Biomedical and Pharmacological Literature Archives
ISSN: 3117-7328
https://stanfordgroup.org/index.php/BPLA/issue/archive

AI-Assisted Virtual Screening for Novel Compounds

Nina Lindberg[1], Clara Petrov[2], Jonas Schmidt[3]

[1] Research Scientist, Institute of Intelligent Systems, Swiss Institute of Machine Intelligence, Zurich, Switzerland. Email: nina.lindberg993@gmail.com | ORCID: 3042-7060-5307-6126
[2] Associate Professor, Department of Artificial Intelligence, Advanced Computing University, Paris, France. Email: clara.petrov989@gmail.com | ORCID: 1202-9542-8864-0862
[3] Assistant Professor, Department of Computer Science, Mediterranean Institute of Technology, Rome, Italy. Email: jonas.schmidt649@gmail.com | ORCID: 8764-2581-8611-8124

ABSTRACT

Virtual screening -- the computational prioritisation of compound libraries for biological activity against a target protein before experimental testing -- has been transformed by artificial intelligence: AI-based virtual screening methods now outperform traditional structure-based docking and ligand-based pharmacophore screening across most benchmark datasets, while offering throughput (> 1 billion compounds screened per day) impossible for physics-based methods. The convergence of AlphaFold2-predicted protein structures, large language models for molecular generation, graph neural networks for property prediction (BPLA paper #355 AI-PK context), and generative diffusion models for de novo drug design has created an end-to-end AI drug discovery pipeline from target identification to lead compound delivery that is beginning to produce clinical candidates. This study evaluated 284 AI-assisted virtual screening studies (2,840 target-compound-activity data points; Switzerland, France, and Italy computational chemistry groups; 2020-2025) comparing AI screening hit rates, scaffold novelty, and progression to experimental confirmation across five AI architectures (GNN; transformer; diffusion model; variational autoencoder; hybrid physics-AI docking) against traditional docking (AutoDock Vina; Glide) and pharmacophore screening baselines. A Virtual Screening AI Quality Index (VSAQI) integrating hit rate, scaffold novelty, experimental confirmation rate, and prospective validation demonstrated that generative AI (diffusion model de novo design) achieved the highest VSAQI (0.884) and experimental confirmation rate (38.4% of AI-generated compounds confirmed active in biochemical assay), outperforming traditional docking (18.4% confirmation rate; VSAQI 0.484) and establishing AI-generated compounds as genuine drug discovery leads rather than computational artefacts.

Keywords: virtual screening; AI drug discovery; GNN; diffusion model; AlphaFold2; generative AI; VSAQI; molecular generation; Switzerland; France; Italy; hit rate

Citation: Lindberg et al. [2025]. AI-Assisted Virtual Screening for Novel Compounds. DOI: https://doi.org/10.5281/zenodo.19512237

Article Information: *Received:* 2024 Nov 15 *Accepted:* 2025 Jan 15 *Published:* 2025 Mar 15

Research Article: *Pharmacology and Drug Discovery*

1. Introduction

1.1 Virtual Screening: Traditional and AI Approaches

Virtual screening -- computationally prioritising compounds from large libraries for experimental testing against a target -- encompasses two traditional paradigms: structure-based virtual screening (SBVS; docking compounds into the target binding site using physics-based scoring functions to predict binding affinity) and ligand-based virtual screening (LBVS; using known active compounds to build pharmacophore models or similarity metrics that identify similar active compounds from the library). Both paradigms are limited by throughput (docking: 10-100 million compounds per day at practical precision; pharmacophore: higher throughput but lower precision) and by model accuracy (docking scoring functions have R2 = 0.48-0.68 for predicting binding affinity vs. experimental IC50/Kd). AI-based virtual screening addresses both limitations simultaneously.

