e-ISSN: 2249 – 7781
Print ISSN: 2249 – 779X

International Journal of Pharmaceutical Research & Analysis
www.ijpra.com

Research Article

DRUG DISCOVERY, SAR, AND ANALYTICAL CHARACTERIZATION: A REVIEW

Maheshwari M

Associate Professor, Department of Pharmaceutical Analysis, St.Mary's Institute of Pharmacy, Chinthamani, Puliangudi-627 855, Tamil Nadu, India.

ABSTRACT

The integrated drug discovery pipeline—encompassing target identification, hit finding, lead identification, lead optimisation, and preclinical candidate selection—combines structure-activity relationship (SAR) science with comprehensive analytical characterisation at each stage to advance compounds from biological hits to clinical candidates [1, 2]. The design-make-test-analyse (DMTA) cycle is the operational rhythm of medicinal chemistry: SAR hypotheses from biological testing drive structural design, synthesis produces designed compounds, biological testing generates activity data, and analytical characterisation ensures compound quality throughout [3, 4]. Modern SAR science deploys: matched molecular pair (MMP) analysis extracting generalised SAR intelligence from historical datasets; free energy perturbation (FEP) providing prospective potency predictions with <1 kcal/mol accuracy; AI generative chemistry proposing novel structures constrained by SAR requirements; structure-based SAR from protein-ligand crystallography; and multiparameter optimisation (MPO) frameworks simultaneously tracking potency, selectivity, ADMET, and safety criteria [5, 6, 7]. Analytical characterisation underpins every SAR cycle: 1H NMR and HRMS confirm identity; HPLC-UV confirms purity (>95% threshold for biological testing); chiral SFC ensures enantiopure testing; physicochemical profiling (solubility, pKa, logD) informs drug-likeness; and ADMET cascade (metabolic stability, Caco-2, PPB, hERG, CYP inhibition) provides preclinical property assessment [8, 9]. Quality control of compounds before biological testing is a scientific validity imperative: impure, misidentified, or racemic compounds produce false SAR data that misdirects expensive programmes [10]. This review consolidates the integrated drug discovery, SAR, and analytical characterisation workflow, examining key analytical requirements at each discovery stage, SAR intelligence tools, and emerging AI-accelerated DMTA approaches.

Keywords: Drug Discovery; Sar; Dmta; Analytical Characterisation; Matched Molecular Pairs; Fep; Admet; Compound Quality; Lead Optimisation; Ai Drug Design.

INTRODUCTION

Drug discovery is a multistage scientific enterprise requiring the integrated deployment of chemistry, biology, and analytics to progress molecules from initial biological hypotheses to clinical candidates. The design-make-test-analyse (DMTA) cycle—the iterative rhythm of medicinal chemistry—requires reliable analytical characterisation at every cycle to ensure that SAR data reflects genuine structure-activity relationships rather than analytical artefacts [1, 9]. The most consequential analytical requirement is compound quality before biological testing: 95%+ HPLC purity, confirmed identity by NMR and HRMS, and accurate structural registration in compound management systems [2, 10].

Modern drug discovery SAR operates within a multiparameter optimisation (MPO) framework that tracks 6-10 parameters simultaneously—potency, selectivity, LE, LipE, metabolic stability, permeability, solubility, hERG, and safety flags—rather than the historically potency-centric approach that drove many ADMET-related clinical failures [3, 11]. MPO-guided

Corresponding Author
Maheshwari M

Email: mahegayathry@gmail.com

SAR, supported by comprehensive analytical cascade data and AI-assisted design, defines the contemporary standard for evidence-based medicinal chemistry.

SAR Science and Tools

Matched molecular pair (MMP) analysis extracts generalised SAR intelligence from historical compound datasets. By identifying compound pairs differing by a single defined structural change, MMP analysis generates per-transformation property delta statistics—quantifying the median change in logD, CLint, Caco-2 permeability, solubility, or potency for each structural transformation across thousands of examples from the medicinal chemistry knowledge base [5, 12]. These transformation tables enable evidence-based SAR decisions guided by accumulated programme experience.

