IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 7, April 2026
Impact Factor: 8.2

AI in Molecular Docking: Paradigms, Architectures, and Clinical Horizons (2024-2025)

Mithilesh M. Rathore[1], Vijay H. Masand[1], Pravin S. Bodkhe[1], Mahima M. Sisodiya[1], Chetan N. Jadhav[1]
[1] Department of Chemistry, Vidya Bharati Mahavidyalaya, Amravati, Maharashtra, India.
Corresponding author: hod_chemistry@vbmv.org

Abstract: *The pharmaceutical industry is currently witnessing a paradigm shift from traditional computer-aided drug design (CADD) to AI-driven drug discovery (AIDD), necessitated by the escalating costs and timelines of therapeutic development. This paper provides a comprehensive review of the evolution of molecular docking technologies through 2024 and 2025, contrasting established physics-based methodologies with emerging deep learning architectures. We trace the lineage of algorithmic innovation from classical search-and-score engines (AutoDock Vina) and CNN-based rescoring (GNINA) to the advent of Geometric Deep Learning. Specifically, we analyze the transition from rapid regression-based models like EquiBind and TankBind to SE(3)-equivariant diffusion models such as DiffDock, which treat docking as a generative denoising process.*

The review highlights the emergence of "co-folding" foundation models in 2024-2025, including AlphaFold 3 and the open-source Boltz-2, which integrate protein structure prediction with ligand binding to address the long-standing challenge of induced fit. Critical attention is devoted to benchmarking integrity; while models like FABFlex demonstrate massive speed advantages (208x faster than prior flexible dockers) and improved accuracy (40.59% success rate on ligand RMSD < 2Å), the "PoseBusters" benchmark reveals persistent issues with the physical validity of AI-generated structures compared to physics-based baselines. Finally, we explore the integration of generative AI with molecular dynamics (MD) simulations and the downstream clinical applications of these technologies, positioning AI not merely as a screening tool but as a foundational component of modern structural biology.

Keywords: Molecular docking, AI, CADD, AIDD

I. INTRODUCTION

The pharmaceutical industry currently faces a paradoxical trajectory known as "Eroom's Law", the observation that drug discovery becomes progressively slower and more expensive over time, despite exponential improvements in technology. The cost to bring a new therapeutic entity to market is estimated to exceed $2.5 billion, with timelines stretching over a decade. A critical bottleneck in this pipeline is the identification of small molecules that can bind with high affinity and specificity to a target protein, a process known as molecular docking. Traditionally, this domain has been dominated by physics-based simulations that, while rigorously grounded in thermodynamics, are computationally demanding and often struggle to navigate the immensity of chemical space. However, the convergence of geometric deep learning, generative artificial intelligence (AI), and massive structural datasets is currently precipitating a paradigm shift of historical magnitude.

As we traverse the mid-2020s, the field is witnessing the transition from "computer-aided drug design" (CADD) to "AI-driven drug discovery" (AIDD).[1] This is not merely a change in nomenclature but a fundamental restructuring of the scientific method applied to biology. The release of AlphaFold 2 in 2020 solved the static protein folding problem for all practical purposes, but the frontier has since moved to the dynamic and interactive realm of intermolecular recognition. The years 2024 and 2025 have been particularly pivotal, marked by the emergence of "co-folding" foundation models

DOI: 10.48175/IJARSCT-33003

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 7, April 2026 Impact Factor: 8.2

like AlphaFold 3 and Boltz-2, which do not simply "dock" a ligand into a rigid pocket but predict the entire multimolecular assembly from scratch.[2]

This report provides an exhaustive analysis of the state of AI in molecular docking as of 2025. It dissects the evolution from classical search-and-score algorithms to the latest SE(3)-equivariant diffusion models. We will explore the architectural innovations that allow machines to "learn" physics, the rigorous benchmarking battles that have exposed the limitations of early deep learning models, and the emerging frontiers where AI integrates with molecular dynamics to capture the breathing motions of life itself. Furthermore, we examine the downstream integration of these technologies into clinical development, including the use of digital twins and synthetic control arms.[4]

1.1 The Docking Problem: A Geometric and Energetic Puzzle

At its core, molecular docking is a high-dimensional optimization problem involving two primary challenges: pose prediction and affinity estimation. Pose prediction seeks the correct orientation and conformation (pose) of a small molecule (ligand) within the binding site of a protein (receptor). Affinity estimation attempts to quantify the strength of this interaction (binding free energy, ΔG).

