# De Novo Ligand Design for Orphan GPCRs with *DrugFlow*

**Myra Haider**[1] **William Lu**[1]
[1]Carnegie Mellon University, Ray and Stephanie Lane Computational Biology Department
{myrah, wtlu}@cs.cmu.edu
myrahaider/GPCR_LigandGen

## Abstract

G protein–coupled receptors (GPCRs) are a family of membrane proteins that are commonly used as therapeutic targets for drug development. However, many GPCRs remain orphan receptors, in which endogenous ligands have not yet been identified. In this work, we employ the generative model DrugFlow to design novel small molecules predicted to bind to orphan GPCR pockets. By conditioning a generative model on receptor pocket structures, DrugFlow can generate drug-like ligands de novo. The generated ligands are then evaluated for structural plausibility and binding affinity using PoseBusters, Autodock Vina, and other cheminformatic metrics. This approach aims to elucidate the ligand structures that can bind to GPCRs, providing new candidates for future experimental validation.

## 1. Introduction

G protein–coupled receptors (GPCRs) are one of the most important classes of membrane proteins in human biology, responsible for transmitting signals from hormones, neurotransmitters, and environmental stimuli into cellular responses. Because of their central role in physiology, GPCRs are targeted by roughly one-third of all approved drugs [3]. However, a large portion of GPCRs are still considered orphans, in which no endogenous receptor ligands has been confirmed [11]. This poses a major gap in our understanding of cell signaling and complicates drug discovery efforts.

Efforts to identify these ligands have previously been limited by lack of structural information and the complexity of receptor conformational dynamics. Traditional experimental screening methods are costly and often fail to capture the full range of possible ligand–receptor interactions. These limitations highlight the need for generative approaches capable of exploring broader chemical search space and predicting potential ligands directly from protein structural features.

To address this, we propose leveraging DrugFlow [21], a diffusion-based generative model to design novel ligands for orphan GPCRs. Diffusion models can learn the underlying distribution of known molecular structures and generate chemically valid compounds conditioned on receptor pocket information. By integrating structural features of the receptor with generative modeling, DrugFlow enables targeted de novo ligand generation, allowing us to identify novel drug candidates for experimental validation.

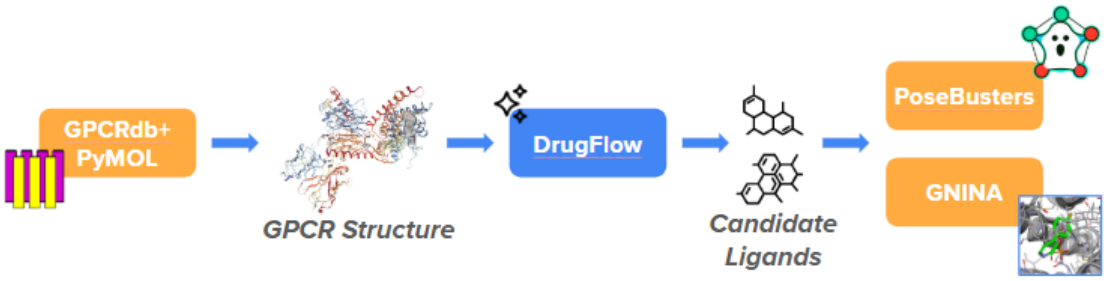


Figure 1 | Outline of our computational ligand generation and evaluation procedure.

## 2. Related Work

Early ligand-discovery methods relied on classical structure-based virtual screening tools such as AutoDock Vina, which depend heavily on existing chemical libraries and cannot explore novel structures. Generative deep-learning models, including RNN-based SMILES generators, GraphVAE, and MolGAN introduced de novo design but often produced invalid molecules and lacked reliable control over 3D geometry. More recent structure-aware approaches such as Pocket2Mol and DiffDock, generated more accurate ligands by conditioning on receptor structure, yet these models still struggle with generating diverse molecules that remain compatible with a specific binding pocket.

DrugFlow advances beyond these limitations by using a diffusion-based 3D generative framework that integrates protein-pocket information and a reference ligand pose, enabling pocket-specific ligand proposals that outperform many prior methods in structural fidelity and target specificity. This is especially valuable for orphan GPCRs where endogenous ligands are unknown, as pocket-conditioned de novo ligand generation can be more informative than traditional screening or purely ligand-based generative models.

## 3. Methodology

### 3.1. Orphan GPCR Selection

Our overall method pipeline is outlined in Figure 1. To select a reference protein for ligand generation, we used the **GPCRdb** Structure database [10] to filter for high-quality human orphan GPCR structures with a co-crystallized ligand, which can be used to define a pocket for conditional ligand generation in DrugFlow[1]. Our filtering criteria led us to select protein MRGPRX4, shown in Figure 2. MRGPRX4 is a Class A rhodopsin-like orphan receptor expressed in sensory neurons involved in chronic itch symptoms in liver disease patients[27], and shares structural similarity to other Class A GPCR orphans with proposed roles in diverse conditions ranging from autoimmune disorders [28], neurodegeneration [4], cancer, and metabolic disease [2].