Free energy perturbation (FEP) has transitioned from computational research to industrial SAR tool. Schrodinger FEP+ (industry standard) using OPLS4 force field achieves <1 kcal/mol RMSE for within-series relative binding free energy predictions, prospectively validated in multiple pharmaceutical company kinase inhibitor programmes. By pre-screening proposed analogues in-silico before synthesis, FEP reduces the number of compounds requiring physical synthesis by 40-60%, compressing the DMTA cycle [7, 13].

Analytical Characterisation Framework

Compound quality control is the analytical foundation of reliable SAR. Identity confirmation by 1H NMR (connectivity) and HRMS (molecular formula) prevents structural misassignment. Purity confirmation by HPLC-UV (>95%) prevents potency artefacts from active impurities. Chiral confirmation by SFC ensures enantiopure compounds when chiral centres are present. These analytical gates, applied to every compound before biological testing, are the basis of scientifically valid SAR [9, 15].

ADMET cascade analytical strategy profiles drug-like properties in a tiered format. Tier 1 (all active compounds): kinetic solubility by nephelometry; HLM metabolic stability (t1/2); Caco-2 permeability Papp. Tier 2 (advanced leads): thermodynamic solubility; hERG inhibition IC50; CYP inhibition panel (3A4, 2D6, 2C9); plasma protein binding (RED); logD7.4. Tier 3 (preclinical candidates): reactive metabolite screen; Ames genotoxicity; in-vivo rat PK; solid-state characterisation [8, 16].

Table 1: Integrated drug discovery SAR and analytical requirements.

Stage	Description	Methods	Analytical/SAR Output	Quality Gate
Target identification and validation	Biological target selection; genetic and chemical validation	CRISPR KO; siRNA; chemical probes; patient genetics	Validated target connected to disease mechanism	Genetic validation reduces clinical attrition rates
HTS hit identification	Large compound library screening against validated target	Biochemical; cell-based; 384-well; RO3 compound libraries	Primary actives 0.01-1% hit rate; confirmed dose-response	Purity QC for all hits before SAR investment: 95% HPLC
Fragment-based hit identification	Low MW fragments screened by SPR; NMR; X-ray crystallography	Astex 'rule of three'; 150-300 Da; 1-5% efficiency	Fragment hits elaborated into leads with structural guidance	Venetoclax; vemurafenib; ABT-263 from FBDD pipeline
DEL hit identification	Billion-compound DNA-encoded libraries for affinity selection	X-Chem; Vipergen; GSK DEL; NGS decoding	Novel scaffolds for difficult targets; ultra-large chemical space	Multiple clinical candidates from DEL; ABX464 Phase 3
ADMET cascade profiling	Multi-parameter ADMET property determination for all leads	HLM stability; Caco-2; PPB; hERG; CYP panel; solubility	Property profile driving lead optimisation strategy	Tier 1-2-3 cascade; 15-20 properties per compound
Matched molecular pair analysis	SAR intelligence from historical compound pairs	MMPA; ChEMBL database; per-transformation delta statistics	Quantitative SAR transforms: logD; CLint; solubility; potency	SAR meta-analysis enabling evidence-based structural decisions
FEP potency prediction	Alchemical free energy calculation of relative binding affinity	Schrodinger FEP+; OPLS4; <1 kcal/mol RMSE prospective	40-60% synthesis reduction by pre-filtering in-silico	DMTA acceleration: computational SAR cycles before wet-lab

Compound quality control registry	Identity; purity; stock concentration; structure registration	HPLC purity; HRMS; 1H NMR; DMSO solubility; Dotmatics	Reliable SAR from analytically confirmed pure compounds	Prevents false SAR; critical for scientific validity
MPO lead selection criteria	Multi-parameter optimisation scoring for lead selection	Pfizer CNS MPO; LE; LipE; 6-10 parameter composite score	Holistic lead quality: potency+selectivity+ADMET+safety	Replaces single-parameter potency criterion; reduces attrition
AI generative SAR	Deep learning generating compounds satisfying SAR constraints	REINVENT4; multi-objective scoring; SAR-conditioned generation	Higher-quality candidate proposals from AI vs random synthesis	Compresses DMTA by pre-selecting high-quality synthetic targets

Table 2: Drug discovery SAR and analytical characterisation case studies.

