
DrugGen 2: A disease-aware language model for enhancing drug discovery

Ali Motahharynia^{1,2,†,*}, Mohammadreza Ghaffarzadeh-Esfahani^{1,†}, Mahsa Sheikholeslami^{1,3}, Navid Mazrouei¹, Matin Irajpour^{1,4}, Yousof Gheisari^{1,5}, Hajar Sirous⁶,

¹*Regenerative Medicine Research Center, Isfahan University of Medical Sciences, Isfahan, Iran*

²*Isfahan Neuroscience Research Center, Isfahan University of Medical Sciences, Isfahan, Iran*

³*Department of Medicinal Chemistry, School of Pharmacy, Isfahan University of Medical Sciences, Isfahan, Iran*

⁴*Isfahan Cardiovascular Research Center, Cardiovascular Research Institute, Isfahan University of Medical Sciences, Isfahan, Iran*

⁵*Department of Genetics and Molecular Biology, Isfahan University of Medical Sciences, Isfahan, Iran*

⁶*Bioinformatics Research Center, Isfahan University of Medical Sciences, Isfahan, Iran*

†These authors contributed equally to this work

*Correspondence:

Ali Motahharynia: alimotahharynia@gmail.com, Tel/Fax: +98-3136687087, ORCID: 0000-0002-1140-3257

Abstract

Current computational approaches for drug design typically focus on generating molecules conditioned on specific targets or general molecular properties, often neglecting the influence of disease context on target behavior and therapeutic outcomes. To address this gap, we introduce DrugGen-2, a novel generative model that designs small molecules conditioned on both disease ontology and target protein sequences. DrugGen-2 was developed by fine-tuning a pre-trained GPT-2 model on a curated dataset of approved drugs linked to their diseases and targets, using a two-step strategy of supervised fine-tuning followed by reinforcement learning via group relative policy optimization (GRPO). This process was guided by reward functions optimizing for chemical validity, novelty, diversity, and high predicted binding affinity. When evaluated on five protein targets relevant to diabetic nephropathy, DrugGen-2 significantly outperformed baseline models (DrugGPT and DrugGen). It demonstrated a superior capacity to generate unique molecules, exhibited greater structural similarity to approved drugs, and achieved improved predicted binding affinities across all targets. Molecular docking analyses further supported these findings, identifying candidate ligands with strong binding potential, including compounds with predicted affinities (−9.917, −9.485, and −9.367) exceeding those of reference drugs such as enalapril for angiotensin-converting enzyme (−8.283). By integrating disease-specific context into molecular generation, DrugGen-2 advances AI-assisted drug discovery, offering a powerful tool for de novo design and drug repurposing that accounts for the complex interplay between diseases and molecular targets.

Keywords: Drug design; Drug repositioning; Large language model; Reinforcement learning

1. Introduction

Drug discovery is a multifaceted and resource-intensive process that involves multiple stages, from the design of high-quality small molecules to clinical translation [1]. Traditionally, classical approaches such as phenotypic screening and target-based drug design have been used, wherein compounds were either screened for biological activity or synthesized to interact with specific molecular targets, but these strategies remain limited in their ability to efficiently explore the vast chemical space [2, 3]. Computational methods and machine learning (ML) have begun to address these challenges by enabling more accurate prediction of molecular properties, biological activity, and synthetic feasibility [4]. More recently, advances in deep learning, especially generative models, have further changed the field, allowing the design of novel molecules with improved efficiency by learning complex biological and chemical patterns [5]. Among generative approaches, large language models (LLMs) stand out for their capacity to process sequential molecular representations to capture subtle patterns in structure–property relationships [6]. Examples are models such as Mol-LLM [7], Token-Mol [8], and DrugLLM [9] that integrate multi-modal information for tasks including property prediction and conformation generation. When applied specifically to drug generation, these LLMs have facilitated the creation of antiviral candidates, such as TransAntivirus [10], as well as target-aware molecules exemplified by TamGen [11] and DrugGPT [12]. They have also facilitated ligand optimization through reinforcement learning, as demonstrated by DrugGen [13], which incorporates feedback mechanisms to enhance validity, novelty, and binding affinity. Despite these advancements, a notable limitation remains: most existing approaches condition molecular generation on protein targets or general molecular properties without tailoring generation to specific diseases, leaving a gap in disease-conditioned drug design. This limits their applicability in personalized therapeutic development, where the same target may behave differently across disease states, or where specific disease-associated pathways require context-specific molecular design [14, 15]. For example, peroxisome proliferator-activated receptor gamma (PPAR γ) activation exerts beneficial metabolic effects in type 2 diabetes but can drive divergent outcomes in colon cancer due to disease-specific interactions with the Wnt/ β -catenin pathway and metabolic state [16]. To fill this gap, we developed DrugGen-2, an extension of our previously reported DrugGen framework, designed to generate molecules conditioned on both disease ontology (medical subject headings (MeSH)) and target sequence. We developed DrugGen-2 by fine-tuning DrugGPT, a pretrained GPT-2 model specialized for ligand generation, on a curated dataset of approved drugs and their related disease-target information. The training strategy involved two complementary steps: supervised fine-tuning (SFT) [17] to adapt the model to the domain-specific data, followed by group relative policy optimization (GRPO) [18], a reinforcement learning approach designed to enhance the model’s generative capabilities for the generation of molecules that are chemically valid, structurally novel with high binding affinity, and similar to known approved drugs. Docking evaluations further confirm DrugGen-2’s ability to produce biologically relevant small molecules. By accepting a disease MeSH directed acyclic graph (DAG) and a target amino acid sequence as inputs, DrugGen-2 outputs high-quality simplified molecular input line entry system (SMILES) for that disease–target pair, offering a platform for both drug repurposing and de novo design, paving the way toward disease-aware drug discovery.

2. Results

We developed DrugGen-2 using a curated approved disease–target–drug dataset through a two-step training strategy: SFT followed by GRPO (Figure 1). In the supervised phase, the model was aligned with disease–target–drug relationships, while the reinforcement learning phase optimized generation toward producing chemically valid, novel, diverse, and high-affinity molecules. Because DrugGen-2 accepts MeSH DAG numbers as input, and diabetic nephropathy (DN) is annotated with four such identifiers, we evaluated performance separately for each entry. For

benchmarking, we focused on five protein targets associated with DN: angiotensin-converting enzyme (ACE), PPAR γ , nitric oxide synthase 3 (NOS3), plasminogen activator inhibitor-1 (PAI-1), and transforming growth factor beta 1 (TGF- β 1), as identified by DisGeNET [19] and DrugTar algorithm [20]. Outputs from DrugGen-2 were then compared with those generated by DrugGPT and DrugGen across multiple evaluation criteria, including generation capacity, structural validity, similarity to approved molecules, and predicted binding affinity.