[1]Per International Union of Basic and Clinical Pharmacology (IUPHAR) standards, a receptor can be considered "deorphanized" after two independent studies report receptor activity after binding of a putative endogenous ligand[11]; a GPCR structure with a bound candidate endogenous ligand or synthetic ligand can be submitted to GPCRdb even as the receptor remains an orphan.

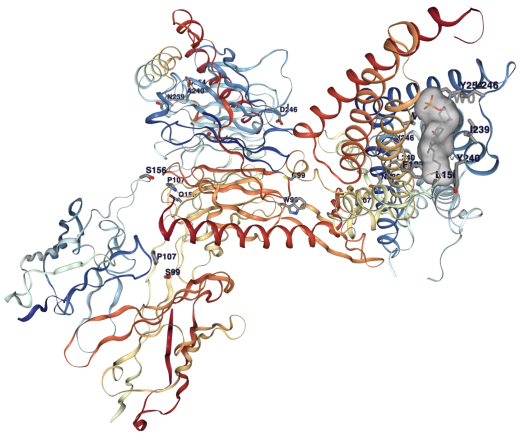

Figure 2 | 3D Structure of G-protein coupled receptor MRGPRX4. The co-crystallized reference ligand DCA-3P is shown in gray.

### 3.2. Ligand Generation

The MRGPRX4 *.pdb* file was downloaded from GPCRdb (**source**). The specific structure file used originated from a recent structural study that produced a solved cryo-EM structure of the MRGPRX4 receptor in complex with a synthetic agonist, DCA-3P, which the authors observed inserting into MRGPRX4's orthostatic site[26]; this pocket corresponds to the canonical central extracellular binding pocket formed by the receptor alpha-helix tranemembrane domains typical of class A GPCRs[12]. The protein pocket structure, consisting of the amino acids surrounding the co-crystallized ligand up to a radius of 6 Å was extracted using pyMOL. The DCA-3P reference ligand and extracted protein pocket were provided as inputs to the DrugFlow model, where we generated 64 novel ligand candidates to assess structural plausibility and binding affinity to the original protein pocket.

### 3.3. Ligand Filtering and GNINA Docking

The 64 generated ligands were initially filtered using evaluation heuristics built-in to DrugFlow, to select for drug-like, valid, and novel generated ligands. The specific metrics are listed and explained in Table 6.

We then used GNINA, a deep-learning enhanced molecular docking tool to evaluate how well each DrugFlow-generated ligand fits into the GPCR binding pocket [15]. For each ligand, GNINA performed an automated docking simulation and output both traditional physics-based scores and CNN-based scores to predict binding affinity. The intuition behind using a CNN is that it can capture non-linear relationships within the protein–ligand interaction, enabling more robust binding predictions.

Each GNINA simulation generates three evaluation metrics:

**Vina Affinity** The standard scoring function used by previous docking tools such as Autodock

Vina, estimating free binding energy. More negative values indicate stronger predicted binding.

**CNN Pose Score** A score based on a pre-trained CNN model predicting the quality of the docked pose. Scores range from 0-1, where 1 indicates high-quality poses with high geometric accuracy.

**CNN Affinity** A raw binding affinity score computed using the CNN-based model. More positive values indicate stronger predicted binding.

After running all simulations, we then generated a table of summary metrics across ligands and identified the candidates with highest Vina and CNN Affinity scores for further analysis.

### 3.4. PoseBusters Validity Checks

PoseBusters is a suite of tools that is widely used to perform a series of quality control checks on ligands produced using generative models such as DrugFlow[6]. Each generated ligand was provided as input to the posebusters.PoseBusters.bust method without a true ligand or the conditioning protein pocket as inputs for evaluation of the validity of each generated ligand in a ligand-only mode. We considered a ligand to have passed ligand-only PoseBusters checks if it satisfied a set of criteria for chemical validity/consistency (e.g. sanitization, connectedness) and intramolecular validity (e.g. bond lengths/angles, steric clashes); the specific listing of criteria used is provided in Table 7. The energy ratio, defined the ratio of the energy of the input molecule compared to the energy of 50 alternative conformations of the input molecule [6], was also reported for each generated ligand.

### 3.5. Tanimoto Similarity Checks

Tanimoto similarity, known in other contexts as Jaccard similarity, is a widely used metric to assess the similarity of two molecules in bit-vector representation [16]. We converted the reference ligand DCA-3P and the generated ligands into Morgan fingerprint representations [18] and computed Tanimoto similarity from the Morgan fingerprint representations using functionalities in the RDKit package.

### 3.6. Drug Suitability Checks

The DrugFlow built-in evaluation metrics also included chemi-informatic metrics, which we used to assess the suitability of the ligands. Specifically, these included synthetic accessibility score (SAscore)[8], lipophilicity measurement (LogP)[20], and aggregate metrics describing whether the compound meets drug-like properties including appropriate H-bond donor/acceptor count, molecular weight, logP etc. as a continuous score (QED)[5] and as a count of rules passed (Lipinski's rule of 5)[14].

