COMPUTATIONAL BIOLOGY

Steering semi-flexible molecular diffusion model for structure-based drug design with reinforcement learning

Xudong Zhang[1,2,3], Sanqing Qu[3], Fan Lu[3], Jianmin Wang[4], Zhixin Tian[5], Shangding Gu[6], Yanping Zhang[7]*, Alois Knoll[8], Shaorong Gao[7], Guang Chen[1,2,3,8]*, Changjun Jiang[3]

Current structure-based molecular generation faces a fundamental dilemma: While static ligand modeling dominates computational approaches, real-world molecular interactions are inherently dynamic. Inspired by the conformational changes ligands undergo during semi-flexible docking, we propose a reinforcement learning (RL)–steered diffusion framework for semi-flexible molecular generation in protein pockets. By defining the denoising process as a Markov decision process, RL dynamically adjusts molecular structures through iterative exploration. Simultaneously, we incorporate multiple molecular properties as conditions to constrain the denoising policy to drug-like regions and perform self-supervised rigid training on both target-free and target-specific molecules. In addition, we propose a fast sampling strategy that accelerates sampling by 20 times, thereby improving the efficiency of training and sampling. Experiments demonstrate that our method outperforms state-of-the-art methods with a Vina score of −7.23 kcal/mol and an 11.53% success rate. Targeting unseen real-world proteins, the generated molecules preserve canonical interaction patterns while discovering previously unknown binding chemotypes.

INTRODUCTION

Structure-based drug discovery (SBDD) is reshaping modern drug development paradigms. Technological advances such as cryo–electron microscopy (cryo-EM) single-particle analysis (*1*, *2*) and artificial intelligence (AI)–driven protein structure prediction methods such as AlphaFold (*3*) have made it possible to determine high-resolution three-dimensional (3D) structures of protein targets at scale. These breakthroughs enable rapid virtual screening of millions of compounds against target pockets (*4*–*6*). However, virtual screening remains confined to a predefined chemical space, limiting its capacity to discover previously unseen chemotypes, particularly for structurally complex or previously unseen binding pockets (*7*).

To overcome the limitations of fixed chemical libraries, deep learning–based generative models have emerged as a powerful alternative, enabling the design of diverse molecules directly conditioned on protein structures (*8*–*10*). Early explorations of molecular generation are primarily conducted in low-dimensional spaces, representing molecules as sequences or 2D graphs. Although these representations allow for rapid sampling through language models (*11*, *12*) and graph neural networks (GNNs) (*13*–*15*), the dimensionality reduction comes with biophysical costs. Because of the lack of 3D spatial configuration, molecules generated through these models may not meet stability and activity requirements in practical applications. The advent of 3D geometric deep learning alleviates this dimensionality barrier, with architectures equivariant to special Euclidean group SE(3) (*16*, *17*) enabling accurate modeling of molecules with 3D structural information and the de novo generation of 3D molecular structures with high spatial fidelity.

Recent research in structure-based molecular generation (SBMG) has integrated protein-ligand complexes into 3D generative architectures (*18*–*20*). Current SBMG models mainly adopt two paradigms: autoregressive and one-shot generation. Autoregressive methods generate molecular structures step-by-step, capturing the complexity and dependencies of molecular sequences. For instance, AR (*21*) uses the protein pocket as a 3D context to estimate the probability density of atoms in space and sequentially samples atoms from the learned distribution. Pocket2Mol (*22*) and ResGen (*23*) use equivariant GNNs to learn the chemical and geometric constraints imposed by protein pockets, considering atom and bond information, and generate molecules in an autoregressive manner. Because of inefficiencies and the risk of generating unrealistic structures when generating atom by atom, FLAG (*24*) learns the geometric structure of protein-ligand complexes through invariant GNNs and introduces fragment priors to conduct autoregressive sampling at the fragment level. However, autoregressive methods suffer from error accumulation and limited structural validity due to their inherently sequential nature.