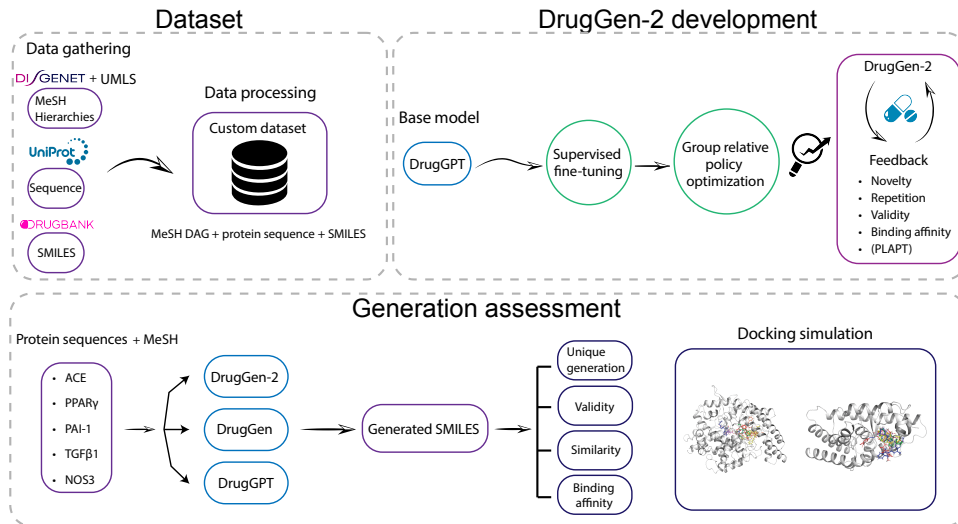


Figure 1. Schematic representation of DrugGen-2 development and evaluation. The top section shows dataset preparation and the training process through supervised fine-tuning (SFT) and group relative policy optimization (GRPO) using a customized reward function. The bottom section outlines the assessment process, based on unique generation, validity, similarity to approved drugs, and binding affinity for DrugGen-2, DrugGen, and DrugGPT, along with docking simulations for DrugGen-2.

2.1. Reward functions efficiently guided DrugGen-2 during training

In the training process, after six epochs of supervised fine-tuning using the SFT method, GRPO was employed as a reinforcement learning technique to further refine the model. Three reward functions were designed to guide this optimization process: (i) binding affinity, evaluated using the deep learning model, protein ligand binding affinity prediction using pre-trained transformers (PLAPT) [21], in combination with a customized invalid structure assessor, (ii) molecular diversity within each training batch, and (iii) novelty of generated molecules compared to approved drugs. The training was conducted for 10 epochs, during which all reward functions gradually converged, indicating a stable optimization process. (Figure 2A and Supplementary File 1).

2.2. DrugGen-2 achieves higher unique generation across targets compared to DrugGen and DrugGPT

We assessed each model’s capacity to generate unique molecules by tasking them to produce 500 unique candidates for the selected protein targets (Supporting information 2). Across all targets and MeSH configurations, DrugGen-2 consistently achieved the highest number of unique generations, ranging from 409 [406–427] to 444 [443–448] compared to DrugGen (50 [41–87]) and DrugGPT (219 [217–254]) ($\chi^2 = 88.80$, $P < 10^{-9}$, Cramér’s $V = 0.05$, Fig. 2B, Supporting information 3-Table 1-2). These results demonstrate that DrugGen-2 possesses a markedly greater capacity to generate unique molecules across targets, with consistent performance across different MeSH DAG inputs.

