Proceedings of ICBioMed 2025 Symposium: AI for Healthcare: Advanced Medical Data Analytics and Smart Rehabilitation
DOI: 10.54254/2753-8818/2025.AU28436

Machine Learning-Driven De Novo Design of DDP-4 Targeted Drug Molecules Using RNN-Based Generative Models and Reinforcement Learning

Jiatong Lou

Cushing Academy, 39 School St., Ashburnham, USA
jasminelou0112@gmail.com

Abstract. We present a fully in silico pipeline for the de novo design of dipeptidyl peptidase-4 (DPP- 4) inhibitors that integrates data-driven curation, transfer learning, and reinforcement learning (RL) within the REINVENT architecture. Activity records for human DPP-4 (CHEMBL284) were programmatically retrieved from ChEMBL, normalized to nanomolar units, filtered at IC50≤100 nM, standardized to canonical SMILES, and consolidated into a high-quality training table; pIC50 values were computed and a top-100 reference set was exported for down- stream novelty control. A REINVENT prior was adapted to the DPP-4 chemical space via maximum-likelihood fine-tuning on 173 nonredundant, high-activity SMILES. The adapted generator was then optimized with an RL objective that combined predicted potency (pIC50), drug-likeness (QED), synthetic accessibility (SA), and novelty penalties relative to the top-100 reference inhibitors. Relative to the transfer-learned baseline, RL increased mean QED by ~ 10%, improved normalized synthetic accessibility (1 - SA)/10 by ~ 15%, and maintained diversity with ~60% novelty, while the composite reward showed a clear upward shift. Structure-based evaluation further corroborated these gains: 100 RL-generated molecules achieved a mean docking score of -9.8 kcal/mol, surpassing both pre-RL de novo samples (-7.7 kcal/mol) and the top 100 reference actives (-8.5 kcal/mol). These results demonstrate that RL fine-tuning can steer a pretrained generator toward DPP-4–relevant regions of chemical space with improved developability surrogates and predicted binding. Future work will integrate ADMET predictors into the reward and prioritize wet-lab validation to confirm biochemical potency and advance selected designs toward lead optimization.

Keywords: DPP-4 inhibitors, de novo drug design, reinforcement learning, generative models, computational drug discovery

1. Introduction

Type 2 diabetes mellitus (T2DM) has become a global epidemic, affecting hundreds of millions of people worldwide and rising rapidly [1]. Over 90% of diabetes cases are T2DM, creating an urgent need for effective therapeutic strategies [1]. One key target for T2DM management is Dipeptidyl Peptidase-4 (DPP-4), a serine protease (also known as CD26) that plays a central role in glucose

DOI: 10.54254/2753-8818/2025.AU28436

homeostasis by degrading incretin hormones (GLP-1 and GIP). Inhibiting DPP-4 prolongs the action of incretins, thereby enhancing insulin secretion and reducing blood glucose levels [1]. DPP-4 inhibitors, commonly called "gliptins," are an established class of oral antihyperglycemic drugs approved for T2DM treatment [2]. Several gliptins (e.g. sitagliptin, saxagliptin, linagliptin, alogliptin, vildagliptin) have been widely used as monotherapy or in combination with other agents to improve glycemic control in T2DM patients [2]. These drugs offer the convenience of once-daily oral dosing and generally have a neutral effect on body weight and a low risk of hypoglycemia compared to some older therapies [2].

Despite their clinical success, current DPP-4 inhibitors have notable limitations that motivate continued drug development. Gliptins typically achieve only moderate reductions in glycated hemoglobin (HbA1c) and have not demonstrated the robust cardiovascular benefits seen with newer drug classes (such as GLP-1 receptor agonists or SGLT2 inhibitors) in outcome trials [2]. Moreover, while DPP-4 inhibitors are well tolerated overall, they have been linked to various adverse effects [1,2]. Common side effects include mild infections (e.g. nasopharyngitis or upper respiratory tract infection) and headaches, but more serious reactions have also been reported [1,2]. Post-marketing surveillance noted rare cases of severe joint pain and immune-mediated reactions (angioedema, Stevens-Johnson syndrome) in patients on gliptins [3]. Of particular concern, one agent (saxagliptin) was associated with an increased risk of hospitalization for heart failure in a large clinical trial, prompting FDA warnings about heart failure risk for saxagliptin (and a similar signal for alogliptin) [3]. These safety and efficacy shortcomings underscore the need to identify new DPP-4 inhibitors with improved profiles [3]. Indeed, even with twelve DPP-4 inhibitors already approved, the development of novel, more potent and selective DPP-4 inhibitors remains an ongoing research priority [3]. The goal is to discover next-generation DPP-4 drugs that can deliver superior glycemic control or additional clinical benefits (such as cardio- or renoprotective effects) while minimizing adverse outcomes [4].

