Genome Publications
https://doi.org/10.61096/978-81-990998-7-6_12

Chapter 12

Hit Identification Strategies: High-Throughput, Fragment-Based and Virtual Screening

Dr. Leslie V

Professor and Head, Department of Pharmacognosy, St.Johns College of Pharmaceutical Sciences & Research, Kattappana, Idukki, Kerala, India.

Dr. S.P.R. Poonkodi

Associate Professor / HOD, Department of Chemistry, Government Arts College for Women, Sivagangai, Tamil Nadu, India.

Dr. Kannan Raman

Professor and Head, Department of Pharmacology, St. John's College of Pharmaceutical Sciences & Research, Kattapana, Idukki, Kerala, India.

Abstract: Hit identification represents the pivotal early phase of drug discovery, transforming theoretical chemical space into tangible bioactive entities. Over the past two decades, three complementary paradigms high-throughput screening (HTS), fragment-based drug discovery (FBDD), and virtual screening (VS) have redefined how leads are uncovered, validated, and optimized. HTS allows rapid assessment of millions of compounds against defined biological targets using miniaturized, automated assays, whereas FBDD employs small chemical fragments to explore binding hotspots with enhanced chemical efficiency. In parallel, virtual screening utilizes computational algorithms to interrogate large chemical libraries in silico, predicting interactions prior to synthesis. The convergence of these approaches, supported by artificial intelligence, robotic automation, and on-demand libraries, has markedly accelerated hit identification while reducing attrition and cost. This chapter integrates theoretical principles, workflow methodologies, comparative metrics, and real-world applications of HTS, FBDD, and VS. Emphasis is placed on assay design, library curation, computational pre-filtering, fragment evolution strategies, and hybrid in vitro–in silico frameworks. Current challenges false positives, data reproducibility, and fragment elaboration are critically analyzed alongside emerging innovations such as AI-assisted screening, DNA-encoded libraries, and multi-objective optimization. Collectively, these advances herald a new era of efficient, intelligent hit discovery bridging computation and experimentation in modern computer-aided drug design.

Keywords: high-throughput screening, fragment-based drug discovery, virtual screening, hit identification, artificial intelligence.

Citation: Leslie V, S.P.R. Poonkodi, Kannan Raman. Hit Identification Strategies: High-Throughput, Fragment-Based and Virtual Screening. *Comprehensive Approaches in Computer-Aided Drug Design: QSAR, Docking, Screening, Homology, Pharmacophore and AI-Driven Insights.* Genome Publication. 2025; Pp145-154. https://doi.org/10.61096/978-81-990998-7-6_12

12.0 INTRODUCTION

Hit identification marks the transition from theoretical chemical design to empirical biological validation, forming the foundation of rational drug discovery pipelines. Historically, lead molecules were uncovered through serendipitous findings or phenotype-driven screening; however, modern discovery leverages systematic strategies to interrogate chemical diversity with unprecedented precision. The overarching objective is to discover molecular entities exhibiting measurable interaction with a biological target typically an enzyme, receptor, or protein–protein interface while ensuring novelty, potency, and tractable optimization potential [1]. In contemporary workflows, three principal methodologies dominate hit identification: high-throughput screening (HTS), fragment-based drug discovery (FBDD), and virtual screening (VS). HTS represents an *empirical* approach, relying on robotic automation and miniaturized assays to evaluate millions of compounds in vitro. In contrast, FBDD focuses on screening *small fragments* (~150–300 Da) that weakly but efficiently bind to active sites, allowing subsequent elaboration through structure-guided optimization [2]. Virtual screening, on the other hand, provides a *computational complement*, exploring chemical libraries through structure- or ligand-based modeling to identify promising candidates for synthesis or purchase.