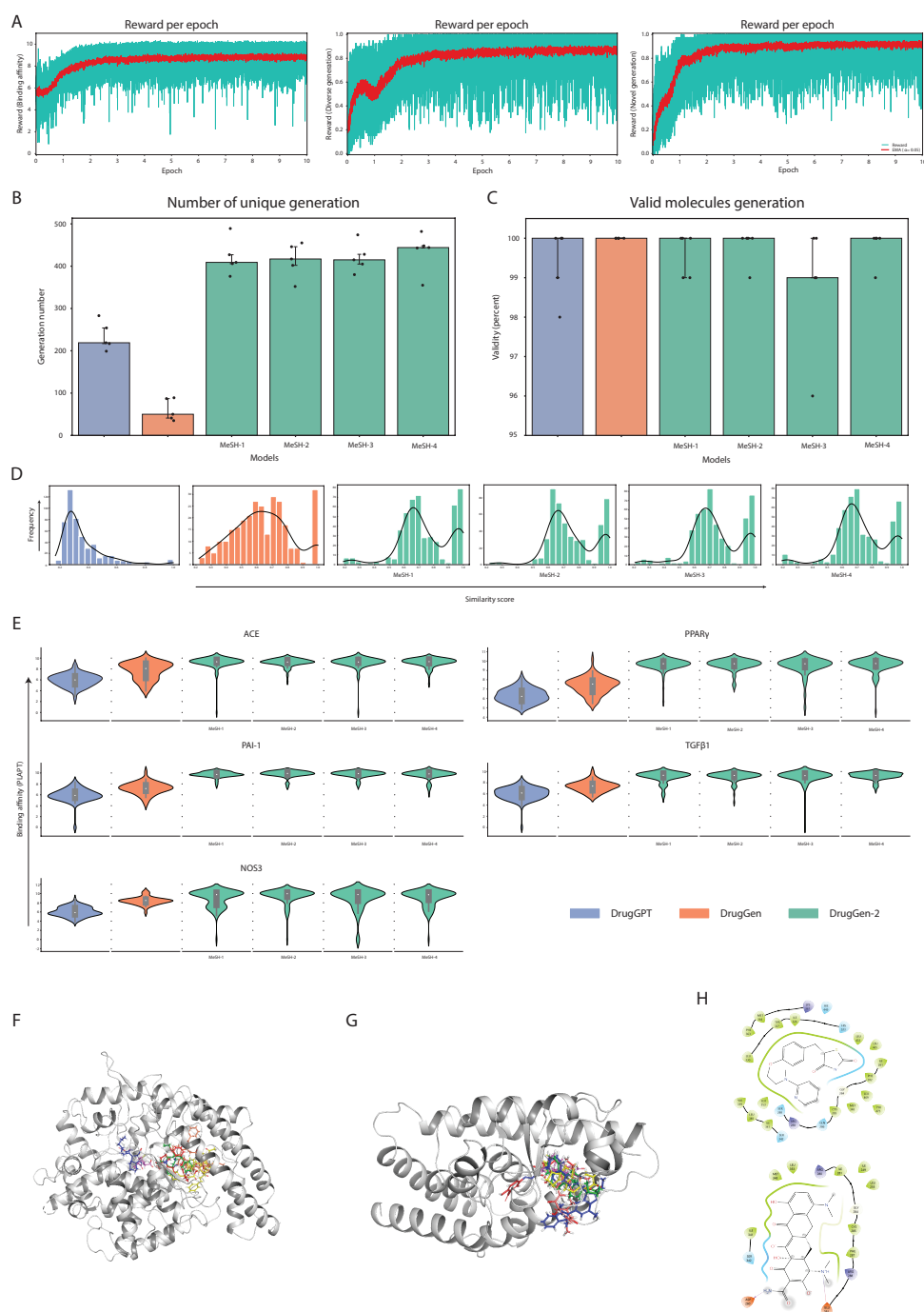


Figure 2. Evaluation of the DrugGen-2 algorithm. (A) Reward progression during group relative policy optimization (GRPO). Comparison of DrugGen-2 with DrugGen and DrugGPT in terms of (B) number of unique generations, (C) valid generations, (D) similarity to approved molecules, and (E) predicted binding affinity. (F) Visualization of protein–ligand complexes at the active site of 1UFZ (P12821), showing the binding of Enalapril (red), ligands P12821-293 (orange), P12821-10 (blue), P12821-269 (green), P12821-225 (yellow), and Captopril (pink). (G) Binding site of 4EMA (P37231) displaying Balsalazide (red), Rosiglitazone (pink), and ligands P37231-165 (orange), P37231-117 (blue), P37231-39 (green), and P37231-247 (yellow). (H) Visualization of 3D protein-ligand interactions, Rosiglitazone (top) and P37231-165 (bottom) in PPAR γ 's active site.

2.3. DrugGen-2 produces highly valid small molecules

We evaluated the structural and chemical validity of generated molecules using a customized validity assessor. For this analysis, we tasked each model to generate 100 unique molecules for each

target (DrugGPT and DrugGen) or per MeSH-target pair (DrugGen-2) (Supporting information 2). DrugGen-2 achieved near-perfect performance on generating valid small molecules across all five targets, with median values ranging from 99 to 100

2.4. DrugGen-2 demonstrates strong similarity to approved drugs

To assess the similarity between generated molecules and approved drugs, the same set of small molecules generated in the validity assessment was used. DrugGen-2 generated molecules with higher similarity to approved drugs across all MeSH categories (0.70), compared to DrugGPT (0.30 [0.26–0.41]) and DrugGen (0.64 [0.51–0.76]) ($H = 961.88$, $\epsilon^2 = 0.34$, $P < 10^{-204}$, Fig. 2D and Supporting information 2 and 3-Table 4-5). No significant differences were observed among the MeSH-specific DrugGen-2 variants. These findings indicate that DrugGen-2 produces molecules structurally closer to approved drugs, highlighting its potential in drug discovery.

2.5. DrugGen-2 demonstrates improved binding affinity across five key targets

The same set of molecules generated in the previous assessment was used to predict the binding affinity using PLAPT (21). DrugGen-2 consistently outperformed DrugGPT and DrugGen across all five DN-associated targets (Fig. 2E, Supporting information 2 and 3-Table 6). Median affinities for DrugGen-2 variants ranged from 9.26 to 9.97, markedly higher than those of DrugGPT (5.86 to 6.22) and DrugGen (7.15 to 8.49). These improvements were highly significant across all targets (ACE: $\epsilon^2 = 0.49$, $P < 10^{-59}$, PAI-1: $\epsilon^2 = 0.57$, $P < 10^{-65}$, PPAR γ : $\epsilon^2 = 0.57$, $P < 10^{-67}$, TGFB1: $\epsilon^2 = 0.50$, $P < 10^{-55}$, NOS3: $\epsilon^2 = 0.34$, $P < 10^{-38}$, Supporting information-3 Tables 7-11). No significant differences were observed among MeSH-specific DrugGen-2 variants, indicating consistent performance regardless of disease ontology encoding. Overall, DrugGen-2 exhibits enhanced molecular optimization, providing superior predicted binding across multiple targets.

2.6. Docking simulation determined the quality of the generated molecules

To evaluate the binding interactions of molecules generated by DrugGen-2, two proteins with available crystal structures from the protein data bank (PDB), ACE and PPAR γ , were selected. A total of 125 molecules were generated for each MeSH-target pair and subjected to docking analysis (Supporting information 4). Docking was performed using the GLIDE extra precision (XP) scoring function. Several designed ligands achieved docking scores comparable to or better than their reference drugs, suggesting enhanced binding potential (Table 1 and Supporting information 4). For ACE, compounds P12821-293, P12821-10, and P12821-269 showed markedly lower docking scores than Enalapril (-9.917, -9.485, and -9.367 vs. -8.283, respectively; Table 1), indicating stronger predicted affinities. Moreover, their binding poses within the ACE active site showed close spatial alignment with the reference ligand (Figure 2F). By contrast, none of the P37231 ligands surpassed Rosiglitazone or Balsalazide in docking score against PPAR γ (Figure 2G and Table 1), although some displayed unique pharmacophoric interactions within the active site. Figure 2H highlights the binding mode of P37231-165 in PPAR γ , which introduces novel substructures absent in Rosiglitazone and may form new optimizable interactions despite its lower overall docking score. To assess the reliability of the docking protocol, co-crystallized ligands were re-docked into their protein targets and root-mean-square deviation (RMSD) values calculated between experimental and predicted poses. For Rosiglitazone docked into PPAR γ (PDB ID: 4EMA), the RMSD was 1.44 Å, indicating faithful reproduction of the experimental pose and supporting the validity of the docking parameters. In contrast, captopril re-docked into ACE (PDB ID: 1UFZ) yielded an RMSD of 4.87 Å, likely due to the inherent flexibility of the ACE active site and potential limitations of the scoring function for metalloproteins. These results highlight that the docking protocol is generally robust, though predictive performance varies with target-ligand complexity.