# 4. Experiments

## 4.1. DrugFlow and Initial Filtering

To generate a set of candidate ligands for our receptor of interest MRGPRX4, DrugFlow was run in inference mode using the DCA-3P reference ligand and PyMOL-extracted protein pocket as inputs; this yielded 64 initial candidate ligands. To identify a subset of candidates most suitable for further evaluation, we used evaluation heuristics built-in to DrugFlow to exclude candidate ligands that did not meet stringent criteria for drug-likeness, validity, and novelty within the inference run; this resulted in 15 docking candidate ligands, depicted in Figure 3, for further study.

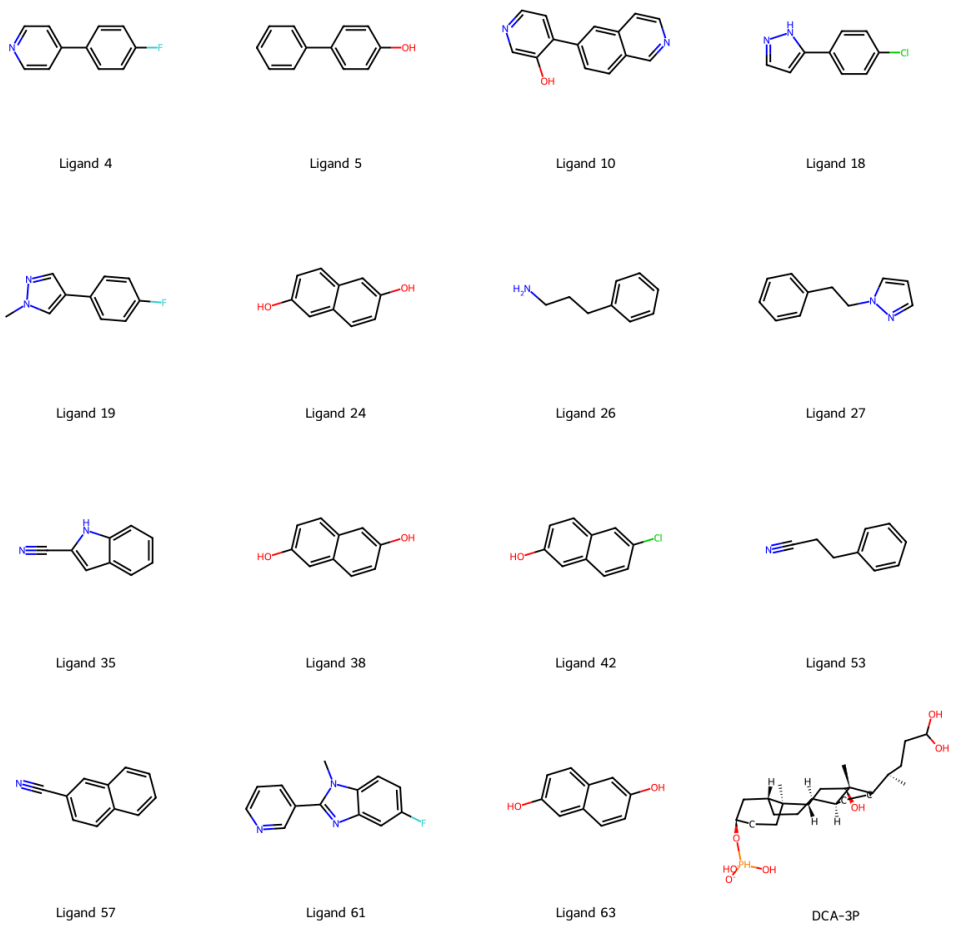


Figure 3 | 2D chemical structures of generated candidate ligands that passed DrugFlow filtering and reference ligand DCA-3P.

## 4.2. Molecular Docking with GNINA

To assess the binding plausibility of the generated ligands, we incorporate GNINA, a deep-learning–enhanced docking framework that combines traditional scoring functions with 3D convo-

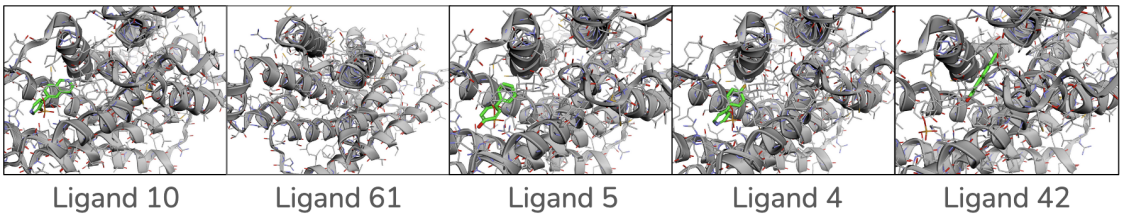


Figure 4 | Docking simulation results for the five ligands with highest Vina Affinity score. Note that Ligand 61 was not able to be visualized with Py3Dmol as its structure was determined to be chemically invalid.

lutional neural networks. [15] For each generated molecule, we performed ligand docking into the GPCR binding pocket and then measured Vina Affinity Score, CNN Pose Score, and CNN Affinity score to evaluate binding likelihood.