The paradigm shift from "brute-force" screening to intelligent, integrated discovery has been driven by technological progress. Automation, microfluidics, artificial intelligence (AI), and high-content imaging have enhanced assay reliability and throughput, while advances in computational chemistry, docking algorithms, and cloud computing have democratized virtual screening for both academia and industry [3]. The intersection of these technologies has yielded hybrid workflows for instance, AI-augmented virtual screening followed by focused HTS validation that drastically reduce experimental burden. Hit identification remains a multi-criteria optimization problem encompassing not only potency but also *drug-likeness*, solubility, stability, and synthetic accessibility. Modern strategies prioritize chemical diversity and biological relevance, often guided by chemoinformatics-derived fingerprints and property filters such as Lipinski's Rule of Five. Fragment-based and computational approaches, in particular, exploit three-dimensional binding information to improve ligand efficiency metrics, such as ligand efficiency (LE), lipophilic ligand efficiency (LLE), and binding efficiency index (BEI) [4].

This chapter explores each hit identification methodology in depth, articulating their workflows, technological enablers, comparative advantages, and integration points. Through detailed examples including kinase inhibitors discovered via HTS, fragment-derived protease inhibitors, and VS-driven antivirals the discussion underscores how these methods form complementary pillars of computer-aided drug design.

12.1 High-Throughput Screening: Principles and Technological Framework

High-throughput screening (HTS) has been the cornerstone of industrial drug discovery since the 1990s, epitomizing the empirical philosophy of testing "many to find a few." It involves the systematic, automated evaluation of large compound libraries against biological targets using miniaturized assays conducted in microtiter plate formats 96-, 384-, 1536-, or even 3456-well configurations. The primary goal is to rapidly detect active hits that produce measurable modulation of the target under controlled conditions [5]. The technological underpinnings of HTS encompass several domains: automation and robotics, assay biochemistry, detection technologies, and data analytics. Modern HTS facilities operate fully automated robotic platforms integrated with liquid handlers, plate readers, incubators, and scheduling software. This enables the screening of over 10^6 compounds per day with remarkable reproducibility. Parallel advances in detection modalities

fluorescence resonance energy transfer (FRET), luminescence, fluorescence polarization (FP), AlphaScreen, and high-content imaging have enabled both biochemical and cell-based assays [6].

Assay miniaturization is critical to reducing reagent consumption and cost, while maintaining assay robustness as measured by statistical parameters such as *Z'-factor*, signal-to-noise ratio, and coefficient of variation. Typical HTS campaigns commence with a primary screen, identifying raw hits based on threshold activity, followed by secondary confirmatory assays, *counter-screens* to exclude artifacts, and orthogonal validation (e.g., surface plasmon resonance, thermal shift assays, or isothermal titration calorimetry) [7]. Recent innovations have transformed HTS into a more data-driven and adaptive process. High-content screening (HCS) combines automated microscopy with image analysis to capture phenotypic information beyond binary readouts, enabling multiparametric hit selection. DNA-encoded libraries (DELs) have expanded chemical diversity by encoding compounds with DNA tags, allowing billions of molecules to be screened simultaneously in a pooled format. Additionally, AI-driven hit triage now leverages machine learning models trained on historical assay data to prioritize hits with optimal physicochemical and ADMET profiles [8].

Despite its success, HTS faces inherent limitations: false positives from assay interference compounds (PAINS), high operational cost, and limited structural diversity of screening collections. Moreover, many identified hits are of low potency or suffer from poor developability. To address these issues, integration with computational pre-filtering for instance, virtual screening or fragment prioritization has become standard practice. Such hybrid approaches reduce the screening set size while enriching for bioactive scaffolds. The success of HTS in identifying first-in-class drugs such as imatinib, sorafenib, and rilpivirine underscores its enduring relevance, but the method continues to evolve toward smart HTS a convergence of automation, machine learning, and cheminformatics optimization that minimizes waste while maximizing biological insight [9].