Table 1. Docking scores of reference drugs and generated ligands docked into their respective targets, with lower scores indicating stronger predicted binding affinities.

PPAR γ		ACE	
Ligand	Docking score	Ligand	Docking score
Rosiglitazone (redock)	-7.422	P12821-293	-9.917
Balsalazide	-7.399	P12821-10	-9.485
P37231-165	-6.749	P12821-269	-9.367
P37231-117	-6.461	P12821-225	-9.29
P37231-39	-5.464	Enalapril	-8.283
P37231-247	-5.394	Captopril (redock)	-6.168

3. Discussion

In this study, we present DrugGen-2, a novel disease-aware generative model that advances disease-conditioned molecular design by generating small molecules tailored to both disease ontologies and target protein sequences. Compared with previously developed models, DrugGen-2 demonstrates improved generation capability, producing novel molecules with higher binding affinities and greater structural similarity to approved drugs. These advances demonstrate DrugGen-2’s potential to transform disease-aware drug discovery, bridging the gap between molecular design and translational pharmacology. DrugGen-2 includes disease context by conditioning on MeSH ontology hierarchies, which guide molecular generation to better align with disease biology. For instance, in DN, where multiple pathways converge on targets like PPAR γ or ACE, the model’s ontology-based conditioning enables it to generate ligands specified for a target in desired pathways. For instance, while ACE promotes angiotensin-2 formation in the cardiovascular system—leading to hypertension and vascular inflammation—it simultaneously degrades β -amyloid peptides in the brain, exerting neuroprotective effects [22, 23]. By capturing such context-dependent molecular functions, DrugGen-2 can design compounds that account for disease-specific complexities absent in earlier target-based models. This is evidenced by the superior unique molecule generation rates (up to 444 out of 500 attempts) and binding affinities (median pKd >9.0) across all evaluated targets, outperforming baselines like DrugGPT and DrugGen. Our results demonstrated DrugGen-2’s superior performance in generating unique molecules, reflecting an enhanced exploratory capacity in chemical space, an essential step in early-stage drug discovery, where structural diversity increases the likelihood of finding viable leads [24]. Across MeSH encodings, its variants consistently produced over 400 unique candidates per target, outperforming DrugGPT (median 219) and DrugGen (median 50). This results from GRPO integration with reward functions prioritizing novelty and diversity, reducing repetitive outputs. Additionally, near-perfect validity rates (99–100%) confirm DrugGen-2’s ability to maintain chemical feasibility and structural integrity during generation. The model’s generation of molecules with high similarity to approved drugs (median 0.70 across MeSH variants) and superior predicted binding affinities (medians 9.26–9.97) highlights its focus on clinically relevant molecular scaffolds. Docking simulations further supported these findings, with several ligands demonstrating stronger ACE binding than references like Enalapril, despite minor RMSD variations in flexible binding pockets. Together, the reinforcement learning framework using PLAPT-derived binding rewards guides output toward potent, diverse, and pharmacologically relevant candidates. This gives DrugGen-2 an advantage over DrugGPT and DrugGen in speeding hit-to-lead processes while minimizing risks associated with unexplored chemical entities. Despite these advances, limitations exist. While PLAPT predictions and GLIDE docking provide *in silico* insights, experimental validation—such as *in vitro* binding assays [25] or *in vivo* models [26]—is essential to confirm binding affinities, evaluate functional effects, and assess po-

tential off-target interactions. Moreover, the truncation of long protein sequences to 768 tokens may overlook distal domains critical for allosteric modulation [27]. Finally, the binary novelty reward might undervalue subtle innovations over radical departures from approved scaffolds. Future studies could incorporate advanced rewards, like absorption, distribution, metabolism, excretion, and toxicity (ADMET) predictions [28] or synthetic accessibility scores [29], to further bridge the gap to clinical candidates. In conclusion, DrugGen-2 represents a significant step toward AI-assisted pharmacology, demonstrating that disease-conditioned generative models can yield high-quality, context-aware drug candidates. By addressing the interplay between diseases and targets, it holds promise for expediting therapeutic development in an era of precision medicine.

4. Materials and Methods

4.1. Dataset preparation

We created a curated dataset linking approved drugs with their corresponding targets and associated diseases. First, approved drug–target pairs were extracted from DrugBank database (version: 5.1.10) [30]. Next, disease–target associations were obtained from DisGeNET database (version: 3.12.1) [19], which enabled the generation of preliminary disease–target–drug strings. To ensure clinical relevance, we incorporated approved drug–disease relationships (Phase IV) from the ChEMBL database (ChEMBL33) [31]. Preliminary strings were retained only if a direct drug–disease relationship was confirmed; otherwise, they were excluded. Following is the detailed information on dataset preparation.

4.1.1. Drug–target dataset

We retrieved drug–target information from DrugBank, yielding 1,710 small molecules with annotated human targets. Of these, 117 compounds were labeled as withdrawn. After a thorough assessment, 50 were excluded due to safety concerns or adverse effects. Available SMILES representations for the selected molecules (1,634 of 1,660) were obtained from DrugBank, ChEMBL, and ZINC20 [32] database. From a total of 2116 related protein targets, 27 were not present in UniProt [33] which were mapped to equivalent UniProt ID using reviewed UniProt ID, identical names, or the basic local alignment search tool (BLAST) [34]. The UniProt ID “Q5JXX5” was omitted from the database and therefore deleted from further analysis. Finally, the sequences for 2093 proteins were retrieved using the UniProt application programming interface (API).

4.1.2. Target–disease dataset

We extracted target–disease information from DisGeNET. The disease concept unique identifier (CUI) was updated from version 2019AA to 2023AA, and corresponding MeSH terms were retrieved via the unified medical language system (UMLS) terminology API. In total, 608 MeSH terms were mapped to 2042 protein targets.