By contrast, one-shot generative models (*25*–*28*) offer improved training efficiency and the ability to model global structure holistically. These models simulate molecular transitions from disorder to in 3D space. Recent diffusion-based SBMG models such as TargetDiff (*29*), DecompDiff (*30*), and IPDIFF (*31*) learn the joint distribution of atom types and coordinates, and MolCRAFT (*32*) models them in a continuous parameter space. These methods generate molecules in continuous 3D space based on protein pockets and achieve invariant likelihood to global translation and rotation of pocket-ligand complexes. Despite these advancements, most SBMG models assume

[1]Shanghai Key Laboratory of Maternal Fetal Medicine, Clinical and Translational Research Center of Shanghai First Maternity and Infant Hospital, School of Computer Science and Technology, Tongji University, Shanghai 200092, China. [2]Shanghai Innovation Institute, Shanghai, China. [3]MOE Key Laboratory of Embedded System and Service Computing, School of Computer Science and Technology, Tongji University, Shanghai, China. [4]The Interdisciplinary Graduate Program in Integrative Biotechnology, Yonsei University, Incheon, Republic of Korea. [5]School of Chemical Science & Engineering and Shanghai Key Laboratory of Chemical Assessment and Sustainability, Tongji University, Shanghai, China. [6]Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA, USA. [7]Shanghai Key Laboratory of Maternal Fetal Medicine, Clinical and Translational Research Center of Shanghai First Maternity and Infant Hospital, Frontier Science Center for Stem Cell Research, School of Life Sciences and Technology, Tongji University, Shanghai 200092, China. [8]Department of Informatics, Technical University of Munich, Munich, Germany.
*Corresponding author. Email: yanpingzhang@tongji.edu.cn (Y.Z.); guang@in.tum.de (G.C.)

rigid ligand conformations. As a result, dynamic coupling effects are often ignored, limiting the ability of models to generate high-quality ligands for novel targets. Moreover, existing methods, constrained by the quality and quantity of receptor-ligand complex data, focus exclusively on generating high-affinity molecules, neglecting the coordination required for multiple molecular property optimization. Consequently, the generated molecules often fail to exceed the quality boundaries of the training set.

To address these limitations, we introduce SeFMol, a Semi-Flexible Molecular diffusion model steered by reinforcement learning (RL). Inspired by semi-flexible docking paradigms (*33*, *34*), SeFMol enables conformational adjustments during training to better accommodate binding sites. This process simulates the restricted degrees-of-freedom optimization commonly used in semi-flexible docking to explore low-energy poses for a fixed protein. Specifically, we formulate the denoising phase of diffusion as a Markov decision process (MDP), allowing RL to guide the structural evolution of molecules based on pocket geometry and property conditions. We condition generation on multiple molecular properties and leverage both target-free and target-specific training to enhance property controllability and target specificity. In addition, we introduce a variable fast sampling strategy that reduces sampling steps by 20 times (from 1000 to 50 steps), markedly improving both training and inference efficiency. We conduct comprehensive experiments to validate the generality and efficacy of SeFMol. Our evaluations span docking performance, drug-like properties, interaction patterns, conformational quality, novelty, and sampling efficiency. Last, we demonstrate SeFMol's generalization to real-world therapeutic targets, highlighting its potential as a robust framework for structure-based drug design.

RESULTS

SBMG with SeFMol

We provide an overview of the semi-flexible diffusion model SeFMol (Fig. 1), elucidating its structural mechanism and detailing the training and sampling processes. The training process of SeFMol consists of a rigid training phase, followed by a semi-flexible optimization phase (Fig. 1A). The rigid training phase uses static training data and a self-supervised learning strategy for molecular reconstruction, concentrating solely on static conformations, which could result in less optimal interaction patterns. The model trained in the rigid phase is further refined through semi-flexible optimization steered by RL (SFRL), which aims to optimize a reward to progressively explore chemically plausible interaction patterns.