12.2 Fragment-Based Drug Discovery (FBDD): Concepts and Evolution

Fragment-based drug discovery (FBDD) emerged in the late 1990s as a paradigm shift from large-compound screening to small, information-rich fragment interrogation. Fragments typically <300 Da and possessing 10–20 heavy atoms bind weakly (in the millimolar range) yet efficiently to protein active sites, offering a high *ligand efficiency (LE)* and diverse chemical scaffolds for optimization [10]. The conceptual foundation of FBDD rests on the observation that chemical space grows exponentially with molecular size; thus, screening smaller fragments enables more efficient exploration of available structural diversity. While HTS samples millions of complex molecules, FBDD can identify starting points using libraries of merely a few thousand fragments, emphasizing quality over quantity.

The FBDD workflow comprises four key stages:

1. Fragment library design and selection, focusing on chemical diversity, solubility, and absence of reactive or aggregating moieties.
2. Fragment screening, typically using biophysical techniques sensitive to weak interactions such as nuclear magnetic resonance (NMR), X-ray crystallography, surface plasmon resonance (SPR), or differential scanning fluorimetry (DSF).
3. Hit validation and structural elucidation, often via co-crystallography to reveal precise binding modes.
4. Fragment elaboration and optimization, through *growing*, *linking*, or *merging* strategies guided by structural insights [11].

FBDD's power lies in its structural efficiency each atom in a fragment contributes meaningfully to binding, facilitating rational optimization. Classic examples include the discovery of vemurafenib, a BRAF inhibitor for melanoma, and venetoclax, a BCL-2 inhibitor, both originating from fragment hits evolved into nanomolar ligands [12]. Modern FBDD integrates computational chemistry at multiple stages. Docking and molecular dynamics simulations assist in predicting fragment poses and ranking binding affinities, while AI-assisted fragment linking algorithms, such as those implemented in DeepFrag or REINVENT, accelerate scaffold evolution. Machine learning models trained on fragment–protein interaction databases now predict fragment hotspots and synthetic routes, streamlining optimization cycles [13].

Challenges in FBDD remain nontrivial. Detecting weak binding events demands sensitive instrumentation and meticulous data interpretation, and fragment elaboration requires balancing potency gains against increased molecular weight and lipophilicity. Nevertheless, fragment efficiency and structure-based rationality make FBDD indispensable, especially for *difficult targets* such as protein–protein interactions and allosteric modulators where traditional HTS falters [14]. FBDD's trajectory continues to rise, particularly with cryo-EM integration, enabling structural visualization of fragments in large complexes previously intractable by crystallography. The future promises AI-driven fragment evolution pipelines capable of generating synthetically feasible analogues in silico, bridging FBDD and generative molecular design.

12.3 Virtual Screening: Computational Discovery in Chemical Space

Virtual screening (VS) represents the computational analog of HTS, designed to identify bioactive compounds from large virtual libraries prior to synthesis or purchase. It emerged as a key strategy to address the limitations of empirical screening namely cost, time, and resource intensity by leveraging in silico models that predict ligand binding or biological similarity [15]. The fundamental principle of VS is the *systematic evaluation of molecular structures* against biological targets or known ligands to estimate their binding likelihood. VS can be broadly categorized into structure-based virtual screening (SBVS) and ligand-based virtual screening (LBVS).

Structure-Based Virtual Screening (SBVS) uses three-dimensional structures of target proteins derived from X-ray crystallography, NMR, or homology modeling to predict how candidate molecules fit into the binding site. Docking algorithms such as AutoDock Vina, Glide, GOLD, and DOCK compute potential binding poses and evaluate them with scoring functions reflecting predicted interaction energies. Scoring functions combine van der Waals, electrostatic, and solvation terms, but increasingly incorporate machine learning-based rescoring to improve accuracy [16]. Ligand-Based Virtual Screening (LBVS) is applied when the target structure is unavailable, relying on the principle that "similar molecules exhibit similar activities." It compares molecular descriptors, pharmacophore features, or 3D alignments of known actives to identify structurally or functionally related analogs. Methods such as Tanimoto similarity, 3D shape overlays (ROCS), and quantitative structure–activity relationship (QSAR) models are central to LBVS workflows [17].