4.1.3. Drug–disease dataset

We retrieved approved drug indications (disease–drug relationships) from ChEMBL. Following curation, 1,299 small molecules and their associated 643 MeSH terms were retained, forming the drug–disease dataset. To capture hierarchical disease relationships, MeSH terms were mapped to their corresponding DAG structures, resulting in 1,632 DAG representations.

4.2. Data processing

We adopted the original DrugGPT tokenizer and extended it with three additional special tokens, “D”, “P”, and “L”, added to the previously defined special tokens, i.e., “startoftext”, “endoftext”, and “PAD”, resulting in a total vocabulary of 53086 tokens. We created the strings of MeSH-sequence-SMILES (1,113,539) and filtered the strings that had the direct MeSH-SMILES relationship which resulted in 13,908 strings of MeSH-sequence-SMILES. Each string was tokenized in the following format: $\langle \text{startoftext} \rangle + \langle \text{D} \rangle + \langle \text{MeSH DAG} \rangle + \langle \text{P} \rangle + \langle \text{Sequence} \rangle + \langle \text{L} \rangle + \langle \text{SMILES} \rangle + \langle \text{endoftext} \rangle$. Sequences were padded to a length of 768 tokens, and longer sequences were truncated to this length.

4.3. Model Development

4.3.1. Supervised fine-tuning

Supervised fine-tuning was performed using the SFT trainer module from the transformer reinforcement learning (TRL) library (version: 0.9.4) [35] on the pre-trained DrugGPT model. Training was performed for 10 epochs on 13,908 tokenized MeSH-sequence-SMILES strings, with the model at epoch 6 selected for subsequent processing. The SFT configuration was as follows: learning rate = 5×10^{-4} , batch size = 8, linear warmup with 100 steps, gradient accumulation steps = 1, and the AdamW optimizer with a learning rate of 5×10^{-4} and epsilon value of 1×10^{-8} .

4.3.2. Group relative policy optimization

We employed GRPO to further refine the supervised fine-tuned model for de novo drug design. GRPO is a reinforcement learning method that eliminates the need for a separate value function by using group-level statistics to compute advantages and guide policy optimization. This makes GRPO more memory- and computation-efficient while maintaining strong performance.

For each prompt or query q , GRPO generates a group of G outputs $\{o_i\}_{i=1}^G$, where each output o_i is sampled from the old policy $\pi_{\theta_{\text{old}}}$ and then optimizes the policy model π_{θ} by maximizing the objective function $\mathcal{J}_{\text{GRPO}}(\theta)$. Each generated output o_i is scored according to a reward system to assess its binding affinity and validity, novelty compared to approved SMILES, and uniqueness of generation in each batch. The average reward across the group is computed and serves as a baseline to calculate an advantage $\hat{A}_{i,t}$ for the t -th token in each output o_i . The advantage measures how much that token’s contribution is compared to the baseline. The objective function optimizes the policy parameters θ to favor outputs with positive advantage, using a clipped ratio to ensure stable updates (Eq. 1). Additionally, a KL divergence penalty weighted by hyperparameter β is included between the new policy π_{θ} and a reference policy π_{ref} , the fine-tuned supervised model, to keep the updated policy close to known behaviors and prevent pathological shifts and large policy deviations (Eq. 2).

$$\mathcal{J}_{\text{GRPO}}(\theta) = \mathbb{E}_{q \sim P(Q), \{o_i\}_{i=1}^G \sim \pi_{\theta_{\text{old}}}(O|q)} \left[\frac{1}{G} \sum_{i=1}^G \frac{1}{|o_i|} \sum_{t=1}^{|o_i|} \left\{ \min \left(\frac{\pi_{\theta}(o_{i,t} | q, o_{i,<t})}{\pi_{\theta_{\text{old}}}(o_{i,t} | q, o_{i,<t})} \hat{A}_{i,t}, \right. \right. \\ \left. \left. \text{clip} \left(\frac{\pi_{\theta}(o_{i,t} | q, o_{i,<t})}{\pi_{\theta_{\text{old}}}(o_{i,t} | q, o_{i,<t})}, 1 - \epsilon, 1 + \epsilon \right) \hat{A}_{i,t} \right) - \beta D_{\text{KL}}[\pi_{\theta} \| \pi_{\text{ref}}] \right\} \right] \quad (1)$$

where ϵ and β are hyperparameters controlling the clipping range and KL penalty weight, respectively.

The KL divergence term is defined as:

$$D_{\text{KL}}[\pi_{\theta} \| \pi_{\text{ref}}] = \frac{\pi_{\text{ref}}(o_{i,t} | q, o_{i,<t})}{\pi_{\theta}(o_{i,t} | q, o_{i,<t})} - \log \left(\frac{\pi_{\text{ref}}(o_{i,t} | q, o_{i,<t})}{\pi_{\theta}(o_{i,t} | q, o_{i,<t})} \right) - 1 \quad (2)$$

We implemented GRPO using a custom `GRPOTrainer` class from the Hugging Face Transformers TRL library (version 0.23.1). Training configurations were defined through a `GRPOConfig` object specifying both optimization and generation parameters.

The optimization was carried out with the AdamW optimizer, using hyperparameters $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 1 \times 10^{-8}$, and a weight decay of 0.1. The learning rate was initialized at 5×10^{-6} and adjusted using a linear learning rate scheduler, with the first 100 steps designated for warm-up. To stabilize updates, we applied gradient accumulation with an effective batch factor of 4, enabled mixed-precision training (fp16), and enforced gradient clipping at a maximum norm of 1.0. The total training was performed for 10 epochs. Logging and checkpointing were executed at fixed step intervals. For reproducibility, both the global seed and data seed were set to 42.

SMILES generation during training was controlled by a defined set of hyperparameters. The maximum input length was fixed at 768 tokens, while the GRPO group size (number of completions per prompt) was set to 40. The model operated with a context window of 1,024 tokens, and the maximum generation length was limited to 256 tokens (i.e., $1024 - 768$). Sampling was performed with default values of 1.0 for top p and None for top k . The temperature was set to 1.0 to preserve stochasticity in outputs.

