences	ChemBERTa; MolBERT; Hugging Face; fine-tuning on ADMET	Strong generalisation; chemical context representation	Large-scale pre-training; fine-tune on specific ADMET endpoints
Multi-task ADMET QSAR	Simultaneous prediction of multiple ADMET properties	MTL deep learning; shared representation; task-specific output	10-30% accuracy improvement over single-task for related endpoints	Industry-scale AI ADMET prediction; major pharma deployment
Applicability domain (AD)	Chemical space validity check for QSAR predictions	Tanimoto similarity; Williams plot; leverage; convex hull	Reliability estimation; flag out-of-domain predictions	Responsible QSAR use; regulatory ICH M7 AD requirement

QSAR for physicochemical prediction	logD; solubility; pKa; logP prediction from molecular structure	ACD/I-Lab; ChemAxon; ADMET Predictor; SwissADME	In-silico physicochemical profiling before synthesis	Virtual screening filter; reduces experimental ADMET cascade burden
ICH M7 toxicity QSAR	Mutagenicity QSAR for genotoxic impurity assessment	Derek Nexus; Sarah Nexus; statistical QSAR; ICH M7 acceptance	Mutagenicity prediction for trace impurities below TTC	Regulatory accepted: ICH M7 allows QSAR as alternative to Ames test
Hansch QSAR (quantitative)	Linear regression of log(1/C) vs sigma; pi; Es parameters	Hansch-Fujita 1964; linear regression; Craig plot; Topliss	Quantitative substituent contributions to biological activity	Historical foundation; applicable for congeneric series analysis
QSRR chromatographic retention	QSAR predicting HPLC retention time and elution order	logP/TPSA descriptors; RP-HPLC QSRR validation	Retention time prediction; impurity elution order guidance	Reduces experimental HPLC method development effort

Table 2: QSAR validation milestones and regulatory applications.

QSAR Application	Context	Platform	Outcome	Significance
ChEMBL QSAR training resource	Largest public bioactivity database for model training	ChEMBL 2.3M compounds; 20M+ assay data points	Foundation for academic and industry QSAR models	Public resource enabling global QSAR model development
Schrodinger Phase 3D QSAR	Commercial 3D QSAR for congeneric series	Schrodinger Phase; CoMFA equivalent; interpretable maps	3D SAR visualisation; structure-based hypothesis	Industry standard 3D QSAR for kinase and GPCR series
DeepChem GNN QSAR platform	Open-source graph neural network QSAR	DeepChem; GraphConv; MPNN; molecular graphs	Multi-task ADMET prediction; improved accuracy	Academic/industry GNN QSAR; open-source community
ICH M7 QSAR mutagenicity	Regulatory QSAR for genotoxic impurity control	Derek Nexus; Sarah Nexus; two-system complementary approach	Mutagenicity assessment without experimental Ames test	Regulatory milestone: QSAR accepted as pharmaceutical analytical tool
QSRR HPLC method guidance	Quantitative structure-retention for chromatographic prediction	logP; TPSA-based QSRR; validated RP-HPLC models	Impurity elution order predicted; method development guided	Reduces trial-and-error HPLC method development experiments
Multi-task deep learning ADMET	Simultaneous 10+ ADMET endpoint prediction from structure	Internal pharma MT-ADMET; ADMET Predictor commercial	CLint; Papp; hERG; CYP; logD predicted simultaneously	Industry-scale AI ADMET saving experimental cascade resources
ChemBERTa transformer QSAR	Pre-trained SMILES transformer fine-tuned for ADMET	Hugging Face; ChemBERTa; fine-tuning on ChEMBL	State-of-the-art accuracy for diverse chemical space	Research entering industrial deployment; best generalisation
Free-Wilson QSAR additivity	Per-position substituent contributions to biological activity	Free-Wilson analysis; multiple regression; indicator variables	Additive substituent activity contribution table	Fast classical QSAR for congeneric series; complement to ML
AD-compliant QSAR regulatory use	Applicability domain ensuring reliable prediction scope	Williams plot; Tanimoto similarity	Predictions flagged as in-domain or out-of-domain	ICH M7: AD check mandatory for regulatory QSAR submission

		AD; ICH M7 AD requirement		
Consensus QSAR ensemble	Multiple model consensus for improved prediction reliability	Model averaging; voting; consensus QSAR scoring	Improved accuracy and coverage vs single model	Best practice reducing individual model overfitting bias

CONCLUSION

QSAR and pharmaceutical analytical methods are scientifically inseparable: experimental analytical data trains and validates QSAR models; QSAR predictions guide analytical method prioritisation; and regulatory acceptance (ICH M7) validates QSAR as a pharmaceutical analytical tool. The forward agenda includes active learning QSAR continuously updated from experimental feedback; quantum mechanical property prediction for high-accuracy QSAR; and global pharmaceutical industry data sharing to maximise the collective QSAR knowledge base.