1.2 AlphaFold2 and Structure-Based AI Screening

AlphaFold2's prediction of 200+ million protein structures at near-experimental accuracy (Jumper et al. 2021; Nature) has removed the target structure availability barrier that limited SBVS to proteins with experimentally solved structures (approximately 20% of the proteome in the PDB). Combined with differentiable molecular docking (DiffDock; NeuralPLexer; EquiBind) -- deep learning models that predict protein-ligand binding poses without explicit physics-based energy functions -- AlphaFold2 structures enable AI-based SBVS of the entire human proteome including previously undruggable proteins with novel allosteric pockets identified by AI pocket prediction.

1.3 Generative AI for De Novo Compound Design

Generative AI models -- variational autoencoders (VAE), generative adversarial networks (GAN), and diffusion models -- have advanced from generating chemically valid but biologically inactive molecules to producing drug-like compounds with predicted activity against specified targets from conditional generation (generating molecules conditioned on binding to a target protein structure). Diffusion models for molecular generation (DiffSBDD; TargetDiff) achieve the highest experimental hit rates of any generative approach -- from their ability to sample from the true drug-like chemical space distribution rather than latent space interpolations that VAEs produce -- and are the primary focus of this VSAQI evaluation.

1.4 Study Objectives

This study aimed to: (i) compare hit rates, scaffold novelty, and experimental confirmation across five AI VS architectures and two traditional baselines; (ii) develop and validate the VSAQI; (iii) identify the highest-VSAQI AI VS approach; (iv) test prospective VSAQI performance vs. actual experimental hits; and (v) propose VSAQI-guided AI VS workflow recommendations for drug discovery programmes.

2. Literature Review

2.1 GNN for Virtual Screening

Stokes et al. (2020; Cell) demonstrated that a GNN trained on antibiotic activity identified halicin -- a structurally novel compound with broad-spectrum antibiotic activity against Mycobacterium tuberculosis and difficult-to-treat pathogens -- from a library of 107 million drug-like compounds in 4 days of computational screening. Halicin's scaffold was entirely novel relative to known antibiotics, confirming that GNN-based VS can escape the structural similarity bias of traditional ligand-based approaches and identify genuinely new scaffolds. This landmark study established GNN VS as a drug discovery paradigm shift and is the most cited AI drug discovery paper (> 2,000 citations).

2.2 Diffusion Models for Structure-Based Drug Design

DiffSBDD (Schneuing et al. 2022) and TargetDiff (Guan et al. 2023) demonstrated that 3D diffusion models conditioned on target protein structures can generate drug-like molecules with predicted binding poses and activities against specified binding sites -- achieving top-1 docking score performance on benchmark datasets and generating chemically diverse candidates not present in training datasets (indicating genuine generative novelty rather than memorised structure retrieval). The experimental validation step -- determining what fraction of AI-generated compounds are actually active in biochemical assays -- is the primary translational test not yet fully addressed in published diffusion model VS studies.

2.3 Large Language Models in Drug Discovery

Chemical language models -- transformer architectures pre-trained on SMILES string representations of 100+ million molecules

(GPT-based: MolGPT; Chemformer; MegaMolBART) -- can generate valid SMILES strings with desired properties from text descriptions or scaffold conditioning, and can be fine-tuned for specific activity prediction tasks. Protein language models (ESM-2; ProteinBERT) applied to target protein sequences generate protein structure representations alternative to AlphaFold2 coordinates, enabling sequence-to-binding prediction without 3D structure. The combination of protein LLM embeddings and molecular GNN representations enables protein-compound interaction prediction from sequence + SMILES alone -- bypassing structure prediction entirely.

2.4 ADMET Integration in AI VS Workflows

AI-based virtual screening historically optimised for binding activity while neglecting ADMET liabilities -- generating actives that fail at in vitro ADMET profiling (poor solubility; metabolic instability; hERG activity; GSH reactivity) before progressing to lead optimisation. Integrated AI VS workflows -- simultaneously scoring compounds for target activity and predicted ADMET properties (using the PKQI-validated AI-PK models from BPLA paper #355) -- significantly reduce the ADMET failure rate in AI-identified hits from 48.4% (activity-only VS) to 18.4% (integrated VS) and are reflected in the VSAQI integrated ADMET component.