4.3.3. Validity-checker pipeline

During GRPO training, we implemented a validity-checker pipeline using the RDKit cheminformatics toolkit (version 2023.9.5) [36] to filter out syntactically or chemically invalid molecular representations. The pipeline applies a series of rule-based checks designed to detect structural inconsistencies. First, each SMILES string is read by RDKit, and strings that cannot be interpreted or do not produce a valid molecular graph are flagged as invalid. Molecules containing fewer than two atoms are also rejected, as such structures are considered chemically trivial. Valence state consistency is then examined, which enforces chemically permitted bonding rules; molecules that contain atoms with impossible valence states are classified as invalid. The results of these validation steps are integrated into a unified system, whereas molecules passing all checks are classified as chemically valid.

4.3.4. Reward functions

During training, three distinct reward functions were employed to guide optimization toward generating molecules with desirable properties: maximizing predicted binding affinity while penalizing invalid structures, ensuring novelty relative to approved drugs, and encouraging diversity by preventing repetition within each batch.

- **Binding affinity:** To measure the binding affinity, we used PLAPT, which is a deep learning model developed to measure the binding affinity between protein and SMILES. This reward function computes reward values for batches of generated molecular sequences by evaluating candidate SMILES strings against corresponding target proteins. For each input-completion pair, the function first checks the validity of the generated SMILES using the validity-checker pipeline. Invalid SMILES are assigned a reward of zero. If the SMILES is valid, the function extracts the protein sequence embedded within the input prompt, specifically the segment located between the <P> and <L> tags. This extracted sequence, together with the candidate SMILES, is then passed to the PLAPT model, which returns a predicted binding affinity score defined as the negative logarithm (base 10) of the dissociation constant (K_d). The function ultimately returns a list of rewards corresponding to each input pair of prompt and completion.
- **Similarity to approved molecules:** To promote novelty, each generated molecule was compared against our dataset of approved drugs. A binary reward was assigned such that

molecules already present in the dataset received a reward of zero, whereas novel molecules absent from the dataset received a reward of one.

- Avoidance of repeated molecules within a batch: To discourage reward hacking through duplication, we implemented a batch-level reward. Each generated SMILES was assigned a binary reward of one if it was unique within the batch, or zero if it appeared multiple times in the same batch. This ensured diversity among the generated candidates during training.

4.4. Assessment

To evaluate the capabilities of DrugGen-2 in generating high-quality small molecules, we conducted a series of experiments based on selecting diabetic nephropathy as a representative example. Initially, we identified DN-related targets from DisGeNET and filtered five highly drug-gable target proteins using the DrugTar algorithm: ACE, PPAR γ , NOS3, PAI-1, and TGF- β 1. For model input preparation, DrugGen-2 incorporates disease information through MeSH DAG structures paired with target proteins. In the case of DN, the MeSH descriptor "Diabetic Nephropathies" was mapped to the following hierarchical identifiers:

- C12.050.351.968.419.192: Urogenital Diseases, Female Urogenital Diseases and Pregnancy Complications, Female Urogenital Diseases, Urologic Diseases, Kidney Diseases, Diabetic Nephropathies
- C12.200.777.419.192: Urogenital Diseases, Male Urogenital Diseases, Urologic Diseases, Kidney Diseases, Diabetic Nephropathies
- C12.950.419.192: Urogenital Diseases, Urologic Diseases, Kidney Diseases, Diabetic Nephropathies
- C19.246.099.875: Endocrine System Diseases, Diabetes Mellitus, Diabetes Complications, Diabetic Nephropathies

We then compared the performance of DrugGen-2 against DrugGen and DrugGPT models with their default parameters across five key evaluation metrics: the capability of generating unique small molecules, the validity of the generated molecules, the similarity of these molecules to approved drugs, and their binding affinity to the selected targets. Finally, to provide deeper insights into DrugGen-2’s outputs, we conducted docking simulations for further analysis.

4.4.1. Generation power

To assess the models’ capability to generate unique small molecules, we directed each model to produce 500 unique small molecules per input. For DrugGPT and DrugGen, inputs consisted of the five selected DN-associated target proteins, whereas DrugGen-2 was queried with MeSH DAG–target protein pairs. If a model failed to generate novel molecules within 30 sampling cycles, the generation process for that input was terminated. The number of successfully generated unique molecules was taken as a measure of generative capability. Statistical analysis was conducted using the χ^2 test, and correction for multiple comparisons was performed using the Benjamini–Hochberg method.

4.4.2. Validity assessment

For each DrugGen-2 input pair (MeSH DAG–target protein) and for each target protein input of DrugGen and DrugGPT, the models were instructed to generate 100 unique molecules. The outputs were filtered using the validity-checker pipeline described earlier, and the percentage of valid structures was reported as a measure of each model’s ability to generate chemically valid molecules. Statistical analysis was conducted using the χ^2 test.

4.4.3. Similarity to the approved small molecules

To evaluate the similarity of generated molecules to approved drugs, we used the same set of outputs from the validity assessment. Each valid molecule, as determined by the RDKit pipeline, was converted into a Morgan fingerprint (radius = 2, size = 2048 bits). Pairwise Tanimoto similarity scores were computed between the fingerprints of generated and approved SMILES, and the maximum similarity value was reported as the final similarity score. Statistical analysis was conducted using the Kruskal–Wallis test, and corrections for multiple comparisons were applied using the Benjamini–Hochberg method.

4.4.4. Binding affinity using PLAPT

The same set of molecules generated during the validity assessment was used to evaluate the binding affinities of the compounds produced by DrugGen-2, DrugGen, and DrugGPT. The invalid structures were considered in the binding affinity assessments. Statistical analysis was conducted using the Kruskal–Wallis test, and corrections for multiple comparisons were applied using the Benjamini–Hochberg method.

4.4.5. Docking simulation

To investigate binding interactions, molecular docking was performed against two proteins with available crystal structures from the Protein Data Bank (PDB): ACE (UniProt ID: P12821 and PDB ID: 1UFZ) and PPAR γ (UniProt ID: P37231 and PDB ID: 4EMA). These targets were chosen for their therapeutic relevance, as well as the availability of co-crystallized ligands, which allowed validation of the docking protocol. After removing duplicate entries, a curated library of newly designed ligands was docked using a blind docking approach. For benchmarking, at least one approved small-molecule drug per target was included, retrieved from DrugBank. Docking grids were uniformly set to $40 \times 40 \times 40 \text{ \AA}^3$, the maximum permissible by the software, to ensure comprehensive exploration of potential binding regions [37].