The highest-scoring candidate ligands differed depending on whether Vina Affinity or CNN Affinity was used as the scoring metric. This is because the scores are computed using two different methodologies. Because GNINA's CNN-based scoring model was designed to capture more complex ligand-receptor interactions, it's possible that the traditional Vina Score is not as robust. On the other hand, if the CNN model was trained on protein structures that are dissimilar to the novel ligands generated by DrugFlow, CNN Affinity may not be the most accurate predictor of binding affinity. Further exploration is needed to determine which is the case. However, both metrics successfully identified 8 of the same ligands within the top 10 scoring candidates, indicating that binding predictions for both methods are fairly consistent (Tables 1 and 2). Docking simulations were visualized using Py3DMol for the highest scoring candidates, as shown in Figure 4.

The docking scores of the top 15 docking candidates are very similar to the full set of 64 ligands. The mean Vina affinity slightly increases from –4.98 to –5.08 kcal/mol, indicating slightly stronger predicted binding among the selected candidates. The CNN pose score also increases marginally from 0.64 to 0.65. Interestingly, the CNN affinity score decreases for the top 15 ligands compared to the entire dataset (4.10 $\rightarrow$ 3.7) (Table 3).

Table 1 | Ligand candidates ranked by Vina Affinity.

| Ligand ID | Vina Affinity | CNN Pose Score | CNN Affinity |
|---|---|---|---|
| 10 | **-6.31** | 0.6009 | 4.667 |
| 61 | -6.23 | 0.5817 | 4.801 |
| 5 | -6.16 | 0.6823 | 3.880 |
| 4 | -5.92 | 0.6469 | 4.140 |
| 42 | -5.66 | 0.7031 | 3.359 |
| 27 | -5.56 | 0.6739 | 3.449 |
| 19 | -5.38 | 0.6457 | 4.175 |
| 38 | -5.35 | 0.6240 | 3.698 |
| 63 | -5.23 | 0.6357 | 3.785 |
| 24 | -5.17 | 0.6467 | 3.765 |

Table 2 | Top 10 filtered ligand candidates ranked by CNN Affinity.

| Ligand ID | Vina Affinity | CNN Pose Score | CNN Affinity |
|---|---|---|---|
| 61 | -6.23 | 0.5817 | **4.801** |
| 10 | -6.31 | 0.6009 | 4.667 |
| 19 | -5.38 | 0.6457 | 4.175 |
| 4 | -5.92 | 0.6469 | 4.140 |
| 5 | -6.16 | 0.6823 | 3.880 |
| 57 | -5.13 | 0.6713 | 3.872 |
| 63 | -5.23 | 0.6357 | 3.785 |
| 24 | -5.17 | 0.6467 | 3.765 |
| 38 | -5.35 | 0.6240 | 3.698 |
| 35 | -4.75 | 0.6083 | 3.690 |

Table 3 | Average GNINA metrics for all generated ligands compared to the top 15 docking candidates

| Ligand Set | Mean Vina Affinity | Mean CNN Pose Score | Mean CNN Affinity |
|---|---|---|---|
| All Ligands (n = 64) | -4.98 | 0.64 | **4.10** |
| Top 15 Docking Candidates | **-5.08** | **0.65** | 3.79 |

### 4.3. PoseBusters Validity Evaluation

To rigorously assess the molecular plausibility of the candidate ligands, the DrugFlow pipeline incorporates Posebusters, a suite of tools that would perform a series of quality control checks on our ligands to evaluate the realism of the structure of each ligand molecule [6]. Overall, 49 out of 64 total generated candidate ligands (not shown) and 14 out of 15 docking candidate ligands passed our PoseBusters checks, with only Ligand 61 passing the DrugFlow built-in filtering but failing at least one PoseBusters check. The PoseBusters energy ratio was less than 8 for all but two docking candidate ligands, one of which was Ligand 61; no docking candidate had an energy ratio exceeding 100 (Table 4).

### 4.4. Tanimoto Similarity

To quantitatively assess whether DrugFlow generated small-molecule ligands that were substantially different from the reference ligand DCA-3P provided as input, we computed the Tanimoto similarity between each of the 15 docking candidate ligands with the reference ligand (Table 4). No docking candidate ligand had a Tanimoto similarity score with the reference ligand exceeding 0.03, and 3 of these ligands had a Tanimoto similarity of 0.