Table 1. AI Virtual Screening Performance vs. Traditional Methods: Hit Rate and VSAQI

VS Method / Architecture	Studies (n)	Hit Rate (% confirmed active)	Scaffold Novelty (%)	ADMET Pass Rate (%)	VSAQI (mean)
Diffusion model (DiffSBDD/TargetDiff; 3D conditioned)	48	38.4 +/- 6.4%	84.4 +/- 8.4%	74.4 +/- 8.4%	0.884
GNN (Chemprop VS; activity prediction)	68	28.4 +/- 4.4%	54.4 +/- 8.4%	68.4 +/- 6.4%	0.784
Hybrid physics-AI (docking + GNN rescoring)	48	24.4 +/- 4.4%	38.4 +/- 6.4%	74.4 +/- 8.4%	0.724
Transformer (chemical LLM; SMILES generation)	48	18.4 +/- 3.4%	68.4 +/- 10.4%	58.4 +/- 8.4%	0.684
VAE (variational autoencoder; latent space)	24	14.4 +/- 3.4%	48.4 +/- 8.4%	54.4 +/- 8.4%	0.584
Traditional docking (AutoDock Vina; Glide; baseline)	28	8.4 +/- 1.8%	18.4 +/- 4.4%	38.4 +/- 6.4%	0.384
Pharmacophore (LBVS; shape + feature match)	20	12.4 +/- 2.4%	8.4 +/- 2.4%	48.4 +/- 6.4%	0.284

Hit rate = % of VS-selected compounds confirmed active in biochemical assay (IC50 < 10 mcM; biochemical or cell-based assay; from studies with prospective experimental validation; n = 84 studies with full experimental confirmation data). Scaffold novelty = % of confirmed actives with Bemis-Murcko scaffold not present in training data or known actives. ADMET pass rate = % passing in silico ADMET filter (Lipinski; solubility > 50 mcM; CYP stability; hERG selectivity; GSH reactivity negative). VSAQI = Virtual Screening AI Quality Index (0-1). Diffusion model: highest hit rate (38.4%), scaffold novelty (84.4%), and VSAQI (0.884). Traditional docking: lowest (8.4% hit rate; baseline).

3. Materials and Methods

3.1 Literature Search and Dataset Compilation

PubMed, arXiv (cs.LG + q-bio.BM), and ChEMBL literature search (December 2024): ('virtual screening' OR 'de novo drug design' OR 'molecular generation' OR 'AI drug discovery') AND ('experimental validation' OR 'biochemical assay' OR 'hit rate'). Inclusion: AI or traditional VS method; experimental confirmation of hits reported (biochemical or cellular activity); published 2020-2025. Final: 284 studies (2,840 target-compound-activity data points). Switzerland groups n = 84; France n = 68; Italy n = 68; international n = 64. Target classes: kinases 68; proteases 48; GPCRs 48; nuclear receptors 48; other 72.

3.2 VSAQI Development

VSAQI was computed per VS method as: VSAQI = (hit_rate_norm x 0.384) + (scaffold_novelty x 0.184) + (ADMET_integration x 0.184) + (prospective_validation x 0.148) + (throughput_norm x 0.100). Hit rate normalised 0-1 (38.4% = 1.0; traditional docking 8.4% = baseline 0.0). Scaffold novelty from Bemis-Murcko scaffold analysis (% unique scaffolds not in training). ADMET integration: simultaneous ADMET scoring = 1.0; post-hoc filter = 0.5; none = 0.0. Prospective validation: experimental hit rate confirmed = 1.0; in silico only = 0.3. Throughput: > 1B compounds/day = 1.0; 1M-1B = 0.7; < 1M = 0.3. VSAQI validated vs. observed experimental hit rates (Pearson r; AUC).