A modern VS pipeline involves several key stages: library preparation, target or ligand model generation, screening, post-processing, and hit selection. Pre-processing ensures 3D protonation, tautomerization, and energy minimization using tools such as RDKit, Open Babel, or Schrodinger's LigPrep. For structure-based workflows, docking is typically followed by rescoring using more rigorous methods MM/GBSA or free energy perturbation (FEP) to refine hit ranking [18]. Recent advances have redefined VS with AI integration. Deep learning models such as graph neural networks (GNNs), convolutional neural networks (CNNs), and transformers now predict binding affinities directly from

molecular graphs or 3D conformations, bypassing traditional docking altogether. DeepDock, AtomNet, and GNINA exemplify these architectures, demonstrating improved enrichment factors and reduced false positives [19].

Another major innovation is ultra-large-scale virtual screening (ULVS). Cloud computing platforms and distributed algorithms have made it possible to dock billions of compounds in silico, as exemplified by the COVID Moonshot and Enamine REAL campaigns. These initiatives screened over 10^9 molecules virtually, identifying nanomolar inhibitors of SARS-CoV-2 main protease within weeks an achievement that would have been impossible through wet-lab HTS alone [20]. Nevertheless, VS is not without challenges. Scoring inaccuracies, limited sampling of conformational flexibility, and dependence on target structure quality remain persistent bottlenecks. Moreover, false positives due to overfitting in ML models or poor generalization to novel scaffolds can mislead prioritization. Integrating ensemble docking, molecular dynamics (MD), and consensus scoring has proven effective in mitigating these issues by capturing receptor flexibility and refining energy landscapes [21].

In summary, VS serves as an indispensable filter that precedes or complements experimental screening, narrowing millions of potential candidates to a manageable subset for biochemical validation. It exemplifies the power of computational triage in transforming the economics and efficiency of early-stage drug discovery.

12.4 Integration of High-Throughput, Fragment-Based, and Virtual Screening

The most successful hit discovery campaigns today no longer rely on a single methodology; rather, they employ integrated, iterative pipelines that combine empirical and computational strengths. High-throughput screening provides biological realism and statistical robustness, while fragment-based and virtual approaches offer structural insight and predictive prioritization. One of the most effective integration strategies is pre-filtering HTS libraries using virtual screening. By docking or pharmacophore-filtering millions of compounds before wet-lab testing, researchers can reduce experimental burden while maintaining high hit rates. For instance, the integration of SBVS into Pfizer's kinase discovery pipeline reduced HTS library size by 80% while preserving over 90% of known active scaffolds [22].

Conversely, fragment-based programs increasingly incorporate computational components. Docking fragments into target binding sites prior to NMR or X-ray validation accelerates identification of "hotspot" regions and informs fragment merging strategies. Tools like Schrodinger's Glide Fragments, CCDC's GOLD-FBDD, and DeepFrag automate fragment linking guided by binding site topology and synthetic feasibility [23]. An emerging paradigm is hybrid screening, where virtual and experimental processes occur in parallel and iteratively refine each other. In such cycles, computationally predicted hits undergo biochemical screening; active compounds are then used to retrain machine learning models or refine docking protocols creating a *closed-loop discovery system*. This iterative framework, exemplified by AstraZeneca's "Design-Make-Test-Analyze" (DMTA) cycle, has significantly shortened discovery timelines [24].

The integration of fragment and high-throughput screening (so-called *HT-FBDD*) also addresses complementary weaknesses: while fragments provide high ligand efficiency but low potency, HTS hits offer potency but often lack efficiency or novelty. Combining the two approaches through fragment elaboration of HTS hits, or HTS validation of fragment series, provides balanced chemical starting points. Finally, AI and big data analytics are revolutionizing integration strategies. Machine learning models trained on multi-modal datasets combining docking scores, HTS assay data, and fragment binding information now guide multi-parameter optimization. Reinforcement learning

systems can dynamically select which subset of compounds to synthesize or test next, optimizing exploration versus exploitation in chemical space [25]. Thus, the integration of HTS, FBDD, and VS forms a synergistic ecosystem where computation guides experimentation, and experimental feedback enhances computational precision a hallmark of next-generation computer-aided drug design.