Mathematically, the ligand's position is defined by its translation (x, y, z), rotation (quaternions or Euler angles), and internal torsion angles (τ). The search space is continuous and rugged, riddled with local minima where a solver might get trapped. Traditional methods address this by sampling thousands of conformations and evaluating them using a scoring function, a mathematical proxy for binding energy. AI approaches, conversely, have evolved from simple scoring improvements to end-to-end generative tasks, mapping the atomic graphs of the protein and ligand directly to **3D coordinates**.

The stakes are high. Accurate docking is the prerequisite for virtual screening, where billions of compounds are tested *in silico* to find "hits." False positives lead to wasted wet-lab resources, while false negatives mean missed cures. The integration of AI promises not just speed, accelerating screening from months to days, but a qualitative leap in our ability to target "undruggable" proteins and explore novel chemical spaces.[1]

II THEORETICAL FOUNDATIONS AND CLASSICAL METHODOLOGIES

To appreciate the trajectory of AI, one must first establish the baseline capabilities and limitations of physics-based methods. For over three decades, tools like AutoDock Vina, Glide, and GOLD have been the workhorses of the industry.

2.1 Classical Mechanics: The AutoDock Vina Standard

AutoDock Vina, and its various forks, represents the pinnacle of the classical approach. It utilizes a semi-empirical scoring function that sums contributions from steric interactions (Van der Waals forces), hydrogen bonding, and hydrophobic effects.[5] The search algorithm typically employs an iterated local search optimizer (e.g., Broyden-Fletcher-Goldfarb-Shanno or genetic algorithms) to explore the landscape of the ligand's degrees of freedom.

While robust, these methods suffer from inherent limitations:

Computational Cost: The iterative search is computationally expensive, scaling linearly with the number of rotatable bonds. High-throughput screening of billion-compound libraries requires massive supercomputing clusters.

Rigid Receptor Assumption: Most classical docking assumes the protein is a rigid statue. In reality, proteins are dynamic entities that undergo "induced fit" conformational changes upon ligand binding. Accounting for this flexibility via ensemble docking increases computational cost exponentially.

Scoring Function Inaccuracies: The approximate nature of scoring functions means they often fail to rank the correct pose as the lowest energy state, or they poorly correlate with experimental binding affinities (K_i or IC_{50}).[5]

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 7, April 2026 Impact Factor: 8.2

2.2 The First Wave: Machine Learning Scoring Functions

The initial entry of AI into this field was not to replace the docking engine, but to improve the scoring function. "Rescoring" methods emerged, where a classical tool (like Vina) generates poses, and a machine learning model evaluates them.

GNINA (Grid-based Neural Network Interaction Assessment) is a prime example of this hybrid era.[6] It utilizes 3D Convolutional Neural Networks (CNNs) to analyze the voxelized representation of the protein-ligand interface. Unlike classical functions that sum explicit terms, GNINA learns to recognize favorable interaction patterns directly from training data (PDBBind).

Mechanism: The protein-ligand complex is discretized into a 3D grid. Each voxel contains channels representing atom types (Carbon, Nitrogen, Oxygen, etc.). The CNN slides over this grid, identifying features like hydrogen bond geometry or pi-stacking without being explicitly programmed with the laws of physics.

Performance: Benchmarks consistently show that GNINA outperforms Vina in pose ranking.[6] It acts as a bridge, retaining the sampling robustness of Vina while leveraging the pattern-recognition power of deep learning.[8]

Limitation: It is still bound by the sampling limitations of the underlying classical engine. If Vina fails to sample the correct pose, GNINA cannot score it.

III. GEOMETRIC DEEP LEARNING: THE SHIFT TO GRAPH NEURAL NETWORKS (GNNS)

The "voxelization" approach of CNNs (like GNINA) has a flaw: empty space. 99% of a protein-ligand grid is empty, leading to computational inefficiency. Furthermore, grids are not rotationally invariant; rotating the molecule changes the input grid, requiring data augmentation to teach the network that a rotated molecule is still the same molecule.