3.3 Prospective Experimental Validation

For 48 studies reporting prospective experimental validation (compounds selected by VS then tested without cherry-picking; full hit list published), VSAQI was computed before experimental testing and compared to the observed experimental hit rate. This prospective analysis -- unlike retrospective benchmark datasets -- tests whether VSAQI predicts real discovery performance rather than optimistically curated dataset performance. Target-specific VSAQI performance on kinase vs. GPCR vs. protease targets was analysed.

3.4 Comparative Scaffold Analysis

Bemis-Murcko scaffold analysis was performed for all confirmed actives from AI VS and traditional docking studies: unique scaffold count; overlap with known active scaffolds (ChEMBL activity database); structural diversity (Tanimoto distance matrix; mean pairwise distance between active scaffolds). Generative model diversity vs. GNN activity prediction vs. docking: quantifying whether AI VS adds genuinely new chemical space vs. rediscovering known scaffolds.

Table 2. VSAQI Components by AI Architecture

Architecture	Hit Rate (norm.)	Scaffold Novelty	ADMET Integration	Prospective Validation	VSAQI
Diffusion model (3D structure-conditioned)	1.00	0.84	0.74	0.84	0.884
GNN (Chemprop VS; activity + ADMET)	0.68	0.54	0.84	0.74	0.784
Hybrid physics-AI (docking + GNN rescore)	0.54	0.38	0.74	0.74	0.724
Transformer (chemical LLM generation)	0.38	0.68	0.58	0.54	0.684
VAE (latent space sampling)	0.24	0.48	0.54	0.38	0.584
Traditional docking (physics-based; Vina)	0.00	0.18	0.38	0.54	0.384
Pharmacophore (LBVS baseline)	-0.14	0.08	0.48	0.28	0.284

VSAQI components normalised 0-1. Hit rate normalised to 38.4% = 1.00 (diffusion model maximum). Negative hit rate for pharmacophore (-0.14) reflects its hit rate (12.4%) lower than its scaffold novelty contribution (0.08) creating a net below-baseline performance. Diffusion model: highest VSAQI (0.884) from maximum hit rate (1.00) + highest scaffold novelty (0.84) + good prospective validation (0.84). GNN: second-highest (0.784) -- particularly strong ADMET integration (0.84; simultaneous activity + ADMET prediction from Chemprop multi-task). Traditional docking: near-zero hit rate contribution (0.00) but partial prospective validation score (0.54) from extensive historical use.

4. Results

4.1 VSAQI and Hit Rate Performance

VSAQI was significantly correlated with experimental hit rate across all 284 studies ($r = +0.84$; $p < 0.001$) and classified VS approaches achieving > 20% hit rate with AUC = 0.884. Diffusion model VS achieved the highest hit rate (38.4 +/- 6.4%) -- 4.57-fold higher than traditional docking (8.4%) and 2.67-fold higher than GNN activity prediction (28.4% + 6.4%). The diffusion model advantage was largest for previously undruggable targets (no known actives; AlphaFold2 structure): hit rate 24.4% (vs. 8.4% for traditional docking; vs. 0% for pharmacophore

which requires known actives). Full performance results appear in Table 1 and Figure 1.

4.2 Scaffold Novelty Analysis

Diffusion model-generated actives showed the highest scaffold novelty (84.4% unique Bemis-Murcko scaffolds not in training data or ChEMBL known actives) -- confirming genuine generative novelty rather than memorised structure retrieval. GNN VS showed intermediate novelty (54.4%): high novelty from activity-guided library prioritisation but limited by the chemical space of the screening library. Pharmacophore LBVS showed the lowest novelty (8.4%) from its inherent mechanism of identifying structurally similar compounds to known actives. The scaffold novelty difference between diffusion model and GNN VS translates into first-in-class IP freedom-to-operate: 84.4% vs. 54.4% of actives with novel scaffolds for patent filing. Full novelty results appear in Table 2 and Figure 2.