### 4.5. Drug-Suitability Evaluation

To assess whether the DrugFlow-generated ligands were suitable as drug candidates, we used evaluation metrics built into DrugFlow that evaluated the drug-like properties of the 15 docking candidate ligands (Table 5). On average, the docking candidates had a favorable synthesizability

| ID | representation.smiles | passed posebusters | energy_ratio | tanimoto_sim_to_ref |
|---|---|---|---|---|
| 4 | Fc1ccc(-c2ccncc2)cc1 | True | 1.563006 | 0.000000 |
| 5 | Oc1ccc(-c2ccccc2)cc1 | True | 1.440248 | 0.014493 |
| 10 | Oc1cnccc1-c1ccc2cnccc2c1 | True | 2.147329 | 0.012195 |
| 18 | Clc1ccc(-c2ccn[nH]2)cc1 | True | 1.313527 | 0.000000 |
| 19 | Cn1cc(-c2ccc(F)cc2)cn1 | True | 1.268150 | 0.025974 |
| 24 | Oc1ccc2cc(O)ccc2c1 | True | 1.726234 | 0.014925 |
| 26 | NCCCc1ccccc1 | True | 3.799574 | 0.013889 |
| 27 | c1ccc(CCn2cccn2)cc1 | True | 7.315127 | 0.012658 |
| 35 | N#Cc1cc2ccccc2[nH]1 | True | 1.744734 | 0.012987 |
| 38 | Oc1ccc2cc(O)ccc2c1 | True | 1.466310 | 0.014925 |
| 42 | Oc1ccc2cc(Cl)ccc2c1 | True | 1.777453 | 0.013699 |
| 53 | N#CCCc1ccccc1 | True | 22.777652 | 0.013699 |
| 57 | N#Cc1ccc2ccccc2c1 | True | 1.176016 | 0.000000 |
| 61 | Cn1c(-c2cccnc2)nc2cc(F)ccc21 | False | 25.555770 | 0.011111 |
| 63 | Oc1ccc2cc(O)ccc2c1 | True | 1.408610 | 0.014925 |

Table 4 | Posebusters validity, energy ratio, and Tanimoto similarity to reference ligand for generated ligands

score and balanced lipophilicity. In terms of aggregated scores, all docking candidates passed each Lipinski rule and had a mean QED score of 0.646.

| Metric | Value |
|---|---|
| Mean medchem.qed | $0.646 \pm 0.0411$ |
| Mean medchem.sa | $1.74 \pm 0.3$ |
| Mean logP | $2.48 \pm 0.46$ |
| % Lipinski 5/5 | 100.0% |
| % Lipinski $\geq$4/5 | 100.0% |

Table 5 | Summary of drug-likeness metrics for docking candidate ligands

## 5. Analysis and Limitations

In this project, we leveraged DrugFlow, a diffusion-based generative model, to design de-novo small molecule ligands against an orphan GPCR. We established a complete pipeline for this procedure that included pocket extraction from the protein structure file to prepare inputs to DrugFlow, and downstream evaluation of the ligands' receptor docking, molecular plausibility, and drug-likeness.

### 5.1. Evaluating Chemical Validity and Novelty of DrugFlow-Generated Ligands

We found that a majority of the generated ligands passed the PoseBusters chemical and molecular plausibility checks, and that each of the docking candidates had a low energy ratio, suggesting that the filtered ligand candidates generated by DrugFlow are likely to be chemically valid small-molecule compounds. In particular, almost all of the filtered generated ligands passed our PoseBusters checks. Interestingly, far fewer of the initial generated candidate ligands passed DrugFlow's built-in validity evaluation metrics than the PoseBusters checks (15 vs. 49), despite using fewer checks overall that operate somewhat heuristically. One explanation for this could be that DrugFlow requires generated ligands to pass REOS filters for all available rule sets, which may be overly rigorous; it has been noted within the chemi-informatics and drug discovery community that some REOS rule sets may be overzealous or arbitrary, and conversely that some FDA-approved drugs contain elements that violate certain REOS filters [24]. Future versions of our pipeline might involve more careful consideration of exactly which REOS rule set or sets to filter on, or potentially to first filter using PoseBusters checks before evaluating the filtered generated ligands for passing REOS. Future work might also involve performing PoseBusters checks with the conditioning pocket, to detect and filter generated ligands for intermolecular clashes.

We found that each of the 15 docking candidate ligands had very low Tanimoto similarity with the reference ligand DCA-3P. This is a somewhat expected result, since DrugFlow is a model designed for generation of small-molecule ligands, while DCA-3P is a lipid molecule agonist of MRGPRX4 [26]. However, this does raise the question of whether our docking candidate ligands, despite showing strong affinity to the binding pocket and favorable pose, actually interact with the pocket in a similar way as DCA-3P and would be expected to produce comparable agonist activity for MRGPRX4. On the other hand, the low similarity between generated ligands and the reference ligand does demonstrate that DrugFlow truly does perform de-novo generation of ligands against our receptor of interest, and does not simply copy the reference ligand.