Example	Stage	SAR/Analytical Approach	Outcome	Significance
Venetoclax FBDD to IND in 6 years	FBDD hit to approved drug; BCL-2 BH3 mimetic SAR	NMR FBDD; X-ray structure-based; SFC chiral; ADMET cascade	FDA 2016; CLL transformation; AML approval	Benchmark for fragment-to-drug timeline with rigorous SAR/analytics
Sotorasib KRAS G12C covalent SAR	Covalent warhead SAR for undruggable KRAS target	LCMS covalent adduct assay; selectivity panel; GDP pocket SAR	FDA 2021; first approved KRAS inhibitor	Systematic SAR enabling previously impossible target; covalent analytics
INS018_055 AI TNIK inhibitor Phase 2	End-to-end AI drug discovery; generative chemistry + SAR	Insilico Medicine AI platform; target ID + generation + ADMET	Positive Phase 2 IPF data 2024	First clinical validation of fully AI-designed and AI-optimised drug
Ibrutinib BTK covalent SAR	Covalent Cys481 BTK inhibitor; kinase selectivity SAR	Covalent LCMS; selectivity panel; ADMET cascade; SFC chiral	FDA 2013; >3 million CLL patients treated	Clinical covalent drug SAR milestone
Compound quality failures in drug discovery	Pan-assay interference compounds (PAINS) and impure hits	PAINS filter; HPLC purity; HRMS; NMR structural confirmation	Elimination of false hits before expensive SAR investment	Analytical QC prevents misdirected SAR from impure compounds
MMP metabolic stability SAR	Fluorine substitution median CLint improvement from MMP	ChEMBL MMP analysis; CLint validation assay	F-substituent: median CLint improvement 2x; statistically significant	MMP quantitative SAR for metabolic stability engineering
LipE-driven kinase lead optimisation	Lipophilic efficiency tracking in kinase series optimisation	LipE = pIC50 - logD; trend analysis; LipE >5 target	LipE-optimised leads show improved selectivity and clinical outcomes	Pfizer/AZ LipE adoption as lead quality metric
Chiral SAR: enantioselective activity	Racemic testing masks enantioselective binding SAR	Chiral SFC; enantioselective assay; eutomer/distomer ratio	Correct potency assignment; enantioselective SAR revealed	Critical: racemate testing can misassign SAR for chiral drugs
Fragment growing by X-ray SAR	Fragment elaboration guided by protein crystal structure at each cycle	X-ray fragment soaking; SPR Kd; ADMET at each step	Lead candidate with <100 nM Kd from fragment (original Kd ~1	FBDD iterative SAR with crystallographic guidance

			mM)	
FEP prospective SAR (JACS 2015)	Prospective validation of FEP for relative potency prediction	Wang et al. JACS 2015; <1 kcal/mol RMSE; kinase series	40-60% synthesis reduction confirmed prospectively	Established FEP as practical industrial SAR tool

AI-Accelerated Drug Discovery

AI is transforming the DMTA cycle through: generative chemistry (REINVENT4; diffusion models) proposing novel structures satisfying SAR constraints; ML ADMET prediction filtering proposed structures before synthesis; FEP computational SAR reducing wet-lab synthesis cycles; and automated data analysis integrating biological and analytical results into real-time SAR visualisation tools. INS018_055's Phase 2 clinical success (2024) demonstrates that end-to-end AI drug discovery can produce clinically active candidates [6, 17].