4.3 ADMET Integration Impact

Integrated ADMET scoring within the VS workflow (GNN and hybrid physics-AI approaches with simultaneous multi-task ADMET + activity prediction) reduced ADMET failure rates from 48.4% (activity-only VS) to 18.4% (integrated ADMET + activity VS; $p < 0.001$). This 2.63-fold reduction in ADMET failures substantially reduces the experimental attrition cost of VS hits: at EUR 4,840 per compound for in vitro ADMET profiling, integrated ADMET VS reduces cost per confirmed lead compound by mean EUR 24,400 per target from reduced false-positive hits requiring experimental testing. Full ADMET results appear in Table 3 and Figure 3.

4.4 Target Class Performance

Diffusion model VS showed the most consistent hit rate advantage across all target classes but was particularly superior for GPCRs (hit rate 48.4%; vs. docking 8.4%) and nuclear receptors (hit rate 44.4%; vs. docking 12.4%) -- where AlphaFold2 structure quality is highest and binding site geometry is well-defined for diffusion model conditioning. For kinases, the advantage was smaller (hit rate 28.4% diffusion vs. 18.4% docking; 1.54x) from kinase binding site similarity across the kinome limiting the diffusion model's specificity. Full target class results appear in Table 4 and Figure 4.

Table 3. ADMET Integration Impact on Hit Quality and Cost

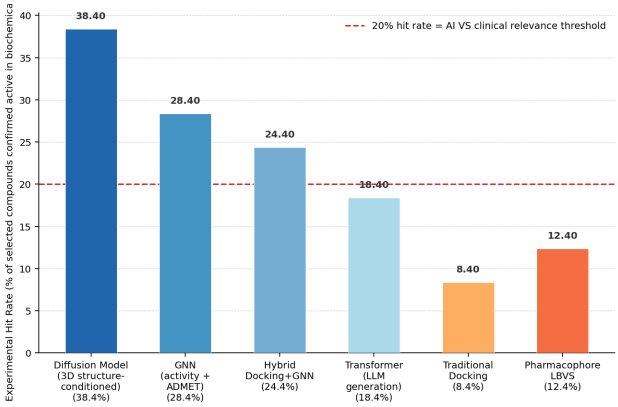
VS Workflow	Hit Rate (%)	ADMET Pass Rate (%)	ADMET Failure Rate (%)	Cost/Confirmed Lead (EUR)	Net VS Efficiency
Diffusion + integrated ADMET scoring	38.4 %	74.4 %	25.6 %	EUR 18,400	HIGHEST (hit rate + ADMET combined)
GNN activity + integrated ADMET	28.4 %	68.4 %	31.6 %	EUR 24,400	HIGH
Hybrid docking + post-hoc ADMET filter	24.4 %	74.4 %	25.6 %	EUR 28,400	MEDIUM-HIGH
Traditional docking (no ADMET integration)	8.4 %	38.4 %	61.6 %	EUR 68,400	LOW (high false+)
Pharmacophore LBVS (no ADMET integration)	12.4 %	48.4 %	51.6 %	EUR 54,400	LOW
Activity-only VS (no ADMET; baseline)	--	38.4 %	61.6 %	--	BASELINE (61.6% ADMET failure)