### 5.2. Drug-Likeness and Synthesizability of Generated Ligands

Finally, we generally found that the docking candidate ligands produced by DrugFlow were suitable as drug targets. Strikingly, the generated ligands have a low SAscore on average, suggesting that it would be quite straightforward to actually chemically synthesize and manufacture them. The generated ligands also exhibit an average LogP measurement that indicates a balance between membrane permeability (lipophilicity) and absorbability (hydrophilicity), a key quality for drug candidates. Interestingly, while there is a growing opinion that Lipinski's rules for drug suitability may be out-of-date and exclude valid drug candidates [22], all of our 15 docking candidate molecules completely satisfied Lipinski's rules of 5. On the other hand, when considering our generated ligands' mean QED of 0.646, our docking candidates are likely suitable drug candiates on balance but fall short of perfect favorability (QED = 1) [5] suggested by the unanimity by which they passed Lipinski's rules. This is likely because the QED score incorporates rules in addition to those used in Lipinski's 5, such as polar surface area, rotatable bond count, and aromatic ring count [5], and that our generated ligands had violations in these additional areas. An enhanced evaluation suite for drug-suitability could specifically measure the metrics incorporated in QED score and Lipinski's rules to identify specifically which rules for drug-likeness are not met.

### 5.3. Discrepancies Between Docking Predictions and Chemical Validity

A key limitation that we observed with in-silico pipelines for ligand generation is that high-scoring computational predictions do not always correspond to chemically valid molecules. Generative models can produce structures that numerically optimize a model's objective function, such as docking affinity, while still violating fundamental chemical rules. As a result, subsequent computational steps may assign strong performance metrics to molecules that would be impossible to synthesize due to structural instability.

We discovered this issue when Ligand 61 emerged as a top-scoring candidate according to GNINA's docking metrics. It achieved the highest CNN Affinity score (4.801) and the second-best Vina affinity score (–6.23 kcal/mol) across all evaluated ligands, suggesting strong predicted binding to the GPCR pocket. However, downstream physical validation revealed substantial issues with this design. PoseBusters analysis yielded an extremely high energy ratio of 25.555, indicating that the ligand is chemically implausible in its generated form. Consistent with this, the structure could not be rendered in Py3Dmol, further suggesting severe structural irregularities, such as broken bonds or invalid stereochemistry, introduced by the DrugFlow generative model.

This discrepancy highlights another known limitation of docking-based scoring: GNINA evaluates interaction patterns within the binding site assuming the ligand geometry is valid, and its CNN-based scoring procedure can overestimate binding affinity when presented with geometrically unrealistic structures. In other words, GNINA "liked" how the malformed ligand fit into the pocket, but lacked the chemical validity checks necessary to penalize impossible conformations. This result emphasizes the importance of combining generative modeling with rigorous physical plausibility checks such as PoseBusters to avoid false positives arising from nonviable molecular structures.

## 6. Future Work

While the natural next step to evaluate the candidate ligands generated by DrugFlow might be to synthesize the compounds and test their activity and function in cell-based assays or in animal models, a number of other in-silico analyses could also be incorporated to strengthen our computational small-molecule ligand generation pipeline. For instance, our current pocket extraction approach relies on the availability of a protein structure co-crystallized with a reference ligand, and would not be successful for receptors for which truly no ligand is known or studied structurally. Future additions to our protein structure preprocessing pipeline could incorporate tools to identify protein pockets from structure alone, either using more traditional physics/geometry based tools like Fpocket [13] or newer machine learning-based methods like DeepPocket [1].

Although our generated ligands all had favorable synthesizability scores, the fact remains that truly novel small-molecules still need chemical synthesis protocols to be established, meaning that traditional screening methods do retain the advantage of identifying existing small molecules that can be repurposed as candidate ligands. Our pipeline could incorporate as a baseline a virtual ligand screening tool such as Pharmit [23] to screen chemical compound databases for existing candidate ligands, which can be evaluated for docking quality with GNINA, chemical plausibility with PoseBusters, and drug-suitability with RDKit to assess whether the existing compounds outperform the generated candidate ligands. In the case that existing ligands do outperform generated ligands, this would either be grounds to direct DrugFlow to more critical use cases or to

motivate improvements in DrugFlow architecture or how inference is run.

Additionally, although our current pipeline incorporates evaluation of docking/pose quality using GNINA, it lacks the capacity to evaluate or predict the functional effect of our generated ligands on the receptor of interest. GPCRs exhibit well-characterized conformational changes in response to ligand binding, from which we can interpret how the receptor's mode of signaling might change [12]. Future versions of our pipeline could incorporate a molecular dynamics simulation step using tools like OpenMM [7], which would give us the capacity to both simulate the temporal stability of ligand binding to the receptor of interest as well as observe conformation changes in the receptor over the timecourse of the simulation to predict whether the ligand might activate, inhibit, or have some more nuanced effect on receptor signaling.

Finally, a key consideration for drug-suitability of our generated candidate ligands is that these ligands should have specific activity against the receptor of interest for which it was generated, and avoid off-target and toxic effects. Future work on this pipeline could involve incorporating tools such as eToxPred [19] that can predict whether generated compounds might have unexpected toxic effects in certain organs or carcinogenicity, among others, to test in-silico whether we expect the generated compounds to be safely tolerated.

