ChatGPT

Open-source Docking Engines (DockStream-Compatible)

- **AutoDock Vina** – Widely used open-source docking; CPU-based (parallelizable) with moderate accuracy. DockStream supports Vina out of the box [1]. GPU-accelerated forks (e.g. Vina-GPU) dramatically speed up throughput (14×–65× on modern GPUs [2]), easing large virtual screens.
- **AutoDock-GPU** – NVIDIA/CUDA version of AutoDock (open-source); accelerates AD4 scoring on GPU. Not natively in DockStream, but outputs (pose + score) could be interfaced similarly.
- **Smina / QuickVina (Q-Vina 2, Q-Vina-W)** – Open-source Vina variants with faster heuristics and tuned scoring. Some (e.g. QuickVina) claim multi-fold speed gains on CPU; Vina-GPU 2.0 shows further 3–4× gains for QuickVina variants [2] while maintaining accuracy.
- **rDock** – Open-source docking (Ghent Univ.); CPU only. Supported by DockStream [1]. Generally fast but gave lower enrichment than commercial tools in benchmarks [3]. Useful as a free alternative, though older scoring.
- **GNINA** – Open-source docking with CNN-based rescoring. Uses AutoDock Vina for pose generation but re-scores with deep networks (GPU-accelerated) [4]. Improves pose ranking on some targets, at cost of heavier compute. Not in DockStream by default.
- **Uni-Dock** – New open GPU docking (DP Technology). Implements Vina/Vinardo/AD4 scoring on GPUs, achieving thousands-fold speedup vs single-CPU Vina [5]. High throughput ideal for VS; would need custom integration with DockStream.
- **Others:** PSO-Vina, UF Dock, etc. In practice, any engine that outputs poses/scores can feed a REINVENT scoring function. DockStream currently only wraps Vina/rDock [1], so others require external calls and score parsing.
- **Speed vs. Accuracy:** GPU variants (Vina-GPU, Uni-Dock) massively increase speed [2] [5]. Open tools (Vina/rDock) tend to underperform commercial programs (Glide/GOLD) in enrichment [3], but they remain valuable for integration. In summary, Vina (with GPU forks), AutoDock-GPU, Smina, GNINA, and Uni-Dock are top open choices; rDock is available but less accurate.

Online Docking Validation Platforms

- **SwissDock** – Free web server (Swiss Institute of Bioinformatics) using *Attracting Cavities* (accurate) and AutoDock Vina (fast) [6]. You upload a target (PDB or ID) and ligand (file/SMILES), and get docked poses and scores. Ideal for cross-checking your docking results.
- **DockThor-VS** – Free Brazilian server for protein–ligand docking (LNCC). Supports preprocessing (protonation, cofactors) and offers redocking, blind-docking, and virtual screening modes [7]. Can validate workflows with its GUI or API.
- **SeamDock** – An interactive web interface (Front. Mol. Biosci. 2021) that lets users dock a ligand with multiple engines (AutoDock4, Vina, Q-Vina, Smina) [8]. Good for quick tests without installing software.
- **Others:** *DockingServer* (commercial); *MolMode* and *MolModulo* (browser tools); *DOCK 3.7/6 (some web options)*; *CABS-dock/HADDOCK* (for peptides/proteins); *ReverseDock (Frontiers)*, etc. These can be used

to see if different backends reproduce your binding modes. In practice, running a few test ligands through SwissDock or DockThor provides an independent sanity check of docking quality.

Modern ML Architectures for PROTAC Linker Design

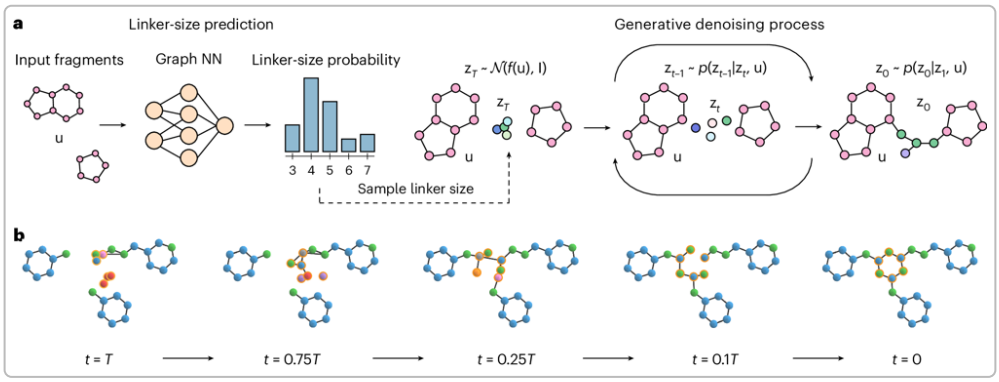

Figure: Diffusion-based linker generation (DiffLinker). (a) A GraphNN estimates the linker length probability from input fragments; (b) Denoising diffusion progressively "grows" the linker atoms (orange) between fragments [9] [10]. Generative models are advancing PROTAC linker design:

- **Diffusion Models:** 3D diffusion networks can build linkers atom-by-atom. *DiffLinker* (Igashov et al. 2024) is an equivariant 3D diffusion model that places and denoises atoms to connect fragment anchors [10]. It handles any number of fragments, predicts linker length and attachment automatically, and respects protein pocket constraints (avoiding steric clashes) [11]. Similarly, *DiffPROTACs* (Li et al. 2024) uses an equivariant graph Transformer in a diffusion framework to generate linkers conditioned on given warhead and E3 ligands [12]. These methods greatly improve 3D realism and chemical validity of linkers. Implementing diffusion in practice requires either using existing libraries (e.g. DiffDock-style frameworks) or calling models (code/servers available).
- **Graph Neural Generators:** Graph VAEs/GANs and graph Transformers can generate linker structures as molecular graphs. For example, Nori et al. (MIT, 2022) trained a graph-based generative model (with policy-gradient RL) to propose PROTAC-like molecules from scratch [13]. Their model learned PROTAC substructures and, after fine-tuning, generated >80% predicted actives with high validity. Models like *DeLinker* (Junction-Tree VAE with 3D context; Lim et al. 2020) [14] show that including 3D distance/orientation between fragments dramatically improves generation (60–200% more high-3D-similarity molecules) than 2D methods. Newer graph-transformer hybrids (e.g. "JTreeformer" [15]) and 3D-graph flows (e.g. 3D-MolT5, DiffDock) could likewise be adapted for linkers. Using graph-based decoders instead of SMILES RNNs may yield more chemically realistic linkers, especially for large PROTACs.
- **Retrieval-Augmented Generation (RAG):** Recent work introduces fragment retrieval into generation. The f-RAG model (Lee et al. 2024) retrieves "hard" fragments (e.g. two binding moieties or fragments of known linkers) and conditions a transformer to "complete" the molecule by generating the linking fragment [16]. This could be applied to PROTACs by treating the warhead and E3 ligand as hard fragments, retrieving similar linkers as soft hints, and using a generative model to produce a novel linker. Such RAG approaches blend database knowledge with neural creativity.
- **Large Language Models (LLMs) / Transformers:** SMILES/SELFIES-based transformer LMs (GPT-like) trained on chemical data (e.g. MolGen-Transformer [17], ChemBERT, etc.) can generate molecules from text-like prompts. In practice, one might fine-tune a chemical GPT on PROTACs or use in-

context learning: for example, a user could prompt ChatGPT or a similar LLM with “Design a linker between ligand A and E3 ligand B to target protein X.” Tools like ChatChemTS [18] show LLMs can even construct reward functions or interpret objectives via chat. While direct SMILES generation from GPT needs careful filtering, pairing LLM outputs with docking/ML scoring in an agent loop (GPT calls docking, gets scores, iterates) is a promising future direction.

- **Multi-Objective / Agent Architectures:** Modern pipelines could use RL or LLM-agents that integrate multiple tasks. For instance, an LLM agent could generate candidate linkers and call DockStream (or another docking engine) as a tool to evaluate binding in each reward loop. Retrieval of known linkers (PROTAC-DB), property predictors (QNICS, ADME models), and docking can all form a composite scoring function. Architecturally, one might build a hierarchical model: first enumerate warhead–E3 combinations, then use a specialized linker generator (graph/diffusion) conditioned on docking score feedback. Using transformers or diffusion models at the linker stage, rather than just RNNs, may capture complex chemistries better.

Learning Resources and Further Reading

- **Reviews/Papers:** Gharbi & Mercado (2024) provide a comprehensive review of ML approaches to PROTAC design [19]. Guo et al. (2023) introduce *Link-INVENT*, an RL-based linker design tool (REINVENT extension) [20]. Igashov et al. (2024) present *DiffLinker* (Nat. Mach. Intel.) [10]; Li et al. (2024) detail *DiffPROTACs* (Briefings Bioinfo) [12]. Lim et al. (2020) describe *DeLinker* (ACS Cent. Sci.) [14]. Lee et al. (2024) cover fragment-RAG (ArXiv) [16]. These papers give in-depth methods and benchmarks.
- **Blogs/Articles:** Harris (2023) provides an accessible Medium article on diffusion models in chemistry [21]. The *Practical Cheminformatics* blog has a layman-friendly post analyzing DiffLinker results and issues [22]. The DockStream GitHub (documentation) and SwissDock tutorial sites explain practical usage.
- **Webinars/Videos:** Look for recent talks on “AI for molecules” (e.g. on YouTube channels of major conferences). Some ML for chemistry workshops cover diffusion and transformers (e.g. NeurIPS, ML4MolecularScience). The CHEMistry AI podcast and AI-focused YouTube channels (e.g. *Artificial Chemist*) often discuss these topics.
- **Tutorials/Code:** The official REINVENT/DockStream repos (GitHub) have tutorials. The DiffLinker and DiffPROTACs GitHub repositories include examples. Uni-Dock’s repo and Vina-GPU release provide code. Transformer model repos (e.g. HuggingFace MolGen) and DiffDock codebases can be studied for architecture ideas.