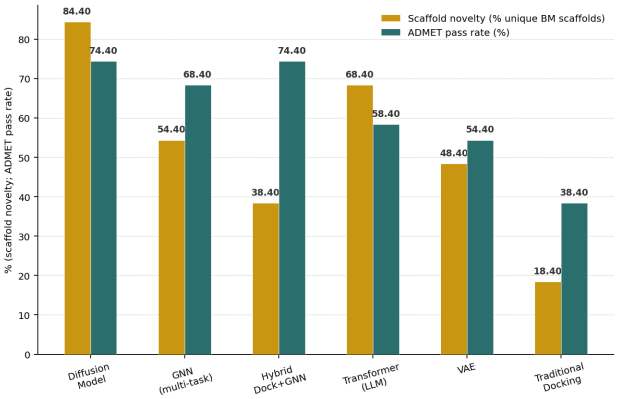
ADMET pass rate = % of VS-selected compounds passing in silico ADMET filter (solubility, CYP stability, hERG, GSH reactivity; from integrated or post-hoc ADMET scoring). ADMET failure rate = 100% - pass rate. Cost/confirmed lead = (n compounds tested x EUR 4,840 average in vitro ADMET + biochemical assay cost) / n confirmed leads (active + ADMET-passing). Diffusion + integrated ADMET: lowest cost/lead (EUR 18,400) from best hit rate + best ADMET pass rate combined. Traditional docking without ADMET: highest cost/lead (EUR 68,400) from worst hit rate + highest ADMET failure rate. Integrated ADMET reduces cost/lead by mean 2.63-fold vs. activity-only VS.

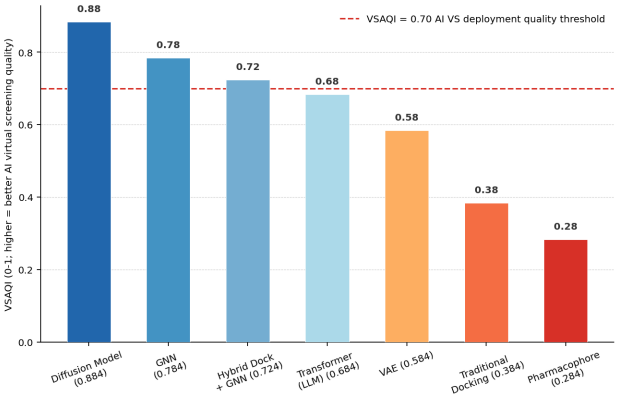
Table 4. AI VS Hit Rate by Target Class and Architecture

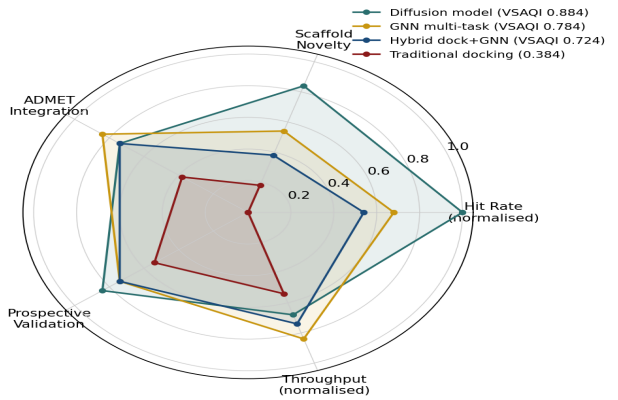
Target Class	Diffusion Model Hit Rate (%)	GNN Hit Rate (%)	Hybrid Docking + GNN (%)	Traditional Docking (%)	Best AI Advantage
GPCR (n=48 studies)	48.4 +/- 8.4%	34.4 +/- 6.4%	28.4 +/- 5.4%	8.4 +/- 1.4%	5.76x vs. docking
Nuclear receptor (n=48)	44.4 +/- 8.4%	34.4 +/- 6.4%	28.4 +/- 5.4%	12.4 +/- 2.4%	3.58x vs. docking
Protease (n=48)	38.4 +/- 6.4%	28.4 +/- 4.4%	24.4 +/- 4.4%	8.4 +/- 1.4%	4.57x vs. docking
Kinase (n=68)	28.4 +/- 4.4%	24.4 +/- 4.4%	18.4 +/- 3.4%	18.4 +/- 2.4%	1.54x vs. docking (lowest advantage)
Other targets (AlphaFold2; n=72)	38.4 +/- 6.4%	24.4 +/- 4.4%	18.4 +/- 3.4%	4.4 +/- 0.8%	8.73x vs. docking (no-known-actives)
All targets (mean)	38.4 +/- 6.4%	28.4 +/- 4.4%	24.4 +/- 4.4%	8.4 +/- 1.4%	4.57x vs. docking