## 7. Conclusion

In this work, we developed an end-to-end computational pipeline for de novo ligand generation against an orphan GPCR using DrugFlow, coupled with downstream evaluation through PoseBusters, GNINA docking, and RDKit-based drug-likeness profiling. Our results demonstrate that diffusion-based generative models can indeed produce chemically plausible candidate ligands, with the majority of DrugFlow outputs passing PoseBusters plausibility checks and satisfying key drug-likeness criteria.

At the same time, our analysis also highlights important limitations of current in-silico generation and evaluation workflows. In particular, we observed that docking-based scoring alone can overestimate binding quality for structurally invalid ligands, whose geometric inconsistencies contrasted sharply with strong GNINA scores. This underscores the necessity of integrating physical plausibility filters and chemical validity checks alongside docking predictions to avoid misleading false positives.

Looking forward, enhancing this pipeline with additional screening methods and model fine-tuning will further strengthen its utility for real-world drug discovery. Our work demonstrates both the promise and the current challenges of generative AI for ligand design, and provides a clear foundation for more reliable computational drug development pipelines in the future.

## 8. Contributions

Myra selected and preprocessed the GPCR pocket structure for ligand generation, implemented DrugFlow to generate candidate ligands, performed docking simulations with GNINA, performed analysis of evaluation metrics, generated 3D visualizations of docked ligands, and helped conduct literature review.

William performed 2D visualization of ligands, implemented PoseBusters for ligand plausibility checks, analyzed RDKit-based drug-likeness metrics, performed Tanimoto similarity analysis, and helped conduct GPCR and cheminformatic literature review.

## References


[1] Rishal Aggarwal et al. "DeepPocket: Ligand Binding Site Detection and Segmentation using 3D Convolutional Neural Networks". In: *Journal of Chemical Information and Modeling* 62.21 (2022). PMID: 34374539, pp. 5069–5079.

[2] Raise Ahmad, Stefanie Wojciech, and Ralf Jockers. "Hunting for the function of orphan GPCRs – beyond the search for the endogenous ligand". In: *British Journal of Pharmacology* 172.13 (2015), pp. 3212–3228.

[3] Peterson SM;Hutchings CJ;Hu CF;Mathur M;Salameh JW;Axelrod F;Sato AK; *Discovery and design of G protein-coupled receptor targeting antibodies*.

[4] Devrim Öz Arslan, Melis Yavuz, and Beki Kan. "Exploring orphan GPCRs in neurodegenerative diseases". In: *Frontiers in Pharmacology* Volume 15 - 2024 (2024). ISSN: 1663-9812.

[5] G. Richard Bickerton et al. "Quantifying the chemical beauty of drugs". In: *Nature Chemistry* 4.2 (Feb. 2012), pp. 90–98. ISSN: 1755-4349.

[6] Martin Buttenschoen, Garrett M. Morris, and Charlotte M. Deane. "PoseBusters: AI-based docking methods fail to generate physically valid poses or generalise to novel sequences". In: *Chem. Sci.* 15 (9 2024), pp. 3130–3139.

[7] Peter Eastman et al. "OpenMM 8: Molecular Dynamics Simulation with Machine Learning Potentials". In: *The Journal of Physical Chemistry B* 128.1 (2024). PMID: 38154096, pp. 109–116.

[8] Peter Ertl and Ansgar Schuffenhauer. "Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions". In: *Journal of Cheminformatics* 1.1 (June 2009), p. 8. ISSN: 1758-2946.

[9] Stephen R. Heller et al. "InChI, the IUPAC International Chemical Identifier". In: *Journal of Cheminformatics* 7.1 (May 2015), p. 23. ISSN: 1758-2946.

[10] Luis P Taracena Herrera et al. "GPCRdb in 2025: adding odorant receptors, data mapper, structure similarity search and models of physiological ligand complexes". In: *Nucleic Acids Research* 53.D1 (Nov. 2024). _eprint: https://academic.oup.com/nar/article-pdf/53/D1/D425/60743681/gkae1 pp. D425–D435. ISSN: 1362-4962.

[11] Amie Jobe and Ranjit Vijayan. "Orphan G protein-coupled receptors: the ongoing search for a home". In: *Frontiers in Pharmacology* Volume 15 - 2024 (2024). ISSN: 1663-9812.

[12] Naomi R. Latorraca, A. J. Venkatakrishnan, and Ron O. Dror. "GPCR Dynamics: Structures in Motion". In: *Chemical Reviews* 117.1 (2017). PMID: 27622975, pp. 139–155.

[13] Vincent Le Guilloux, Peter Schmidtke, and Pierre Tuffery. "Fpocket: An open source platform for ligand pocket detection". In: *BMC Bioinformatics* 10.1 (June 2009), p. 168. ISSN: 1471-2105.

[14] Christopher A Lipinski et al. "Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings1PII of original article: S0169-409X(96)00423-1. The article was originally published in Advanced Drug Delivery Reviews 23 (1997) 3–25.1". In: *Advanced Drug Delivery Reviews* 46.1 (2001). Special issue dedicated to Dr. Eric Tomlinson, Advanced Drug Delivery Reviews, A Selection of the Most Highly Cited Articles, 1991-1998, pp. 3–26. ISSN: 0169-409X.