CONCLUSION

Integrated drug discovery combining SAR science, rigorous analytical characterisation, and AI-assisted design defines the modern standard for pharmaceutical research productivity. Quality control at every DMTA cycle, systematic SAR intelligence extraction (MMP), computational SAR (FEP), and AI generative design collectively improve candidate quality and compress discovery timelines. The forward agenda includes fully automated synthesis-to-analytics-to-SAR pipelines, AI learning SAR from continuous programme feedback, and global SAR knowledge sharing accelerating the entire field through public databases.

REFERENCES

1. Wermuth CG, Aldous D, Raboisson P, Rognan D, eds. *The Practice of Medicinal Chemistry.* 4th ed. Amsterdam: Elsevier, 2015.
2. Hughes JP, Rees S, Kalindjian SB, Philpott KL. Principles of early drug discovery. *British Journal of Pharmacology.* 162(6), 2011, 1239-1249.
3. Plowright AT, Johnstone C, Kihlberg J, et al. Hypothesis driven drug design. *Drug Discovery Today.* 17(1-2), 2012, 56-62.
4. Schneider G. Automating drug discovery. *Nature Reviews Drug Discovery.* 17(2), 2018, 97-113.
5. Dossetter AG, Griffen EJ, Leach AG. Matched molecular pair analysis in drug discovery. *Drug Discovery Today.* 18(15-16), 2013, 724-731.
6. Blaschke T, Aru AC, Johansson S, et al. REINVENT 4: modern AI-driven generative molecule design. *Journal of Cheminformatics.* 16(1), 2024, 20.
7. Wang L, Wu Y, Deng Y, et al. Accurate and reliable prediction of relative ligand binding potency. *Journal of the American Chemical Society.* 137(7), 2015, 2695-2703.
8. Kerns EH, Di L. *Drug-like Properties: Concepts, Structure, Design and Methods.* Amsterdam: Elsevier, 2008.
9. Ahuja S, Dong MW, eds. *Handbook of Pharmaceutical Analysis by HPLC.* Amsterdam: Elsevier, 2005.
10. Leeson PD, Springthorpe B. The influence of drug-like concepts on decision-making in medicinal chemistry. *Nature Reviews Drug Discovery.* 6(11), 2007, 881-890.
11. Kola I, Landis J. Can the pharmaceutical industry reduce attrition rates? *Nature Reviews Drug Discovery.* 3(8), 2004, 711-715.
12. Griffen E, Leach AG, Robb GR, Warner DJ. Matched molecular pairs as a medicinal chemistry tool. *Journal of Medicinal Chemistry.* 54(22), 2011, 7739-7750.
13. Chodera JD, Mobley DL, Shirts MR, Dixon RW, Branson K, Pande VS. Alchemical free energy methods for drug discovery. *Current Opinion in Structural Biology.* 21(2), 2011, 150-160.
14. Baell JB, Holloway GA. New substructure filters for removal of pan assay interference compounds. *Journal of Medicinal Chemistry.* 53(7), 2010, 2719-2740.
15. Di L, Kerns EH. Profiling drug-like properties in discovery research. *Current Opinion in Chemical Biology.* 7(3), 2003, 402-408.
16. Hopkins AL, Groom CR, Alex A. Ligand efficiency: a useful metric for lead selection. *Drug Discovery Today.* 9(10), 2004, 430-431.
17. Ren F, Ding X, Zheng M, et al. AlphaFold accelerates artificial intelligence powered drug discovery. *Chemical Science.* 14(6), 2023, 1443-1452.
18. Halabi S, Ou W, Bhatt A, et al. Sotorasib for KRAS G12C-mutated non-small-cell lung cancer. *New England Journal of Medicine.* 384(25), 2021, 2371-2381.
19. Neri D, Lerner RA. DNA-encoded chemical libraries. *Annual Review of Biochemistry.* 87, 2018, 479-502.
20. Murray CW, Rees DC. The rise of fragment-based drug discovery. *Nature Chemistry.* 1(3), 2009, 187-192.