Advances in artificial intelligence (AI) and computational chemistry offer a promising avenue to accelerate the discovery of improved DPP-4 inhibitors. Drug discovery is often likened to finding a needle in a haystack, given the astronomically large chemical space (estimated 1060–10100 possible drug-like molecules) that must be searched [5]. Traditional trial-and-error synthesis and screening can sample only a tiny fraction of this space [5]. By contrast, AI-driven de novo design allows researchers to efficiently explore virtual chemical libraries and focus on candidates most likely to succeed [5]. In particular, deep generative models have emerged as powerful tools for in silico drug design [5]. Recurrent neural network (RNN) models trained on large collections of known compounds

can learn the "chemical language" of drug-like molecules and generate novel structures that obey learned syntax and patterns [5]. By fine-tuning such models on molecules active against a specific target, one can bias the generation toward promising chemotypes for that target [5]. For example, Olivecrona et al. (2017) showed that an RNN model fine-tuned on dopamine D2 receptor ligands could produce new molecules of which over 95% were predicted to be active, including some not present in the training set [5]. Reinforcement learning (RL) can further enhance generative design by optimizing multiple objectives jointly [6]. In an RL framework, a generative model (e.g. an RNN or variational autoencoder) is guided with feedback from a custom "reward" function that scores each generated molecule on desired properties [6]. This approach makes it feasible to incorporate complex drug-like criteria – such as predicted bioactivity, binding affinity, pharmacokinetic properties, and synthetic accessibility – into the design process. Recent studies have demonstrated that deep RL algorithms can steer molecular generation toward higher-scoring regions of chemical space, yielding

DOI: 10.54254/2753-8818/2025.AU28436

candidates that satisfy multiple pharmaceutical design objectives simultaneously [5]. Given the abundance of data available for DPP-4 (thousands of known inhibitors with activity values in databases ChEMBL [7]) and the availability of high-resolution crystal structures of DPP-4-inhibitor complexes [1], this target is especially well suited for an AI-driven discovery effort. An AI model can leverage the rich structure–activity data to learn what makes a compound a potent DPP-4 inhibitor, and use the protein structure to assess binding interactions, thereby generating novel inhibitor designs that exploit the DPP-4 active site. Computational design not only expands the search for new DPP-4 inhibitor chemotypes beyond known scaffolds, but also significantly reduces the time and cost of early-stage development by prioritizing only the most promising candidates for synthesis and testing [5].

In this work, we present a de novo drug design pipeline for DPP-4 inhibitors that integrates deep learning and molecular modeling in a no wet-lab (fully in silico) approach. We employ a generative neural network based on the REINVENT architecture [5,8] – a sequence-to-sequence model that produces novel molecular SMILES strings – to explore the chemical space of potential DPP-4 inhibitors. The model is first pre-trained on a large corpus of drug-like molecules to learn general chemical syntax, then fine-tuned with transfer learning on known DPP-4 active molecules (and inactive) to impart knowledge of the target-specific structure–activity landscape [5]. After this supervised training stage, we further optimize the generator using reinforcement learning. A multi-objective reward function is designed to drive the model toward compounds with high predicted DPP-4 inhibitory activity and favorable drug-like properties. In our reward, we incorporate both ligand-based and structure-based metrics: for instance, a QSAR predictor (e.g. an XGBoost model [9]) provides an estimated bioactivity (IC50) for DPP-4, and a docking score is computed by virtually docking the generated molecule into the DPP-4 binding pocket (using a known crystal structure such as PDB 6B1E) to evaluate its binding affinity. Additional penalty or bonus terms ensure that the generated compounds satisfy key medicinal chemistry filters – including Lipinski's drug-likeness rules, acceptable polarity (TPSA), low predicted toxicity (e.g. avoiding hERG liability), and synthetic accessibility scores – and that they are structurally novel (dissimilar to existing DPP-4 inhibitors). By integrating these components, the AI model effectively optimizes each design iteration for potency, selectivity, and developability. We iterate this RL-based generation process to produce a library of virtual DPP-4 inhibitor candidates that meet our criteria. Finally, we evaluate the top-scoring molecules from our generative model via more detailed in silico analyses (such as rescoring with more accurate binding free energy calculations and ADMET [10] property prediction) to triage the best candidates.