This led to the adoption of **Geometric Deep Learning (GDL)**, specifically Graph Neural Networks (GNNs), which treat atoms as nodes and bonds/contacts as edges. These architectures are naturally sparse (ignoring empty space) and can be designed to be SE(3)-equivariant (understanding that the physics of the molecule doesn't change if you rotate the reference frame).[9]

3.1 EquiBind: Geometry Matching and Keypoints

EquiBind, developed by Stärk et al. (2022), represented a conceptual breakthrough. Instead of iteratively searching a landscape, it formulated docking as a direct regression problem.[9]

Architecture: It employs an SE(3)-equivariant Graph Neural Network (EGNN). The model processes the ligand graph and the receptor graph to predict "keypoints" (functional landmarks) on both structures.

Mechanism: Once keypoints are identified, the model applies the **Kabsch algorithm**, a method from linear algebra used to compute the optimal rotation matrix that aligns two sets of points. It essentially "snaps" the ligand into the pocket by aligning the predicted ligand keypoints with the receptor keypoints.

Impact: EquiBind was exponentially faster than Vina (approx. 1000x), enabling the screening of massive libraries in seconds.

Analysis: However, speed came at a cost. EquiBind treated the docking as a rigid body transformation. It often produced "physically implausible" structures, ligands passing through protein backbones or atoms overlapping (steric clashes).[9] It lacked the "physical intuition" to respect atomic radii, leading to poor performance on validity benchmarks like PoseBusters.[9]

3.2 TankBind: Trigonometry and Distance Matrices

Following EquiBind, **TankBind** (2022) offered a different geometric solution. Instead of keypoint alignment, it focused on predicting the **interatomic distance matrix**.[9]

Architecture: Trigonometry-aware Neural Networks.

Mechanism: The model predicts the pairwise distances between every atom of the ligand and every residue of the pocket. It effectively draws a map of "where every atom should be relative to the protein."

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 7, April 2026 Impact Factor: 8.2

Reconstruction: It then uses multidimensional scaling (optimization) to generate coordinates that satisfy these predicted distances.

Performance: Like EquiBind, TankBind excelled at identifying the binding pocket (blind docking) but struggled with the fine-grained atomic details required for high-fidelity pose prediction. The regression approach tends to output the "average" of possible conformations, which, in a multimodal distribution of poses, might result in a physically impossible structure (e.g., an atom floating halfway between two valid binding sites).[9]

IV. THE DIFFUSION REVOLUTION: GENERATIVE MODELS TAKE CENTER STAGE

The critical flaw of regression models (EquiBind, TankBind) is their tendency to predict the "mean" pose. Molecular docking is multimodal; a ligand might bind in Pose A or Pose B, but the average of A and B is a clash. This realization paved the way for **Generative AI**, specifically **Diffusion Models**, which are designed to sample from complex distributions rather than averaging them.

4.1 DiffDock: Denoising the Molecular Landscape

DiffDock (Corso et al., 2023) is widely considered the first "state-of-the-art" deep learning docking model that could legitimately challenge physics-based methods on blind docking tasks.[10]

The Diffusion Concept: In image generation (e.g., Midjourney), diffusion models learn to remove noise from a static image to reveal a clear picture. In DiffDock, the model learns to remove "structural noise" from a ligand. The training process takes a crystal structure, randomizes the ligand's position (translation, rotation) and conformation (torsion angles), and trains a network to reverse this process.

Architecture: It utilizes a **Score-Based Generative Model (SGM)** over the group of rigid transformations $SE(3)$ and the torus of torsion angles $SO(2)^m$.

SE(3)-EGNN: The core neural network is an SE(3)-equivariant GNN that predicts a "score" (a vector field) pointing towards the high-probability binding pose.

Langevin Dynamics: During inference, the model starts with a random ligand conformation and iteratively updates the position using the predicted score, effectively "sliding" the molecule into the binding pocket.

Performance: DiffDock achieved a massive leap in accuracy on the PDBBind benchmark compared to EquiBind and TankBind. Crucially, it produced significantly more physically plausible structures because it samples *explicit* poses rather than averaging them.[9]

Blind Docking Supremacy: DiffDock excels at "blind docking", scenarios where the binding site is unknown. While Vina requires a user-defined "box" (search space), DiffDock processes the whole protein surface. Studies show that while traditional methods struggle without a defined box, diffusion models can locate cryptic pockets with high success rates.[9]

4.2 Limitations of the Diffusion Paradigm

Despite its success, the original DiffDock had limitations:

Rigid Protein: Like Vina, it treated the protein as a rigid object. This fails in cases of significant induced fit (e.g., loop movements).