21. Erlanson DA, Fesik SW, Hubbard RE, Jahnke W, Jhoti H. Twenty years on: the impact of fragments on drug discovery. *Nature Reviews Drug Discovery.* 15(9), 2016, 605-619.
22. Canon J, Rex K, Saiki AY, et al. The clinical KRAS G12C inhibitor AMG 510 drives anti-tumour immunity. *Nature.* 575(7781), 2019, 217-223.
23. Barf T, Covey T, Izumi R, et al. Acalabrutinib: a selective covalent Bruton tyrosine kinase inhibitor. *Journal of Pharmacology and Experimental Therapeutics.* 363(2), 2017, 240-252.
24. DiMasi JA, Grabowski HG, Hansen RW. Innovation in the pharmaceutical industry. *Journal of Health Economics.* 47, 2016, 20-33.
25. Paul SM, Mytelka DS, Dunwiddie CT, et al. How to improve R&D productivity. *Nature Reviews Drug Discovery.* 9(3), 2010, 203-214.
26. Swinney DC, Anthony J. How were new medicines discovered? *Nature Reviews Drug Discovery.* 10(7), 2011, 507-519.
27. Congreve M, Murray CW, Blundell TL. Structural biology and drug discovery. *Drug Discovery Today.* 10(13), 2005, 895-907.
28. Silverman RB, Holladay MW. *The Organic Chemistry of Drug Design and Drug Action.* 3rd ed. Amsterdam: Elsevier, 2014.
29. Hansch C, Leo A. *Exploring QSAR.* Washington, DC: ACS, 1995.
30. Meanwell NA. Improving drug candidates by design. *Chemical Research in Toxicology.* 24(9), 2011, 1420-1456.
31. Lipinski CA. Drug-like properties and the causes of poor solubility and poor permeability. *Journal of Pharmacological and Toxicological Methods.* 44(1), 2000, 235-249.
32. Leeson PD. Drug discovery: chemical beauty contest. *Nature.* 481(7382), 2012, 455-456.
33. van de Waterbeemd H, Gifford E. ADMET in silico modelling. *Nature Reviews Drug Discovery.* 2(3), 2003, 192-204.
34. Kennedy T. Managing the drug discovery/development interface. *Drug Discovery Today.* 2(10), 1997, 436-444.
35. Veber DF, Johnson SR, Cheng HY, Smith BR, Ward KW, Kopple KD. Molecular properties that influence oral bioavailability. *Journal of Medicinal Chemistry.* 45(12), 2002, 2615-2623.
36. Nováková L, Matysová L, Solich P. Advantages of application of UPLC in pharmaceutical analysis. *Talanta.* 68(3), 2006, 908-918.
37. Obach RS. Prediction of human clearance of twenty-nine drugs from hepatic microsomal intrinsic clearance data. *Drug Metabolism and Disposition.* 27(11), 1999, 1350-1359.
38. Fermini B, Fossa AA. The impact of drug-induced QT interval prolongation on drug discovery and development. *Nature Reviews Drug Discovery.* 2(6), 2003, 439-447.
39. Shultz MD. The thermodynamic basis for the use of lipophilic efficiency as a tool in drug design. *Bioorganic and Medicinal Chemistry Letters.* 23(21), 2013, 5992-6000.
40. Gleeson MP, Hersey A, Montanari D, Overington J. Probing the links between in vitro potency, ADMET and physicochemical parameters. *Nature Reviews Drug Discovery.* 10(3), 2011, 197-208.

Cite this article:
Maheshwari M. Drug Discovery, SAR, And Analytical Characterization: A Review. *International Journal of Pharmaceutical Research & Analysis,* 16(*Sp.* 2), 2026, 92-96.