Overall, our study demonstrates how modern AI techniques can be harnessed to accelerate lead discovery for an important diabetes drug target without any laboratory experiments. We show that the AI-driven approach can rapidly generate novel DPP-4 inhibitor structures with predicted high potency and improved properties, highlighting specific example compounds. These virtual leads can then be prioritized for synthesis and biological testing in future work. By bridging computational modeling and medicinal chemistry, this work is intended to benefit both data scientists and drug discovery researchers, illustrating a workflow where deep learning models serve as creative in silico chemists to propose new therapeutic candidates. Through this case study on DPP-4, we aim to pave the way for broader adoption of AI-assisted, cost-effective drug design strategies in the fight against T2DM and other diseases.

DOI: 10.54254/2753-8818/2025.AU28436

2. Methods

2.1. Dataset construction and preprocessing

Small-molecule inhibitors of dipeptidylpeptidase-4 (DPP-4) were collected directly from the ChEMBL [7] database using programmatic queries. The human DPP-4 target was specified by its identifier CHEMBL284, and all activity records containing IC50 values were retrieved. To ensure consistency across entries, IC50 values were converted into nanomolar (nM) units according to their re- ported measurement units, and only records with clearly defined relations ("=", "<", or "<=") were retained. Compounds with IC50 values greater than 100 nM were discarded, leaving only the most potent inhibitors for subsequent modeling.

The molecular structures were then standardized to remove salts and normalize chemical representations. Canonical SMILES were regenerated, and molecules that could not be parsed or sanitized were excluded. To avoid redundancy, duplicate structures were collapsed by retaining the entry with the most potent IC50 value. In addition to SMILES, we encoded each compound into the SELFIES representation to facilitate downstream generative modeling.

Finally, pIC50 values were calculated from the normalized IC50 values, and the resulting dataset of high-quality DPP-4 inhibitors was saved as a CSV file for use in model training. For benchmarking and novelty assessment during reinforcement learning, the top 100 compounds ranked by pIC50 were extracted and written into a separate SMILES file. These top actives serve as a reference when evaluating the similarity of newly generated molecules.

2.2. Base model and pretraining

We adopted the REINVENT4 sequence model as the generative backbone, operating on SMILES strings. The model was run in the same environment used throughout the project and consumed the canonical SMILES produced by the preprocessing pipeline. Although SELFIES strings were also stored during curation for potential downstream use, the generative model in this work was trained and sampled exclusively with SMILES.

No bespoke pretraining was carried out within this project. Instead, we initialized the generator from the published REINVENT "prior" checkpoint, which is already pretrained on a large, general drug-like chemical corpus. This choice ensured immediate high validity and diversity in sampled structures without first training a new prior from scratch. The resulting base model served as the starting point for all target-specific learning that followed.

To verify that the base-plus-transfer workflow produced chemically well-formed outputs, we later sampled molecules from the DPP-4–adapted model and collected the valid SMILES into a single file; the corresponding log shows a total of 19,815 valid samples were produced in that run, which were subsequently screened by our postfilter.py pipeline. The DPP-4 fine-tuning set itself was prepared as a SMILES list containing 173 high-quality inhibitors extracted from the curated activity table; this file is referenced downstream when describing target-specific transfer learning.

2.3. Target-specific fine-tuning

To adapt the REINVENT prior to the chemistry of potent DPP-4 inhibitors, we performed transfer learning on a compact, activity-enriched training set assembled from the curated ChEMBL pull. From the standardized activity table, we extracted a nonredundant list of canonical SMILES for high-activity compounds (IC50 $\leq$100 nM) and saved them as a plain SMILES file for training. The

DOI: 10.54254/2753-8818/2025.AU28436

resulting file contained 173 unique structures, which served as the sole supervision signal for this stage.

Fine-tuning followed the standard REINVENT maximum-likelihood procedure on SMILES strings: starting from the public prior checkpoint, the model weights were updated by continuing next-token likelihood training on the 173-sequence corpus. No labels beyond the sequences themselves were used at this step; the goal was to nudge the generator's distribution toward scaffolds and substituent patterns characteristic of strong DPP-4 inhibitors while retaining the chemical validity and diversity afforded by the pretrained prior.