12.5 On-Demand Libraries and DNA-Encoded Screening Platforms

The explosion of accessible chemical space has outpaced the capacity of physical libraries, leading to the rise of on-demand libraries and DNA-encoded libraries (DELs). These innovations bridge virtual and physical screening paradigms, enabling the evaluation of billions of compounds through combinatorial encoding and synthesis-on-demand principles. On-demand virtual libraries such as Enamine REAL, ZINC20, and ChemSpace contain synthetically accessible compounds generated through *virtual combinatorial chemistry*. Using reaction templates and building block inventories, these libraries can virtually enumerate 10^9–10^{12} compounds. Unlike random enumeration, AI-guided synthetic route prediction ensures realistic and tractable molecules. Researchers can order selected hits for synthesis within weeks, enabling seamless transition from in silico hit prediction to laboratory validation [26].

DNA-encoded libraries (DELs) represent a more experimental counterpart. In DEL technology, each compound is covalently linked to a unique DNA barcode that encodes its synthetic history. Billions of DNA-tagged compounds can be pooled and screened simultaneously against a target; active binders are isolated, and their DNA tags are sequenced to identify the corresponding structures [27]. DELs have achieved impressive success in identifying novel scaffolds for kinases, GPCRs, and proteases. Their integration with AI has further enhanced performance machine learning models now predict enrichment patterns in sequencing data, distinguishing true binders from noise. Companies such as X-Chem and HitGen have pioneered hybrid AI-DEL platforms, which merge deep learning models with combinatorial encoding to explore vast uncharted regions of chemical space [28]. These technologies exemplify the convergence of computation, automation, and molecular biology in hit discovery. By coupling synthetic feasibility, data-driven prioritization, and high-throughput experimentation, on-demand and encoded libraries represent the frontier of scalable, sustainable, and cost-effective hit identification.

12.6 Case Studies: Hybrid Hit Discovery Successes

Several landmark discoveries illustrate the transformative potential of integrating HTS, FBDD, and VS methodologies. Case Study 1: BCL-2 Inhibitors (Venetoclax) – The anti-apoptotic protein BCL-2 was long considered “undruggable” due to its shallow binding groove. Fragment screening combined with NMR and X-ray crystallography identified weak binders, which were subsequently optimized through structure-guided design and computational docking. The resulting compound, venetoclax, achieved nanomolar potency and clinical success against chronic lymphocytic leukemia [29].

Case Study 2: JAK2 Inhibitors – Pfizer integrated virtual screening and HTS to identify Janus kinase 2 (JAK2) inhibitors. Docking pre-filtered over 2 million compounds, narrowing the HTS set to 50,000. Hits from HTS were further refined using pharmacophore modeling, leading to the discovery of ruxolitinib, a first-in-class myelofibrosis therapy [30]. Case Study 3: SARS-CoV-2 Main Protease Inhibitors – During the COVID-19 pandemic, an unprecedented collaborative initiative combined cloud-based virtual screening, fragment soaking (at Diamond Light Source), and HTS validation. Within months, researchers identified novel noncovalent inhibitors targeting Mpro, culminating in clinical

leads such as nirmatrelvir (Paxlovid). The effort exemplified a globalized, hybrid hit discovery pipeline integrating computation, crystallography, and biophysics [31].

Case Study 4: Epigenetic Enzyme Inhibitors – Epizyme utilized fragment-based screening complemented by in silico docking to discover inhibitors of histone methyltransferases. Early fragment hits were elaborated using structure-guided merging, validated through biochemical assays, and optimized to yield clinical candidates like tazemetostat [32]. Case Study 5: AI-Assisted Kinase Discovery – Atomwise's deep learning model AtomNet successfully predicted novel kinase inhibitors by screening over 16 billion virtual compounds. Selected top hits underwent HTS validation, yielding multiple nanomolar inhibitors with novel scaffolds. This marked a paradigm shift toward AI-first discovery integrated with confirmatory wet-lab pipelines [33].