Hallucinations: While better than regression models, it can still generate poses with subtle steric clashes or improbable bond angles if the score function is not perfectly calibrated.[12]

Generalization: Performance drops significantly on protein families not well-represented in the PDBBind training set (e.g., ion channels vs. kinases).[9]

V. THE ERA OF CO-FOLDING AND FOUNDATION MODELS (2024-2025)

The distinction between "protein structure prediction" (AlphaFold) and "molecular docking" (Vina/DiffDock) began to blur in 2024. The new paradigm is **Co-Folding**: predicting the structure of the protein and the ligand *simultaneously* from

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 7, April 2026 Impact Factor: 8.2

the amino acid sequence and the SMILES string (chemical formula). This allows the AI to model the protein accommodating the ligand (induced fit) naturally.

5.1 AlphaFold 3 (AF3): The Unified Biomolecular Model

Released by Google DeepMind in 2024, **AlphaFold 3** represents a monumental leap. It is not just a protein folder; it is a general-purpose biomolecular structure predictor.[2]

Architecture: AF3 replaces the "Evoformer" module of AF2 with a simpler "Pairformer" and relies heavily on **Diffusion** for the final structure generation. It operates directly on atomic coordinates for proteins, DNA, RNA, and small molecules.

Performance on Ligands: On the rigorous **PoseBusters benchmark** (v1), AlphaFold 3 achieved a success rate of **76.4%** (defined as RMSD < 2Å and physically valid).

This compares to ~52.6% for Vina (with Gold docking).

The "blind" nature of AF3 is notable; it doesn't need a binding box. However, providing pocket information improves its accuracy by a further ~10%.[13]

Halogenated Ligands: For ligands containing fluorine or chlorine, common in modern drugs, AF3 shows robust performance, although "Strong Baselines" (Gnina-rescored Vina) can sometimes outperform it on these specific subsets.[13]

Implication: AF3 demonstrates that deep learning can outperform physics-based methods even on their home turf (accuracy), not just speed. It effectively handles the flexibility of the protein side-chains, a long-standing hurdle for Vina.

5.2 Boltz-2: The Open-Source Challenger

In mid-2025, a collaboration between MIT and Recursion Pharmaceuticals released **Boltz-2**, a fully open-source foundation model designed to rival AF3.[2]

Multimodal Capabilities: Boltz-2 creates a unified representation of proteins, DNA, RNA, and small molecules. It uses a **contrastive learning** objective to distinguish binders from non-binders (decoys), training on both positive and negative data.[15]

Accuracy: In the **Polaris Antiviral Drug Discovery 2025** competition, Boltz-2 demonstrated >80% docking accuracy on SARS-CoV-2 and MERS-CoV datasets, significantly outperforming traditional pipelines.[8]

Affinity Prediction: Perhaps most groundbreaking is Boltz-2's claim to approach the accuracy of **Free Energy Perturbation (FEP)** calculations for binding affinity. FEP is a rigorous physics simulation that takes days per molecule; Boltz-2 attempts to approximate this in seconds.[15]

Bias and Limitations: Independent benchmarks have noted that Boltz-2 can exhibit bias toward "canonical" protein shapes seen in training. For example, in the case of the **WRN helicase**, Boltz-2 struggled to predict the alternate pose induced by the inhibitor HRO-761 because the training data was dominated by ATP-bound structures.[17] This highlights that even foundation models can struggle to "escape" their training distribution without explicit prompting or templates.

Scalability: Being open-weights and optimized for inference, Boltz-2 is designed for "virtual screening at scale," enabling the processing of billion-compound libraries that would be cost-prohibitive with FEP or even Vina.[3]

5.3 FABFlex: Mastering Flexibility with Speed

While foundation models like AF3 are powerful, they are computationally heavy. For ultra-high-throughput screening, speed is paramount. **FABFlex** (Fast and Accurate Blind Flexible Docking), introduced in 2025, targets this niche.[18]

Approach: It uses a regression-based multi-task learning framework. It decomposes the problem into three modules: Pocket Prediction, Ligand Docking, and Pocket Docking (predicting the *holo* or bound shape of the pocket from the *apo* or unbound shape).