Upon completion of transfer learning, we sampled a large batch from the adapted model to verify validity and to gauge the distributional shift. The project logs and subsequent filtering run indicate that 19,815 valid SMILES were produced in this sampling pass. These samples were then process cessed by the project's post-generation screen to remove trivial or undesired structures before any reinforcement learning or docking-based comparisons. The fine-tuned generator obtained here was used as the starting policy for the downstream reinforcement learning stage.

2.4. Reinforcement learning optimization

Starting from the transfer-learned generator, we applied reinforcement learning to further bias the sampling process toward molecules that satisfy multiple design objectives simultaneously. The optimization was carried out within the REINVENT policy-gradient framework, where the generative model is treated as a stochastic policy πθ over the space of SMILES sequences.

The composite reward R(m) for a molecule m was defined as a weighted sum of four terms reflecting activity, drug-likeness, synthetic accessibility, and novelty:

$$R(m) = \alpha \bullet \hat{y}_{pIC50}(m) + \beta \bullet QED(m) - \gamma \bullet SA(m) + \delta \bullet Novelty(m)$$

where pIC50 (m) is the predicted potency from the XGBoost regression model trained on the curated

DPP-4 dataset, QED(m) is the quantitative estimate of drug-likeness, SA(m) is the synthetic accessibility score, and Novelty(m) penalizes excessive similarity to the top 100 highest-pIC50 reference molecules extracted during preprocessing. The coefficients α, β, γ, δ control the relative weighting of each design criterion. In practice, molecules similar to any entry in the top-100 list (exported as a SMILES file from the curated ChEMBL data) received lower novelty scores, ensuring that the agent generated new scaffolds rather than memorized actives.

The learning objective was to maximize the expected reward under the current policy:

$$J(\theta) = \mathbb{E}_{m\sim\pi\theta}[R(m)]$$

Following the REINVENT implementation of the REINFORCE algorithm, the model parameters θ were updated via the gradient

$$\nabla_\theta J(\theta) = \mathbb{E}_{m\sim\pi\theta}[R(m)\nabla_\theta log\pi_\theta(m)$$

DOI: 10.54254/2753-8818/2025.AU28436

This update rule shifts the policy toward regions of chemical space with higher composite reward while preserving the SMILES validity guarantees inherited from the pretrained prior.

During each RL epoch, the model sampled batches of candidate SMILES, computed their individual rewards as defined above, and applied the policy gradient update. After sampling, molecules were standardized and passed through the same filtering pipeline used in earlier stages (including PAINS, SA, and novelty checks) to exclude trivial or invalid outputs. The final RL-optimized generator produced the set of molecules that were subsequently benchmarked against both pre-RL samples and the experimentally validated high-pIC50 inhibitors using docking-based evaluation.

2.5. Docking-based evaluation

To evaluate binding potential, molecular docking was performed against the DPP-4 protein structure obtained from the Protein Data Bank (PDB). Docking simulations were carried out on three sets of molecules:

1. Molecules generated from the pretrained model (before RL).
2. Molecules generated after RL optimization.
3. Reference high-pIC50 molecules from the curated dataset.

Docking scores were compared across the three groups. Pre-RL molecules generally exhibited worse binding scores than reference inhibitors, while RL-optimized molecules achieved docking scores comparable to or exceeding those of experimentally validated inhibitors. This confirmed the effectiveness of the RL strategy in biasing the model toward chemically feasible and biologically relevant structures.

3. Results and discussion

Application of reinforcement learning to the DPP-4–adapted generator yielded substantial improvements in the overall quality of generated molecules as measured by multiple design criteria. Quantitative estimate of drug-likeness (QED) values increased on average by approximately 10% relative to the transfer-learned baseline, while synthetic accessibility, normalized as (1 - SA)/10, showed a mean improvement of 15%. Importantly, this optimization did not come at the cost of diversity: the novelty of the generated set, measured as the fraction of molecules remaining dissimilar to the top 100 high-pIC50 reference inhibitors, was maintained at roughly 60%. The composite reward scores used during training showed a clear upward trend, confirming that the agent was successfully guided by the multi-objective optimization signal.