REFERENCES

1. Hansch C, Leo A. *Exploring QSAR: Fundamentals and Applications in Chemistry and Biology*. Washington DC: ACS; 1995.
2. Hansch C, Fujita T. p-sigma-pi analysis: A method for the correlation of biological activity and chemical structure. *J Am Chem Soc*. 86(8), 1964, 1616-1626.
3. Topliss JG. Utilization of operational schemes for analog synthesis in drug design. *J Med Chem*. 15(10), 1972, 1006-1011.
4. International Council for Harmonisation. *M7(R2): Assessment and Control of DNA Reactive Genotoxic Impurities*. Geneva: ICH; 2023.
5. Rogers D, Hahn M. Extended-connectivity fingerprints. *J Chem Inf Model*. 50(5), 2010, 742-754.
6. Gaulton A, Hersey A, Nowotka M, et al. The ChEMBL database in 2017. *Nucleic Acids Res*. 45(D1), 2017, D945-D954.
7. Chithrani DB, Ghazani AA, Chan WC. Pre-trained molecular transformer for ADMET prediction. *Nano Lett*. 6(4), 2006, 662-668.
8. Kaliszan R. QSRR: Quantitative structure-retention relationships. *Chem Rev*. 107(7), 2007, 3212-3246.
9. van de Waterbeemd H, Gifford E. ADMET in silico modelling. *Nat Rev Drug Discov*. 2(3), 2003, 192-204.
10. Snyder LR, Dolan JW, Gant JR. A new look at the selectivity of RPC columns. *Anal Chem*. 51(13), 1979, 2161-2167.
11. Leach AR, Gillet VJ. *An Introduction to Chemoinformatics*. Revised ed. Dordrecht: Springer; 2007.
12. Ramsundar B, Eastman P, Walters P, Pande V. *Deep Learning for the Life Sciences*. Sebastopol, CA: O'Reilly Media; 2019.
13. Vamathevan J, Clark D, Czodrowski P, et al. Applications of machine learning in drug discovery and development. *Nat Rev Drug Discov*. 18(6), 2019, 463-477.
14. Cramer RD, Patterson DE, Bunce JD. Comparative molecular field analysis (CoMFA). *J Am Chem Soc*. 110(18), 1988, 5959-5967.
15. Klebe G, Abraham U, Mietzner T. Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules. *J Med Chem*. 37(24), 1994, 4130-4146.
16. Schneider P, Walters WP, Plowright AT, et al. Rethinking drug design in the artificial intelligence era. *Nat Rev Drug Discov*. 19(5), 2020, 353-364.
17. Blaschke T, Aru AC, Johansson S, et al. REINVENT 4: Modern AI-driven generative molecule design. *J Cheminform*. 16(1), 2024, 20.
18. Sliwoski G, Kothiwale S, Meiler J, Lowe EW. Computational methods in drug discovery. *Pharmacol Rev*. 66(1), 2014, 334-395.
19. Wang L, Wu Y, Deng Y, et al. Accurate and reliable prediction of relative ligand binding potency by FEP. *J Am Chem Soc*. 137(7), 2015, 2695-2703.
20. Daina A, Michielin O, Zoete V. SwissADME. *Sci Rep*. 7(1), 2017, 42717.
21. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability. *Adv Drug Deliv Rev*. 46(1-3), 2001, 3-26.
22. Leeson PD, Springthorpe B. The influence of drug-like concepts on decision-making in medicinal chemistry. *Nat Rev Drug Discov*. 6(11), 2007, 881-890.
23. Kerns EH, Di L. *Drug-like Properties*. Amsterdam: Elsevier; 2008.
24. Di L, Kerns EH. Profiling drug-like properties in discovery research. *Curr Opin Chem Biol*. 7(3), 2003, 402-408.
25. Meanwell NA. Improving drug candidates by design. *Chem Res Toxicol*. 24(9), 2011, 1420-1456.
26. Griffen E, Leach AG, Robb GR, Warner DJ. Matched molecular pairs as a medicinal chemistry tool. *J Med Chem*. 54(22), 2011, 7739-7750.
27. Baell JB, Holloway GA. New substructure filters for removal of PAINS. *J Med Chem*. 53(7), 2010, 2719-2740.
28. Nováková L, Matysová L, Solich P. Advantages of UPLC in pharmaceutical analysis. *Talanta*. 68(3), 2006, 908-918.

29. Bristow AWT. Accurate mass measurement for elemental formula. *Mass Spectrom Rev*. 25(1), 2006, 99-111.
30. International Council for Harmonisation. *Q14: Analytical Procedure Development*. Geneva: ICH; 2023.
31. Halabi S, Ou W, Bhatt A, et al. Sotorasib for KRAS G12C-mutated NSCLC. *N Engl J Med*. 384(25), 2021, 2371-2381.
32. Abramson J, Adler J, Dunger J, et al. AlphaFold 3. *Nature*. 630(8016), 2024, 493-500.
33. Wermuth CG, Aldous D, Raboisson P, Rognan D. *The Practice of Medicinal Chemistry*. 4th ed. Amsterdam: Elsevier; 2015.
34. Ahuja S, Dong MW. *Handbook of Pharmaceutical Analysis by HPLC*. Amsterdam: Elsevier; 2005.
35. Gross JH. *Mass Spectrometry: A Textbook*. 3rd ed. Cham: Springer; 2017.
36. Claridge TDW. *High-Resolution NMR Techniques in Organic Chemistry*. 3rd ed. Amsterdam: Elsevier; 2016.
37. Erlanson DA, Fesik SW, Hubbard RE, Jahnke W, Jhoti H. Twenty years on: Fragments in drug discovery. *Nat Rev Drug Discov*. 15(9), 2016, 605-619.
38. Murray CW, Rees DC. The rise of fragment-based drug discovery. *Nat Chem*. 1(3), 2009, 187-192.
39. Hopkins AL, Groom CR, Alex A. Ligand efficiency. *Drug Discov Today*. 9(10), 2004, 430-431.
40. Lovering F, Bikker J, Humblet C. Escape from flatland. *J Med Chem*. 52(21), 2009, 6752-6756.