Iterative Update: A mechanism allows the ligand and pocket modules to communicate, refining the fit iteratively.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 7, April 2026 Impact Factor: 8.2

Performance: FABFlex boasts a **208x speed advantage** over DynamicBind (another flexible docker). It achieves a ligand RMSD < 2Å in **40.59%** of cases on standard benchmarks, outperforming many competitors in the "flexible blind docking" category.[19]

VI. BENCHMARKING AND CRITICAL EVALUATION

The rapid proliferation of models necessitated a "reality check." Early claims of "90% accuracy" often evaporated when models were tested on rigorous, independent datasets.

6.1 The PoseBusters Benchmark

The **PoseBusters** paper (2024) was a watershed moment. It criticized the ML community for focusing solely on RMSD (geometric proximity) while ignoring chemical reality. An AI might predict a ligand is in the right place (low RMSD), but with bond lengths of 5Å (impossible) or atoms overlapping.

The Test: PoseBusters evaluates "validity" (stereochemistry, planarity, clashes) alongside RMSD.

Results:

Deep Learning (Early): EquiBind and TankBind had validity scores near 0-2%. They essentially produced "ghost molecules".[9]

Diffusion (Middle): DiffDock improved validity to ~12-38% (depending on the version and strictness).

Foundation (Current): AlphaFold 3 achieved **76.4%** success (RMSD < 2Å + Valid), setting a new standard.[13]

Physics Baseline: Traditional methods like Vina/Gold naturally produce valid molecules (100% validity) because their sampling is constrained by physics, but they often get the *wrong* pose (lower RMSD success).

6.2 The Generalization Gap and Data Bias

A persistent critique of AI docking is "memorization." The PDBBind dataset, used to train almost all models, is heavily biased.

Kinase Bias: A large portion of the PDB consists of kinases. AI models perform exceptionally well on kinases but often fail on **GPCRs** or **Ion Channels**, which are underrepresented.[9]

Sequence Similarity: If the test set contains proteins similar to the training set, the AI might just be retrieving a memorized interaction pattern rather than learning the physics of binding.

Analysis: Comparisons using strict "time-split" or "cluster-split" cross-validation (like in the CASF-2016 benchmark) reveal that while AI models are catching up, physics-based methods (which don't require training data) still hold an advantage in generalization to completely novel protein folds.[7]

6.3 The "Strong Baseline" Challenge

In response to the AlphaFold 3 release, researchers at Inductive Bio demonstrated that a "Strong Baseline" constructed from open-source tools could approach SOTA performance. By combining **AutoDock Vina** (for sampling) with **GNINA** (for rescoring) and using an ensemble of starting ligand conformers, they achieved results within 5-10% of AF3 on many metrics, and actually outperformed the *blind* version of AF3 on specific subsets like "Other Ligands" (non-nucleotide/non-peptide).[13] This underscores that reports of the death of physics-based docking are premature; hybrid systems remain highly competitive.

VII. GENERATIVE AI AND DE NOVO DESIGN

The ultimate promise of AI is not just to screen existing libraries (which is limited by what we have synthesized), but to imagine new molecules.

7.1 From Finding to Designing

Generative AI (Diffusion, Flows, VAEs) allows for **target-aware molecule generation**.[22] Instead of docking a billion existing molecules, the AI generates a molecule *inside* the pocket, atom by atom, conditioned on the protein structure.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 7, April 2026 Impact Factor: 8.2

Pocket-Conditioned Generation: Models like **Pocket2Mol** or **DiffLinker** generate ligands that perfectly complement the electrostatic and geometric features of the binding site.
Novelty: A 2025 ACS review highlighted that structure-based generative models yield significantly higher structural novelty (17.9% with Tcmax > 0.4) compared to ligand-based models, which tend to generate variations of known drugs.[1]
Synthesizability: The major hurdle remains synthesizing these "dreamed" molecules. Integrated labs (e.g., Xaira Therapeutics, Recursion) are building closed loops where AI designs a molecule, robots synthesize it, and the biological result is fed back to retrain the AI.[2]

7.2 Case Studies in Generative Design

Recent literature highlights tangible successes:
Antibiotics: Generative models were used to design novel inhibitors for *Acinetobacter baumannii*, a multi-drug resistant pathogen. These compounds demonstrated *in vivo* efficacy, validating the generative approach beyond theoretical benchmarks.[23]
Multi-Target Design: In oncology, resistance often develops against single-target therapies. New generative models are being deployed for **multi-target drug design**, creating "poly-pharmacological" agents that can simultaneously inhibit multiple pathways (e.g., KRAS and MGLL), reducing the likelihood of resistance.[26]

VIII. INTEGRATION WITH MOLECULAR DYNAMICS (MD)

Static docking, even with flexible side-chains (AF3), captures only a snapshot of the binding event. Biological reality is dynamic. The integration of AI with Molecular Dynamics (MD) is the next frontier.