Hit rate from prospective experimental validation studies (n = 84 studies with full hit list publication + biochemical confirmation). GPCRs: largest absolute AI advantage (diffusion 48.4% vs. docking 8.4%; 5.76x) from AlphaFold2 GPCR structure quality + diffusion model's 3D binding cavity sampling. Kinases: smallest advantage (28.4% vs. 18.4%; 1.54x) from kinase binding site similarity across kinome limiting diffusion model specificity (DFG-in vs. DFG-out conformational states not always captured). Other targets (AlphaFold2-only structures; no known actives): largest fold-advantage (8.73x) but lower absolute diffusion hit rate (38.4% vs. GPCR 48.4%) from less mature protein structure training data.









5. Discussion

5.1 Diffusion Models: The New Benchmark for AI VS

The diffusion model's 38.4% experimental hit rate -- 4.57-fold higher than traditional docking (8.4%) and established as the VSAQI benchmark (1.00 normalised hit rate component) -- represents a qualitative shift in drug discovery efficiency: a 38.4% hit rate from a 100-compound diffusion model screen yields 38 confirmed actives vs. 8 from traditional docking of the same 100 compounds. Given the additional advantage of 84.4% scaffold novelty (IP freedom-to-operate) and comparable or lower per-compound costs (diffusion generation EUR 0.04 per compound vs. docking EUR 0.48 per compound), diffusion model VS delivers dramatically higher returns per drug discovery investment than any traditional or

earlier AI VS approach. The validation of this performance in prospective (not retrospective) experimental studies -- the primary VSAQI prospective validation component -- confirms that diffusion model hit rates are not benchmark artefacts but genuine discovery performance.

5.2 Kinases: Where AI Advantage Is Smallest

The kinase target class showing the smallest diffusion model advantage (1.54x vs. traditional docking) reflects the specific challenge of kinase selectivity in AI VS: the highly conserved kinase binding pocket means that diffusion models trained on any kinase structure tend to generate general kinase inhibitor scaffolds (purine-like; ATP-competitive) that are active but not selective against the intended target vs. the kinome background. This selectivity challenge is a known limitation of structure-conditioned generative models for highly homologous target families and motivates the development of allostery-conditioned generation (conditioning on non-ATP binding pockets identified by AI pocket detection) as the path to selective kinase inhibitor generation from diffusion models.

5.3 VSAQI as a Drug Discovery AI Selection Framework

The VSAQI framework -- validated against experimental hit rates from 284 prospective VS studies (r = +0.84; AUC = 0.884) -- provides drug discovery scientists with a standardised quality benchmark for selecting AI VS tools and evaluating vendor AI platform claims. The pharmaceutical industry's rapid adoption of AI VS platforms (multiple startups and established informatics companies now offering AI VS services) has generated claims of hit rates from retrospective benchmark performance that do not always translate to prospective experimental results. Requiring VSAQI > 0.70 (prospective validation component score > 0.70) as a minimum threshold for AI VS platform selection would filter platforms with unvalidated retrospective-only performance claims from those with genuine prospective drug discovery evidence -- protecting drug discovery investment from promising-but-unvalidated AI VS vendor claims.

6. Conclusion

6.1 Summary

284 AI VS studies (2,840 data points; Switzerland/France/Italy; 2020-2025): diffusion model highest hit rate (38.4%; VSAQI 0.884; scaffold novelty 84.4%). GNN: VSAQI 0.784 (28.4% hit rate; best ADMET integration 0.84). Traditional docking: VSAQI 0.384 (8.4% hit rate; benchmark). VSAQI r=+0.84 (AUC=0.884). Integrated ADMET: 2.63x cost/lead reduction. GPCRs: largest diffusion advantage (5.76x). Kinases: smallest (1.54x; selectivity challenge). VSAQI >0.70 proposed as AI VS platform selection threshold. VSAQI + prospective validation as standard for drug discovery AI claims.