Docking experiments further substantiated the impact of reinforcement learning. A set of 100 molecules sampled from the RL-optimized generator achieved a mean docking score of -9.8 kcal/mol against the DPP-4 binding site. This represents a marked improvement over the mean docking score of -7.7 kcal/mol obtained from an equal-sized set of de novo molecules generated before RL. Moreover, the RL-generated molecules slightly outperformed the experimentally validated high-pIC50 reference set, who's top 100 entries averaged -8.5 kcal/mol. These results indicate that reinforcement learning not only improved surrogate metrics such as QED and synthetic accessibility but also produced compounds with more favorable predicted binding energies than both the pre-RL baseline and the strongest inhibitors in the original dataset.

Taken together, these findings highlight the effectiveness of reinforcement learning in shaping generative chemical models toward multiple, practically relevant objectives. The ability to achieve improvements simultaneously in drug-likeness, synthetic feasibility, and docking-predicted binding

DOI: 10.54254/2753-8818/2025.AU28436

affinity underscores the potential of this approach for designing high-quality DPP-4 inhibitors beyond those present in current experimental databases.

4. Conclusion

In this study, we reported a reinforcement learning fine-tuning strategy for molecular generation using the REINVENT architecture. By curating a high-quality dataset of potent DPP-4 inhibitors and guiding the generative model with a composite reward, we successfully biased the sampling process toward compounds with improved drug-like properties. The RL-optimized molecules demonstrated higher QED values, better synthetic accessibility, and maintained a substantial level of novelty compared to reference actives. In addition, docking evaluation revealed that the new molecules achieved more negative binding scores than both pre-RL generated compounds and the top experimental inhibitors, suggesting relatively stronger predicted binding affinity to DPP-4.

These results highlight the potential of reinforcement learning to refine pretrained generative models toward therapeutically relevant objectives in drug design. As future work, we plan to extend the framework by incorporating explicit ADMET10 predictors into the reward design, further balancing potency with pharmacokinetic and safety considerations. Beyond computational assessments, experimental validation through wet-lab synthesis and biochemical assays will be essential to con- firm the practical viability of the generated candidates and to advance them toward the stage of lead optimization.

References

[1] Istrate, D.; Crisan, L. Dipeptidyl peptidase 4 inhibitors in type 2 diabetes mellitus management: Pharmacophore virtual screening, molecular docking, pharmacokinetic evaluations, and conceptual DfT analysis. Processes 2023, 11, 3100.
[2] Green, B. D.; Flatt, P. R.; Bailey, C. J. Dipeptidyl peptidase IV (DPP IV) inhibitors: a newly emerging drug class for the treatment of type 2 diabetes. Diabetes and vascular disease re- search 2006, 3, 159–165.
[3] Petrov, V.; Aleksandrova, T.; Pashev, A. Synthetic Approaches to Novel DPP-IV Inhibitors— A Literature Review. Molecules 2025, 30, 1043.
[4] Hossain, D.; Saghapour, E.; Chen, J. Y. NeSyDPP4-QSAR: Discovering DPP-4 Inhibitors for Diabetes Treatment with a Neuro-symbolic AI Approach. Frontiers in Bioinformatics 2025, 5, 1603133.
[5] Olivecrona, M.; Blaschke, T.; Engkvist, O.; Chen, H. Molecular de-novo design through deep reinforcement learning. Journal ofcheminformatics 2017, 9, 48.
[6] Popova, M.; Isayev, O.; Tropsha, A. Deep reinforcement learning for de novo drug design. Science advances 2018, 4, eaap7885.
[7] Gaulton, A.; Bellis, L. J.; Bento, A. P.; Chambers, J.; Davies, M.; Hersey, A.; Light, Y.; McGlinchey, S.; Michalovich, D.; Al-Lazikani, B.; Overington, J. P. ChEMBL: a large-scale bioactivity database for drug discovery. Nucleic acids research 2012, 40, D1100–D1107.
[8] Loeffler, H. H.; He, J.; Tibo, A.; Janet, J. P.; Voronov, A.; Mervin, L. H.; Engkvist, O. Reinvent 4: modern AI–driven generative molecule design. Journal of Cheminformatics 2024, 16, 20.
[9] Chen, T.; Guestrin, C. Xgboost: A scalable tree boosting system. Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining. 2016; pp 785–794.
[10] Van De Waterbeemd, H.; Gifford, E. ADMET in silico modelling: towards prediction paradise? Nature reviews Drug discovery 2003, 2, 192-204.