8.1 AI-Accelerated MD and Ensemble Docking

MD simulations solve Newton's equations of motion for every atom, capturing the "movie" of binding. However, they are femtosecond-step processes requiring nanoseconds or microseconds of data, making them computationally prohibitive for screening.
Surrogate Models: AI models are being trained on MD trajectories to predict the "next frame" or the "equilibrium state" thousands of times faster than physics solvers.
Ensemble Generation: A workflow described in recent literature involves using MD to generate a diverse ensemble of protein conformations (clustering trajectories). AI docking (like DiffDock or FABFlex) is then run across this ensemble. This "Ensemble Docking" significantly improves hit rates for difficult targets with cryptic pockets.[27]
Deep Learning Force Fields: AI is also improving the *accuracy* of MD itself. Neural network potentials (like ANI or DeepMD) approach quantum-mechanical accuracy at empirical force-field speeds, allowing for highly precise refinement of docked poses.[29]

IX. DOWNSTREAM HORIZONS: CLINICAL INTEGRATION

The impact of improved docking extends downstream into clinical development. As noted in recent reviews [4], the data generated by AI discovery platforms is feeding into:
Digital Twins: Virtual representations of patients used to simulate drug responses before actual administration.
Synthetic Control Arms: Using real-world data and AI modeling to simulate the placebo group in a clinical trial, potentially reducing the number of human participants required.
Patient Stratification: AI models that predict which patient genetic profiles will respond to the specific binding characteristics of the designed molecule.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 7, April 2026
Impact Factor: 8.2

X. COMPARATIVE ANALYSIS OF KEY ARCHITECTURES

Model Class	Representative Tools	Mechanism	Key Advantage	Primary Limitation
Physics-Based	AutoDock Vina, Glide	Search & Score (Genetic Algorithms)	Interpretable, physically valid poses, no training bias.	Slow, struggles with large scale, assumes rigid protein.
ML Scoring	GNINA	CNN on Voxel Grids	Improved ranking over Vina.	Bound by Vina's sampling; grid inefficiencies.
Geometric DL (Regression)	EquiBind, TankBind	SE(3)-GNN, Keypoints, Distance Matrix	Extremely fast (1000x Vina).	Low physical validity (clashes), averaging problem.
Generative Diffusion	DiffDock	SE(3)-Equivariant Diffusion (SDEs)	SOTA in blind docking, handles ligand flexibility well.	Can hallucinate interactions; rigid protein assumption (early versions).
Co-Folding Foundation	AlphaFold 3, Boltz-2	Multimodal Transformer + Diffusion	High accuracy (PoseBusters >76%) , handles induced fit.	Computationally heavy, data-hungry, "black box" nature.
Flexible Regression	FABFlex	Multi-task Learning	Fast flexible docking (208x DynamicBind).	Newer architecture, less community validation than AF3.

Table 1: Summary of major molecular docking paradigms as of 2025.

XI. CONCLUSION

The field of AI in molecular docking has evolved with breathtaking speed. In just five years, we have moved from CNNs that "looked" at voxel grids (GNINA) to diffusion models that "denoise" molecular geometry (DiffDock), and finally to foundation models that "co-fold" life's machinery (AlphaFold 3, Boltz-2).

The implications are profound. The "activation energy" for starting a drug discovery program is lowering. What used to require a cluster of supercomputers and weeks of time can now be approximated in minutes on a cloud GPU. However, the role of the expert is not diminishing but shifting. The focus is moving from *running* the dock to *curating* the data and *interpreting* the biological relevance of the AI's hallucinations.

As we look toward 2030, the integration of these digital minds with robotic wet-labs, the concept of "Self-Driving Laboratories", suggests a future where the bottleneck is no longer the *discovery* of a molecule, but the speed of clinical validation. The "hype cycle" [24] is real, but beneath it lies a bedrock of technological transformation that is irreversibly changing how we design medicine.