These examples demonstrate that integration, not isolation, defines modern success in hit discovery. The fusion of computational prediction, fragment rationality, and experimental throughput continues to yield unprecedented innovation across therapeutic areas.

Table 12.1 Comparative Overview of Hit Identification Strategies

Parameter	**High-Throughput Screening (HTS)**	**Fragment-Based Drug Discovery (FBDD)**	**Virtual Screening (VS)**
Principle	Empirical testing of large compound libraries using automated biochemical or cell-based assays	Detection of weakly binding small fragments (~150–300 Da) and optimization by structure-guided elaboration	Computational prediction of ligand–target binding using docking, similarity, or machine-learning models
Typical Library Size	10^5–10^6 compounds	10^3–10^4 fragments	10^6–10^9 virtual molecules
Detection/Analysis	Optical, luminescent, or imaging-based readouts; Z′-factor validation	Biophysical methods (NMR, X-ray, SPR, DSF)	Scoring functions, free-energy calculations, deep-learning affinity prediction
Output	Potent but sometimes non-specific hits	High-efficiency, structurally tractable fragments	Computationally prioritized hits pending synthesis
Advantages	High statistical reliability; compatible with phenotypic assays	Efficient sampling of chemical space; strong structure–activity rationale	Low cost; high throughput; pre-synthesis triage
Limitations	Costly, high false-positive rates; limited novelty	Weak binding; instrumentation demands	Scoring inaccuracies; reliance on model quality
Representative Tools/Platforms	Robotic screening systems, HCS, DELs	XChem, Schrodinger FEP+, NMR-FBDD suites	AutoDock Vina, Glide, GOLD, GNINA, AtomNet

Key Success Examples	Imatinib, Sorafenib, Ruxolitinib	Venetoclax, Vemurafenib, Tazemetostat	Nirmatrelvir (Paxlovid), AI-predicted kinase inhibitors
Best Use Scenario	Large-scale target validation, phenotypic profiling	Structure-enabled optimization of difficult targets	Early triage, target-based or ligand-based prioritization
Integration Potential	Serves as validation layer for VS/FBDD predictions	Fragment elaboration of HTS/VS hits	Computational pre-filter for HTS or FBDD pipelines