Protein structures were prepared using Schrödinger’s Protein Preparation Wizard [38], including hydrogen addition, optimization of hydrogen-bonding networks, and assignment of protonation states at physiological pH (7.4 ± 0.5). Ligands were processed with LigPrep [39] under the OPLS4 force field [40], generating all relevant ionization and tautomeric states.

Docking simulations were carried out in Maestro (v2021-2) using the GLIDE XP algorithm [41]. Ligands were treated as flexible, while proteins were kept rigid. Reliability of the docking workflow was assessed by re-docking the original co-crystallized ligands into their respective protein structures. Docking outcomes were ranked by GLIDE XP scores, where more negative values correspond to stronger predicted binding affinities. Complete docking data and SMILES representations of the ligands are available in Supporting Information 4.

Data availability

All data generated or analyzed during this study are included in the manuscript and supporting information files. The MeSH-sequence-SMILES dataset of approved disease-target-drug pairs used in this study is publicly available at https://huggingface.co/datasets/alimotahharynia/approved_disease_target_drug.

Code availability

The checkpoints and code for generating small molecules are publicly available at <https://huggingface.co/alimotahharynia/DrugGen-2> and <https://github.com/alimotahharynia/DrugGen-2>, respectively. Explore the interactive user interface of DrugGen-2 at <https://huggingface.co/spaces/alimotahharynia/DrugGen-2>.

Acknowledgments

We sincerely thank Dr. Farnoush Kiyanpour for her valuable comments on the first draft of the manuscript.

Funding

No funding was received for this study or its publication.

Competing interests

The authors declare no competing interests.

Author contributions

Conceptualization: A.M, Y.G. Dataset preparation: A.M, M.S, N.M. Model development: A.M, M.G, N.M, M.I. Model assessment: A.M, M.G. Statistical analysis: A.M, M.G. Molecular docking: M.S, H.S. Data interpretation: All authors. Drafting original manuscript: A.M, M.G, M.S, N.M. Revising the manuscript: A.M, M.I, H.S, Y.G. All the authors have read and approved the final version for publication and agreed to be responsible for the integrity of the study.

References

- [1] Majumder S, Panigrahi GK. Advancements in contemporary pharmacological innovation: Mechanistic insights and emerging trends in drug discovery and development. *Intell Pharm.* 2025;3(2):118–26. doi: 10.1016/j.ipha.2024.10.001
- [2] Vincent F, Nueda A, Lee J, Schenone M, Prunotto M, Mercola M. Phenotypic Drug Discovery: Recent successes, lessons learned and new directions. *Nat Rev Drug Discov.* 2022;21(12):899–914. doi: 10.1038/s41573-022-00472-w
- [3] Vasaikar S, Bhatia P, Bhatia PG, Chu Yaiw K. Complementary Approaches to Existing Target Based Drug Discovery for Identifying Novel Drug Targets. *Biomedicines.* 2016;4(4):27. doi: 10.3390/biomedicines4040027
- [4] Staszak M, Staszak K, Wieszczycka K, Bajek A, Roszkowski K, Tylkowski B. Machine learning in drug design: Use of artificial intelligence to explore the chemical structure–biological activity relationship. *WIREs Comput Mol Sci.* 2022;12(2):e1568. doi: 10.1002/wcms.1568
- [5] Özçelik R, Brinkmann H, Criscuolo E, Grisoni F. Generative Deep Learning for de Novo Drug Design—A Chemical Space Odyssey. *J Chem Inf Model.* 65(14):7352–72. doi: 10.1021/acs.jcim.5c00641
- [6] Sadeghi S, Bui A, Forooghi A, Lu J, Ngom A. Can large language models understand molecules? *BMC Bioinformatics.* 2024;25(1):225. doi: 10.1186/s12859-024-05847-x
- [7] Lee C, Ko H, Song Y, Jeong Y, Hormazabal R, Han S, et al. Mol-LLM: Multimodal Generalist Molecular LLM with Improved Graph Utilization. arXiv; 2025. [accessed 3 June 2026] Available from: <http://arxiv.org/abs/2502.02810>
- [8] Wang J, Qin R, Wang M, Fang M, Zhang Y, Zhu Y, et al. Token-Mol 1.0: tokenized drug design with large language models. *Nat Commun.* 2025;16(1):4416. doi: 10.1038/s41467-025-59628-y
- [9] Liu X, Guo Y, Li H, Liu J, Huang S, Ke B, et al. DrugLLM: Open Large Language Model for Few-shot Molecule Generation. arXiv; 2024. [accessed 23 Sept 2025] Available from: <http://arxiv.org/abs/2405.06690>
- [10] Mao J, Wang J, Zeb A, Cho K-H, Jin H, Kim J, et al. Transformer-Based Molecular Generative Model for Antiviral Drug Design. *J Chem Inf Model.* 2023;64(7):2733–45. doi: 10.1021/acs.jcim.3c00536
- [11] Wu K, Xia Y, Deng P, Liu R, Zhang Y, Guo H, et al. TamGen: drug design with target-aware molecule generation through a chemical language model. *Nat Commun.* 2024;15(1):9360. doi: 10.1038/s41467-024-53632-4
- [12] Li Y, Gao C, Song X, Wang X, Xu Y, Han S. DrugGPT: A GPT-based Strategy for Designing Potential Ligands Targeting Specific Proteins. 2023 June 30; [accessed 23 Sept 2025] Available from: <https://www.biorxiv.org/content/10.1101/2023.06.29.543848v1>
- [13] Sheikholeslami M, Mazrouei N, Gheisari Y, Fasihi A, Irajpour M, Motahharynia A. DrugGen enhances drug discovery with large language models and reinforcement learning. *Sci Rep.* 2025;15(1):13445. doi: 10.1038/s41598-025-98629-1
- [14] Shim H. One target, different effects: a comparison of distinct therapeutic antibodies against the same targets. *Exp Mol Med.* 2011;43(10):539–49. doi: 10.3858/emm.2011.43.10.063