REFERENCES

1. Xie, S., Zhu, H., & Huang, N. (2025). AI-Designed Molecules in Drug Discovery, Structural Novelty Evaluation, and Implications. Journal of Chemical Information and Modeling, 65(17), 8924-8933. https://doi.org/10.1021/acs.jcim.5c00921
2. Gong, J., Zhao, Z., Niu, X., Ji, Y., Sun, H., Shen, Y., Chen, B., & Wu, B. (2025). AI reshaping life sciences: intelligent transformation, application challenges, and future convergence in neuroscience, biology, and medicine. Frontiers in Digital Health, 7. https://doi.org/10.3389/fdgth.2025.1666415

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 7, April 2026
Impact Factor: 8.2

3. Passaro, S., Corso, G., Wohlwend, J., Reveiz, M., Thaler, S., Somnath, V. R., Getz, N., Portnoi, T., Roy, J., Stark, H., Kwabi-Addo, D., Beaini, D., Jaakkola, T., & Barzilay, R. (2025). Boltz-2: Towards Accurate and Efficient Binding Affinity Prediction. https://doi.org/10.1101/2025.06.14.659707
4. Ocana, A., Pandiella, A., Privat, C., Bravo, I., Luengo-Oroz, M., Amir, E., & Gyorffy, B. (2025). Integrating artificial intelligence in drug discovery and early drug development: a transformative approach. Biomark Res, 13(1), 45. https://doi.org/10.1186/s40364-025-00758-2
5. Ma, Y., & Chen, X. (2025). Deep-Learning vs Physics-Based Docking Tools for Future Coronavirus Pandemics. Journal of Chemical Information and Modeling, 65(21), 12006-12013. https://doi.org/10.1021/acs.jcim.5c02029
6. McNutt, A. T., Li, Y., Meli, R., Aggarwal, R., & Koes, D. R. (2025). GNINA 1.3: the next increment in molecular docking with deep learning. Journal of Cheminformatics, 17(1). https://doi.org/10.1186/s13321-025-00973-x
7. Buccheri, R., & Rescifina, A. (2025). High-Throughput, High-Quality: Benchmarking GNINA and AutoDock Vina for Precision Virtual Screening Workflow. Molecules, 30(16). https://doi.org/10.3390/molecules30163361
8. Ma, Y., & Chen, X. (2025). Deep-Learning vs Physics-Based Docking Tools for Future Coronavirus Pandemics. Journal of Chemical Information and Modeling, 65(21), 12006-12013. https://doi.org/10.1021/acs.jcim.5c02029
9. Lee, J., Hao Nguyen, C., & Mamitsuka, H. (2025). Beyond rigid docking: deep learning approaches for fully flexible protein-ligand interactions. Brief Bioinform, 26(5). https://doi.org/10.1093/bib/bbaf454
10. Nguyen, N.-Q., & Kang, J. (2025). EquiCPI: SE(3)-Equivariant Geometric Deep Learning for Structure-Aware Prediction of Compound-Protein Interactions. Journal of Chemical Information and Modeling, 65(13), 7252-7262. https://doi.org/10.1021/acs.jcim.5c00773
11. Wang, Y., Ma, Y., Chang, Y., Yan, J., Zhang, J., Cai, M., & Wei, K. (2025). Diffusion Models at the Drug Discovery Frontier: A Review on Generating Small Molecules Versus Therapeutic Peptides. Biology, 14(12). https://doi.org/10.3390/biology14121665
12. Li, Y., Yi, J., Li, H., Li, K., Kang, F., Deng, Y., Wu, C., Fu, X., Jiang, D., & Cao, D. (2025). Decoding the limits of deep learning in molecular docking for drug discovery. Chemical Science, 16(37), 17374-17390. https://doi.org/10.1039/d5sc05395a
13. Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A. J., Bambrick, J., Bodenstein, S. W., Evans, D. A., Hung, C.-C., O'Neill, M., Reiman, D., Tunyasuvunakool, K., Wu, Z., Žemgulytė, A., Arvaniti, E., . . . Jumper, J. M. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature, 630(8016), 493-500. https://doi.org/10.1038/s41586-024-07487-w
14. Morehead, A., Giri, N., Liu, J., Neupane, P., & Cheng, J. (2025). Assessing the potential of deep learning for protein-ligand docking. ArXiv. https://www.ncbi.nlm.nih.gov/pubmed/38827451
15. Bret, G., Sindt, F., & Rognan, D. (2026). Assessing Boltz-2 Performance for the Binding Classification of Docking Hits. Journal of Chemical Information and Modeling. https://doi.org/10.1021/acs.jcim.5c02630
16. Lukauskis, D., Kashif-Khan, N., Tame, C., & Potterton, A. Optimizing Molecular Glues Using Free Energy Perturbation and Cofolding Methods. https://doi.org/10.26434/chemrxiv-2025-tb29n
17. Ille, A. M., Markosian, C., Burley, S. K., Pasqualini, R., & Arap, W. (2025). Human protein interactome structure prediction at scale with Boltz-2. https://doi.org/10.1101/2025.07.03.663068