Summary: In practice, we recommend starting with DockStream’s supported tools (Vina/rDock) for docking integration, possibly augmenting with Vina-GPU or AutoDock-GPU for speed. Use webservers like SwissDock to sanity-check a few poses. For modeling, explore diffusion and graph-based generative models for linkers (DiffLinker, DiffPROTACs, graph VAEs) rather than only SMILES RNNs. Consider hybrid strategies (retrieval-augmented or LLM-based) to leverage existing PROTAC data. The cited reviews and blog posts above offer deeper guidance on these modern ML approaches [19] [21] [22].

---

[1] GitHub - MolecularAI/DockStream: DockStream: A Docking Wrapper to Enhance De Novo Molecular Design
https://github.com/MolecularAI/DockStream

2 Vina-GPU 2.0：further accelerating AutoDock Vina and its derivatives with GPUs | Biological and Medicinal Chemistry | ChemRxiv | Cambridge Open Engage
https://chemrxiv.org/engage/chemrxiv/article-details/634f64c73399729e97905ba7

3 DockStream: a docking wrapper to enhance de novo molecular design | Journal of Cheminformatics | Full Text
https://jcheminf.biomedcentral.com/articles/10.1186/s13321-021-00563-7

4 GNINA 1.3: the next increment in molecular docking with deep ...
https://pmc.ncbi.nlm.nih.gov/articles/PMC11874439/

5 GitHub - dptech-corp/Uni-Dock: Uni-Dock: a GPU-accelerated molecular docking program
https://github.com/dptech-corp/Uni-Dock

6 SwissDock 2024: major enhancements for small-molecule docking with Attracting Cavities and AutoDock Vina - PubMed
https://pubmed.ncbi.nlm.nih.gov/38686803/

7 DockThor
https://dockthor.lncc.br/v2/

8 Frontiers | SeamDock: An Interactive and Collaborative Online Docking Resource to Assist Small Compound Molecular Docking
https://www.frontiersin.org/journals/molecular-biosciences/articles/10.3389/fmolb.2021.716466/full

9 10 11 Equivariant 3D-conditional diffusion model for molecular linker design | Nature Machine Intelligence
https://www.nature.com/articles/s42256-024-00815-9?error=cookies_not_supported&code=5adea210-c60b-466f-8a1a-397999f50b2e

12 (PDF) DiffPROTACs is a deep learning-based generator for proteolysis targeting chimeras
https://www.researchgate.net/publication/382882394_DiffPROTACs_is_a_deep_learning-based_generator_for_proteolysis_targeting_chimeras

13 arxiv.org
https://arxiv.org/pdf/2211.02660

14 Deep Generative Models for 3D Linker Design - PMC
https://pmc.ncbi.nlm.nih.gov/articles/PMC7189367/

15 JTreeformer: Graph-Transformer via Latent-Diffusion Model for Molecular Generation
https://arxiv.org/html/2504.20770v1

16 Molecule Generation with Fragment Retrieval Augmentation
https://arxiv.org/html/2411.12078v1

17 MolGen-Transformer: A molecule language model for the generation and latent space exploration of pi-conjugated molecules | Theoretical and Computational Chemistry | ChemRxiv | Cambridge Open Engage
https://chemrxiv.org/engage/chemrxiv/article-details/67bce95d81d2151a02e708ba

18 Large Language Models Open New Way of AI-Assisted Molecule Design for Chemists | Theoretical and Computational Chemistry | ChemRxiv | Cambridge Open Engage
https://chemrxiv.org/engage/chemrxiv/article-details/66220456418a5379b0297f8d

19 A comprehensive review of emerging approaches in machine learning for de novo PROTAC design - Digital Discovery (RSC Publishing) DOI:10.1039/D4DD00177J
https://pubs.rsc.org/en/content/articlehtml/2024/dd/d4dd00177j

20 Link-INVENT: generative linker design with reinforcement learning - Digital Discovery (RSC Publishing) DOI:10.1039/D2DD00115B
https://pubs.rsc.org/en/content/articlehtml/2023/dd/d2dd00115b

21 Diffusion Models in Generative Chemistry for Drug Design | by Charlie Harris | Medium
https://medium.com/@cch57/exploring-the-promise-of-generative-models-in-chemistry-an-introduction-to-diffusion-models-31530e9d1dcb

22 Generative Molecular Design Isn't As Easy As People Make It Look
http://practicalcheminformatics.blogspot.com/2024/05/generative-molecular-design-isnt-as.html