[15] Andrew T. McNutt et al. "Gnina 1.3: The next increment in molecular docking with deep learning". In: *Journal of Cheminformatics* 17.1 (2025).

[16] C.L. Mellor et al. "Molecular fingerprint-derived similarity measures for toxicological read-across: Recommendations for optimal use". In: *Regulatory Toxicology and Pharmacology* 101 (2019), pp. 121–134. ISSN: 0273-2300.

[17] David Mendez et al. "ChEMBL: towards direct deposition of bioassay data". In: *Nucleic Acids Research* 47.D1 (Nov. 2018). _eprint: https://academic.oup.com/nar/article-pdf/47/D1/D930/27437436/gky pp. D930–D940. ISSN: 0305-1048.

[18] H. L. Morgan. "The Generation of a Unique Machine Description for Chemical Structures-A Technique Developed at Chemical Abstracts Service." In: *Journal of Chemical Documentation* 5.2 (1965), pp. 107–113.

[19] Limeng Pu et al. "eToxPred: a machine learning-based approach to estimate the toxicity of drug candidates". In: *BMC Pharmacology and Toxicology* 20.1 (Jan. 2019), p. 2. ISSN: 2050-6511.

[20] Dipankar Roy and Chandan Patel. "Revisiting the Use of Quantum Chemical Calculations in LogPoctanol-water Prediction". In: *Molecules* 28.2 (2023). ISSN: 1420-3049.

[21] Arne Schneuing et al. "Multi-domain distribution learning for de novo drug design". In: *arXiv preprint arXiv:2508.17815* (2025).

[22] Michael D. Shultz. "Two Decades under the Influence of the Rule of Five and the Changing Properties of Approved Oral Drugs". In: *Journal of Medicinal Chemistry* 62.4 (2019). PMID: 30212196, pp. 1701–1714.

[23] Jocelyn Sunseri and David Ryan Koes. "Pharmit: interactive exploration of chemical space". In: *Nucleic Acids Research* 44.W1 (Apr. 2016). _eprint: https://academic.oup.com/nar/article-pdf/44/W1/W442/7631816/gkw287.pdf, W442–W448. ISSN: 0305-1048.

[24] Pat Walters. *Filtering Chemical Libraries*. https://practicalcheminformatics.blogspot.com/2018/08/filtering-chemical-libraries.html. Blog post. Aug. 2018.

[25] W. Patrick Walters and Mark Namchuk. "Designing screens: how to make your hits a hit". In: *Nature Reviews Drug Discovery* 2.4 (Apr. 2003), pp. 259–266. ISSN: 1474-1784.

[26] Jun Yang et al. "Structure-guided discovery of bile acid derivatives for treating liver diseases without causing itch". In: *Cell* 187.25 (2024), 7164–7182.e18. ISSN: 0092-8674.

[27] Huasheng Yu et al. "MRGPRX4 is a bile acid receptor for human cholestatic itch". In: *eLife* 8 (2019). Ed. by David D Ginty, Gary L Westbrook, and Diana M Bautista, e48431. ISSN: 2050-084X.

[28] Mingming Zhao et al. "The Roles of Orphan G Protein-Coupled Receptors in Autoimmune Diseases". In: *Clinical Reviews in Allergy & Immunology* 60.2 (Apr. 2021), pp. 220–243. ISSN: 1559-0267.

## A. Additional Details

| Metric | Explanation/Interpretation | Ref |
|---|---|---|
| reos = 1 | REOS (Rapid Elimination of Swill) refers to sets of rules designed to filter out drug candidates for having undesirable functional groups or features. DrugFlow's source code implements this using the useful_rdkit_utils package, which includes 8 distinct rule sets developed by different academic institutions/pharmaceutical companies/etc; reos=1 if a compound passes all 8 rule checks. | [25] |
| chembl_ring_systems = 1 | DrugFlow's source code uses functionality in the useful_rdkit_utils package to look up whether each ring system occurring in the generated ligand occurs in ligands in the manually curated ChEMBL database of small molecules; chembl_ring_systems=1 if all rings in the ligand are found. | [17] |
| novel = True | DrugFlow compares each generated ligand against previously generated ligands in a given inference run and considers a ligand to be novel if it has not been previously generated in this inference run. | |

Table 6 | DrugFlow built-in metrics and interpretation- based on source GitHub repo unless otherwise cited

| Chemical Validity/Consistency | Intramolecular Stability |
|---|---|
| RDKit Sanitizable | No Radicals |
| Convertible to InChI Strings[9] | No Invalid Bond Lengths |
| Molecule is Connected | No Invalid Bond Angles |
| Passes Kekulization | No Internal Steric Clashes |
| | Planar Aromatic Rings |
| | Planar Double Bonds |
| | Non-planar Non-Aromatic Rings |
| | Energy Ratio $< 100$ |

Table 7 | PoseBusters chemical validity and intramolecular stability checks- a ligand must meet every criteria listed above to be considered to have passed PoseBusters checks.