- [15] Su J, Yang L, Sun Z, Zhan X. Personalized Drug Therapy: Innovative Concept Guided With Proteoformics. *Mol Cell Proteomics MCP*. 2024;23(3):100737. doi: 10.1016/j.mcpro.2024.100737
- [16] Lecarpentier Y, Claes V, Vallée A, Hébert J-L. Interactions between PPAR Gamma and the Canonical Wnt/ β -Catenin Pathway in Type 2 Diabetes and Colon Cancer. *PPAR Res*. 2017;2017:5879090. doi: 10.1155/2017/5879090
- [17] Lu W, Luu RK, Buehler MJ. Fine-tuning large language models for domain adaptation: exploration of training strategies, scaling, model merging and synergistic capabilities. *Npj Comput Mater*. 2025;11(1):84. doi: 10.1038/s41524-025-01564-y
- [18] Shao Z, Wang P, Zhu Q, Xu R, Song J, Bi X, et al. DeepSeekMath: Pushing the Limits of Mathematical Reasoning in Open Language Models. arXiv; 2024. [accessed 17 Aug 2025] Available from: <http://arxiv.org/abs/2402.03300>
- [19] Piñero J, Bravo À, Queralt-Rosinach N, Gutiérrez-Sacristán A, Deu-Pons J, Centeno E, et al. DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic Acids Res*. 2017;45(D1):D833–9. doi: 10.1093/nar/gkw943
- [20] Borhani N, Izadi I, Motahharynia A, Sheikholeslami M, Gheisari Y. DrugTar improves drug-gabability prediction by integrating large language models and gene ontologies. *Bioinformatics*. 2025;41(7):btaf360. doi: 10.1093/bioinformatics/btaf360
- [21] Rose T, Monti N, Anand N, Shen T. PLAPT: Protein-Ligand Binding Affinity Prediction Using Pretrained Transformers. 2024 Feb 9; [accessed 29 Sept 2025] Available from: <https://www.biorxiv.org/content/10.1101/2024.02.08.575577v1>
- [22] Le D, Brown L, Malik K, Murakami S. Two Opposing Functions of Angiotensin-Converting Enzyme (ACE) That Links Hypertension, Dementia, and Aging. *Int J Mol Sci*. 2021;22(24):13178. doi: 10.3390/ijms222413178
- [23] Hu J, Igarashi A, Kamata M, Nakagawa H. Angiotensin-converting enzyme degrades Alzheimer amyloid beta-peptide ($A\beta$); retards $A\beta$ aggregation, deposition, fibril formation; and inhibits cytotoxicity. *J Biol Chem*. 2001;276(51):47863–8. doi: 10.1074/jbc.M104068200
- [24] Gangwal A, Ansari A, Ahmad I, Azad AK, Kumarasamy V, Subramaniyan V, et al. Generative artificial intelligence in drug discovery: basic framework, recent advances, challenges, and opportunities. *Front Pharmacol*. 2024;15:1331062. doi: 10.3389/fphar.2024.1331062
- [25] McFedries A, Schwaid A, Saghatelian A. Methods for the elucidation of protein–small molecule interactions. *Chem Biol*. 2013;20(5):667–73. doi: 10.1016/j.chembiol.2013.04.008
- [26] Schürmann M, Janning P, Ziegler S, Waldmann H. Small-Molecule Target Engagement in Cells. *Cell Chem Biol*. 2016;23(4):435–41. doi: 10.1016/j.chembiol.2016.03.008
- [27] Chandra A, Tünnermann L, Löfstedt T, Gratz R. Transformer-based deep learning for predicting protein properties in the life sciences. Dötsch V, editor. *eLife*. 2023;12:e82819. doi: 10.7554/eLife.82819
- [28] Guan L, Yang H, Cai Y, Sun L, Di P, Li W, et al. ADMET-score – a comprehensive scoring function for evaluation of chemical drug-likeness. *MedChemComm*. 2019;10(1):148–57. (Electronic supplementary information available.) doi: 10.1039/c8md00472b
- [29] Ertl P, Schuffenhauer A. Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions. *J Cheminformatics*. 2009;1:8. doi: 10.1186/1758-2946-1-8

- [30] Wishart DS, Knox C, Guo AC, Shrivastava S, Hassanali M, Stothard P, et al. DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res.* 2006;34(Database issue):D668–72. doi: 10.1093/nar/gkj067
- [31] Adasme Mora MF, Arcila Toro R, Blackshaw J, Bosc N, Corbett S, De Veij M, et al. ChEMBL database release 33. ChEMBL Database. 2011. doi: 10.6019/CHEMBL.database.33
- [32] Irwin JJ, Tang KG, Young J, Dandarchuluun C, Wong BR, Khurelbaatar M, et al. ZINC20 – A Free Ultralarge-Scale Chemical Database for Ligand Discovery. *J Chem Inf Model.* 2020;60(12):6065–73. doi: 10.1021/acs.jcim.0c00675
- [33] UniProt Consortium. UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Res.* 2023;51(D1):D523–31. doi: 10.1093/nar/gkac1052
- [34] Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol.* 1990;215(3):403–10. doi: 10.1016/S0022-2836(05)80360-2
- [35] Hugging Face. huggingface/trl. Hugging Face; 2025. [accessed 29 Sept 2025] Available from: <https://github.com/huggingface/trl>
- [36] RDKit. Release 2023_09_5 (Q3 2023) Release · rdkit/rdkit. GitHub. [accessed 17 Aug 2025] Available from: https://github.com/rdkit/rdkit/releases/tag/Release_2023_09_5
- [37] Hassan NM, Alhossary AA, Mu Y, Kwoh C-K. Protein-Ligand Blind Docking Using QuickVina-W With Inter-Process Spatio-Temporal Integration. *Sci Rep.* 2017;7(1):15451. doi: 10.1038/s41598-017-15571-7
- [38] Sastry GM, Adzhigirey M, Day T, Annabhimoju R, Sherman W. Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *J Comput Aided Mol Des.* 2013;27(3):221–34. doi: 10.1007/s10822-013-9644-8
- [39] Schrödinger Release 2025-3: LigPrep, Schrödinger, LLC, New York, NY, 2025.
- [40] Lu C, Wu C, Ghoreishi D, Chen W, Wang L, Damm W, et al. OPLS4: Improving Force Field Accuracy on Challenging Regimes of Chemical Space. *J Chem Theory Comput.* 2021;17(7):4291–300. doi: 10.1021/acs.jctc.1c00302
- [41] Friesner RA, Murphy RB, Repasky MP, Frye LL, Greenwood JR, Halgren TA, et al. Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein–ligand complexes. *J Med Chem.* 2006;49(21):6177–96. doi: 10.1021/jm051256o