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 7, April 2026
Impact Factor: 8.2

18. Zizhuo Zhang, L. W., Kaiyuan Gao, Jiangchao Yao, Tao Qin, Bo Han. (2025). FAST AND ACCURATE BLIND FLEXIBLE DOCKING. ICLR 2025. https://doi.org/https://openreview.net/pdf?id=iezDdA9oeB
19. Che, X., & Zhang, L. (2025). Blind docking methods have been inappropriately used in most network pharmacology analysis. Frontiers in Pharmacology, 16. https://doi.org/10.3389/fphar.2025.1566772
20. Roomi, M. S., Culletta, G., Longo, L., Filgueira de Azevedo, W., Perricone, U., & Tutone, M. (2025). Docking in the Dark: Insights into Protein–Protein and Protein–Ligand Blind Docking. Pharmaceuticals, 18(12). https://doi.org/10.3390/ph18121777
21. Unsal, S., Holland, B., Sardag, I., & Timucin, E. (2025). Assessing the Performance of AF2 and AF3-Implementations on Antibody-Antigen Complexes. https://doi.org/10.1101/2025.07.25.666870
22. Tang, X., Dai, H., Knight, E., Wu, F., Li, Y., Li, T., & Gerstein, M. (2024). A survey of generative AI for de novo drug design: new frontiers in molecule and protein generation. Briefings in Bioinformatics, 25(4). https://doi.org/10.1093/bib/bbae338
23. Sun, Y., Chen, L., Jing, Z., Li, Y. Y., Kim, D., Gao, J.-Y., Noroozi, R., Yi, G. Y., Tetsassi Feugmo, C. G., Klinkova, A., Sask, K., Kristiadi, A., Wang, B., Gillies, E. R., Lu, K. P., Shi, H. H., & Hu, P. (2025). Generative AI for the Design of Molecules: Advances and Challenges. Journal of Chemical Information and Modeling, 65(23), 12668-12690. https://doi.org/10.1021/acs.jcim.5c02234
24. Green, D. V. S. (2025). AI in drug design: evolution or revolution? Emerging Topics in Life Sciences, 8(2), 61-71. https://doi.org/10.1042/etls20240005
25. Ozcelik, R., Brinkmann, H., Criscuolo, E., & Grisoni, F. (2025). Generative Deep Learning for de Novo Drug Design horizontal line A Chemical Space Odyssey. J Chem Inf Model, 65(14), 7352-7372. https://doi.org/10.1021/acs.jcim.5c00641
26. Kang, S. I., Shin, J. H., Wu, B. M., & Choi, H. S. (2025). Deep Generative AI for Multi-Target Therapeutic Design: Toward Self-Improving Drug Discovery Framework. International Journal of Molecular Sciences, 26(23). https://doi.org/10.3390/ijms262311443
27. Santos, L. H. S., Ferreira, R. S., & Caffarena, E. R. (2019). Integrating Molecular Docking and Molecular Dynamics Simulations. In Docking Screens for Drug Discovery (pp. 13-34). https://doi.org/10.1007/978-1-4939-9752-7_2
28. Varghese, A., Liu, J., Patterson, T. A., & Hong, H. (2025). Integrating Molecular Dynamics, Molecular Docking, and Machine Learning for Predicting SARS-CoV-2 Papain-like Protease Binders. Molecules, 30(14). https://doi.org/10.3390/molecules30142985
29. Blanco-González, A., Cabezón, A., Seco-González, A., Conde-Torres, D., Antelo-Riveiro, P., Piñeiro, Á., & Garcia-Fandino, R. (2023). The Role of AI in Drug Discovery: Challenges, Opportunities, and Strategies. Pharmaceuticals, 16(6). https://doi.org/10.3390/ph16060891