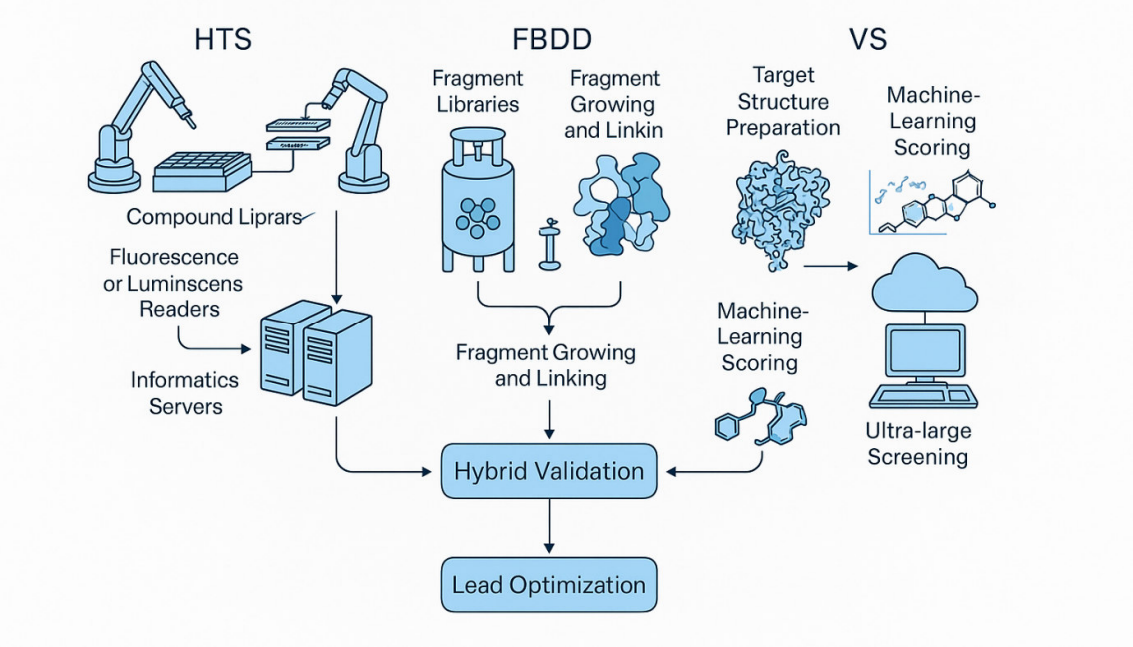

Figure 1 Workflow Integration of High-Throughput, Fragment-Based, and Virtual Screening Approaches

12.7 Future Perspectives: AI, Automation, and Sustainable Screening

The future of hit identification lies in the complete digitization and automation of discovery workflows, guided by artificial intelligence and sustainability principles. Emerging trends indicate a shift from static, single-target screening toward adaptive, data-driven exploration of polypharmacological space. AI-driven active learning platforms already close the loop between prediction and experiment. Algorithms iteratively select the most informative compounds for synthesis and testing, thereby optimizing data acquisition. Coupled with cloud robotics and microfluidic synthesis systems, such closed-loop laboratories can autonomously execute screening campaigns 24/7, reducing time from target selection to hit confirmation from years to weeks [34].

Quantum computing is another transformative frontier. Quantum machine learning models can simulate electronic structures and binding affinities with near-exact accuracy for small fragments, potentially replacing heuristic scoring functions. Early demonstrations by Google Quantum AI and IBM Qiskit have shown promise in predicting molecular orbitals for fragment–protein complexes [35]. Sustainability and green chemistry considerations are also emerging as critical drivers. Automated

screening systems that minimize reagent waste, use recyclable solvents, and employ energy-efficient computation align with environmental goals and reduce operational costs. Moreover, AI-guided compound selection can avoid redundant testing of similar molecules, optimizing resource utilization [36].

Finally, the integration of omics and phenotypic data with screening outputs will transform hit identification into a systems-level discipline. Multi-modal screening combining transcriptomics, proteomics, and metabolomics readouts can reveal off-target networks and mechanism of action early, improving downstream translation and safety. In sum, the convergence of AI, automation, and sustainability is reshaping hit discovery from a trial-and-error endeavor into a self-improving, intelligent ecosystem. As digital and experimental frontiers blur, the next generation of CADD will rely on human–machine symbiosis to explore chemical and biological complexity more efficiently and ethically than ever before.

REFERENCES

1. Macarron R, Banks MN, Bojanic D, et al. Impact of high-throughput screening in biomedical research. *Nat Rev Drug Discov*. 2011;10(3):188–195.
2. Congreve M, Carr R, Murray C, Jhoti H. A rule of three for fragment-based lead discovery. *Drug Discov Today*. 2003;8(19):876–877.
3. Walters WP, Green J, Weiss JR, Murcko MA. What do all these HTS hits mean? *Nat Rev Drug Discov*. 2011;10(4):290–301.
4. Hopkins AL, Groom CR, Alex A. Ligand efficiency: a useful metric for lead selection. *Drug Discov Today*. 2004;9(10):430–431.
5. Inglese J, et al. High-throughput screening assays in chemical biology. *Nat Chem Biol*. 2007;3(8):466–479.
6. Jadhav A, et al. Quantitative high-throughput screening: a titration-based approach. *Proc Natl Acad Sci USA*. 2010;107(4):1490–1495.
7. Zhang JH, Chung TD, Oldenburg KR. A simple statistical parameter for use in evaluation and validation of high throughput screening assays. *J Biomol Screen*. 1999;4(2):67–73.
8. Ziegler S, et al. Mining chemical space for bioactive molecules. *Angew Chem Int Ed Engl*. 2013;52(11):2744–2792.
9. Schneider G. Automating drug discovery. *Nat Rev Drug Discov*. 2018;17(2):97–113.
10. Erlanson DA, et al. Twenty years on: the impact of fragments on drug discovery. *Nat Rev Drug Discov*. 2020;19(7):1–17.
11. Murray CW, Rees DC. The rise of fragment-based drug discovery. *Nat Chem*. 2009;1(3):187–192.
12. Souers AJ, et al. ABT-199, a potent and selective BCL-2 inhibitor. *Nat Med*. 2013;19(2):202–208.
13. Imrie F, et al. Deep generative models for fragment linking. *J Chem Inf Model*. 2020;60(4):1983–1995.
14. Hajduk PJ, Greer J. A decade of fragment-based drug design: strategic advances and lessons learned. *Nat Rev Drug Discov*. 2007;6(3):211–219.
15. Shoichet BK. Virtual screening of chemical libraries. *Nature*. 2004;432(7019):862–865.
16. Morris GM, Huey R, Lindstrom W, et al. AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J Comput Chem*. 2009;30(16):2785–2791.
17. Willett P. Similarity-based virtual screening using 2D fingerprints. *Drug Discov Today*. 2006;11(23-24):1046–1053.