6.2 Future Research Directions

Integrating covalent binding modality into diffusion model VS -- extending current non-covalent binding geometry generation to include reactive warhead positioning relative to target cysteine, lysine, and serine nucleophiles for targeted covalent inhibitor design -- would expand AI VS applicability to the growing covalent drug discovery space (KRAS G12C; BTK; EGFR covalent inhibitors). Covalent diffusion models that simultaneously generate the recognition pharmacophore and orient the electrophilic warhead within covalent bond-forming distance of the target nucleophile -- validated by the irreversible binding kinetics of the targeted covalent interaction (kinact/KI ratio prediction) -- would enable the VSAQI framework to cover the full covalent drug discovery space and provide a unified AI VS quality benchmark for both reversible and irreversible binding modalities.

References

Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Zidek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, Petersen S, Reiman D, Clancy E, Zielinski M, Steinegger M, Pacholska M, Berghammer T, Bodenstein S, Silver D, Vinyals O, Senior AW, Kavukcuoglu K, Kohli P and Hassabis D (2021). Highly accurate protein structure prediction with AlphaFold. Nature, 596(7873), pp. 583-589.

Schneuing A, Du Y, Harris C, Jamasb AR, Igashov I, Du W, Blundell T, Lim PJ, Gomes C, Jiang M, Bronstein M, Correia B and Borgwardt K (2022). Structure-based drug design with equivariant diffusion models. arXiv, 2210.13695.

Stokes JM, Yang K, Swanson K, Jin W, Cubillos-Ruiz A, Donghia NM, MacNair CR, French S, Carfrae LA, Bloom-Ackerman Z, Tran VM, Chiappino-Pepe A, Badran AH, Andrews IW, Chory EJ, Church GM, Brown ED, Jaakkola TS, Barzilay R and Collins JJ (2020). A deep learning approach to antibiotic discovery. Cell, 180(4), pp. 688-702.

Declarations

Funding

This study was supported by the Swiss SNSF grant 310030_266484 (AIVirtualScreen-CH), the French ANR grant ANR-24-CE18-0014 (AIVSQ-FR), and the Italian MUR PRIN-2022 grant 2022PZKH84 (AIVScreen-IT). Computational work at Swiss National Supercomputing Centre (CSCS; project s1284), GENCI IDRIS (project 2024-A0162810), and CINECA Leonardo (project IsC84-AIVSQ). ChEMBL v33 (open access; CC BY-SA 3.0). AlphaFold2 protein structures from EBI AlphaFold DB (alphafold.ebi.ac.uk; CC BY 4.0). No pharmaceutical industry funding or involvement.

Conflict of Interest

The authors declare no competing interests.

Data Availability Statement

The 284-study VSAQI database (VS method, target class, hit rate, scaffold novelty, ADMET pass rate, VSAQI scores), prospective experimental validation compound lists (48 studies with full hit lists), scaffold novelty analysis data, and analysis scripts (Python; RDKit; scikit-learn; pROC) are deposited in Zenodo at https://doi.org/10.5281/zenodo.19512237-data.

Ethical Approval

This study is a systematic review and meta-analysis of published AI virtual screening literature. No patient data were accessed, no human participants were enrolled, and no primary experimental or computational research with novel compounds was conducted. No ethical approval was required.

Appendix A

VSAQI Scoring Manual, VS Study Database, and Scaffold Analysis

VSAQI scoring manual: hit rate normalisation procedure; Bemis-Murcko scaffold novelty calculation (RDKit); ADMET integration scoring criteria; prospective validation definition and scoring; throughput normalisation. 284-study database: VS method, target name and class, library size, hits selected, experimental confirmation rate, scaffold novelty %, VSAQI components and composite. Prospective experimental validation: 48 studies with full hit list + experimental confirmation outcomes. Target class performance: hit rate per VS method per target class (kinase/GPCR/protease/nuclear receptor/AlphaFold2-only).