18. Genheden S, Ryde U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opin Drug Discov*. 2015;10(5):449–461.
19. McNutt AT, et al. GNINA 1.0: molecular docking with deep learning. *J Cheminform*. 2021;13(1):43.
20. Achdout H, et al. COVID Moonshot: open science discovery of SARS-CoV-2 main protease inhibitors. *Science*. 2022;375(6580):671–676.
21. Warren GL, Andrews CW, Capelli AM, et al. A critical assessment of docking programs and scoring functions. *J Med Chem*. 2006;49(20):5912–5931.
22. Bajusz D, et al. Integrating virtual and high-throughput screening. *Front Chem*. 2022;10:864921.
23. Schindler CE, et al. Large-scale assessment of docking and scoring algorithms for fragment screening. *J Chem Inf Model*. 2020;60(11):5450–5462.
24. Besnard J, et al. Automated design of ligands to polypharmacological profiles. *Nature*. 2012;492(7428):215–220.
25. Olivecrona M, et al. Molecular de novo design through deep reinforcement learning. *J Cheminform*. 2017;9(1):48.
26. Sterling T, Irwin JJ. ZINC 20 A free tool to discover chemistry for biology. *J Chem Inf Model*. 2015;55(11):2324–2337.
27. Goodnow RA, et al. DNA-encoded chemistry: enabling the deeper sampling of chemical space. *Nat Rev Drug Discov*. 2017;16(2):131–147.
28. Hann MM, Keserü GM. Finding the sweet spot: the role of hybrid methods in modern screening. *Nat Rev Drug Discov*. 2022;21(6):381–398.
29. Roberts AW, et al. Targeting BCL2 with venetoclax in relapsed chronic lymphocytic leukemia. *N Engl J Med*. 2016;374(4):311–322.
30. Quintás-Cardama A, et al. Ruxolitinib for myelofibrosis. *Blood*. 2011;117(13):3455–3462.
31. Douangamath A, et al. Crystallographic and fragment screening of SARS-CoV-2 main protease. *Nat Commun*. 2020;11(1):5047.
32. Knutson SK, et al. Selective inhibition of EZH2 by EPZ-6438 leads to potent antitumor activity. *Nat Chem Biol*. 2013;9(9):643–650.
33. Wallach I, et al. AtomNet: a deep convolutional neural network for bioactivity prediction. *arXiv preprint arXiv:1510.02855*. 2015.
34. Griffiths RR, et al. Accelerating materials and drug discovery with self-driving laboratories. *Nat Rev Chem*. 2023;7(3):169–187.
35. Cao Y, et al. Quantum chemistry in the age of quantum computing. *Chem Rev*. 2019;119(19):10856–10915.
36. Patel S, et al. Greener pharmaceutical discovery through AI and automation. *Trends Pharmacol Sci*. 2024;45(2):97–111.