The two-stage rigid training process of SeFMol (Fig. 1B) is founded on a conditional diffusion model. During the pretraining phase, 1 million molecules without protein targets were selected from the Molecule3D dataset (*35*), and eight molecular properties were computed using RDKit (*36*). These properties were input into a linear layer to obtain property embeddings y. Subsequently, during the diffusion process at time step t, small Gaussian noise and uniform noise were added to the atomic coordinates and types of 3D molecules according to a fixed variance schedule of a Markov chain. These noises were processed through two linear layers to obtain embedded features x_t and h_t. The property features y were then integrated with the coordinate and atomic type features, and the fused features were input into a denoiser constructed using an equivariant GNN to predict the original coordinates and atomic types, thereby imparting property bias to the denoising model. In the fine-tuning phase, 100,000 protein-ligand pairs and their corresponding molecular properties from the CrossDocked2020 dataset (*18*) were used to fine-tune the pretrained diffusion model. In this stage, the geometric spatial and interaction information of the protein-ligand pairs were taken into account, and noise injection and removal for the molecules were continued. This process ensures that the denoiser maintains property bias while generating target-aware 3D molecular structures. Hereafter, the term rigid model refers to the trained model obtained through this process.

Figure 1C illustrates the semi-flexible optimization of the rigid model steered by RL. In this phase, the denoising step is modeled as an MDP, and gradient descent is used to optimize the reward signal. At time step t, the current molecular state M_t serves as state s_t and is simultaneously input into the reference denoiser (with frozen parameters) and the policy denoiser (in a fine-tuning state). To prevent excessive deviation of the policy denoiser from the reference denoiser, a Kullback-Leibler (KL) constraint is applied to the predicted positions by these two denoisers. In addition, the policy denoiser executes the current denoising action a_t, and the molecular result processed by a_t becomes the next molecular state M_{t-1}. A molecule is reconstructed based on M_{t-1}, and a reward is calculated using a reward function. This reward is then combined with the output of the value network to update the parameters of the policy denoiser. Intermediate molecular states during denoising lack practical significance and cannot calculate affinity, resulting in sparse reward signals that affect exploration efficiency. To address this, we incorporate the original property biases of the rigid model, designating various molecular properties as conditional signals to enhance the controllability of properties and affinity. In addition, diffusion model sampling typically requires numerous steps, leading to excessive redundant intermediate states and sparse rewards. To mitigate this, we introduce a fast sampling strategy during optimization, reducing the sampling steps to $^1/_{20}$ of the original, thereby improving the training and sampling efficiency of the molecular diffusion model.

During the sampling phase, SeFMol retains the carefully fine-tuned molecular diffusion model, discarding auxiliary modules used during training. The desired molecular properties of the target molecule are set, and sampling begins from state $\mathcal{N}(0, I)$ to initialize the state. Subsequently, the transition probability for the given target protein is obtained by the denoiser of SeFMol. Next, $p_\theta\left(M_{t-1} \mid M_t, y, \mathcal{P}\right)$ is used iteratively to generate the next state. Guided by the specified properties, the newly generated state exhibits lower entropy than the previous one. Last, molecule M_0 is generated by iterative sampling of M_{t-1} for T steps.

Common properties evaluation of generated molecules

We sampled 100 molecules for each protein pocket in the test set, followed by a comprehensive performance evaluation of SeFMol and baseline models, focusing on docking performance, drug-like properties, and diversity. As presented in Table 1, SeFMol surpasses other baselines in all affinity-related metrics. Regarding the Vina score, SeFMol generates molecules with an average high affinity for protein pockets (−7.23 kcal/mol), representing a 13.7% improvement compared to reference molecules (−6.36 kcal/mol). Conversely, other methods face challenges in generating molecules with superior affinity relative to reference molecules.

The statistical test results at the pocket level are shown in table S1. SeFMol, Pocket2Mol, FLAG, TargetDiff, and MolCRAFT all achieved a hit pocket of 100, successfully generating molecules for 100 protein

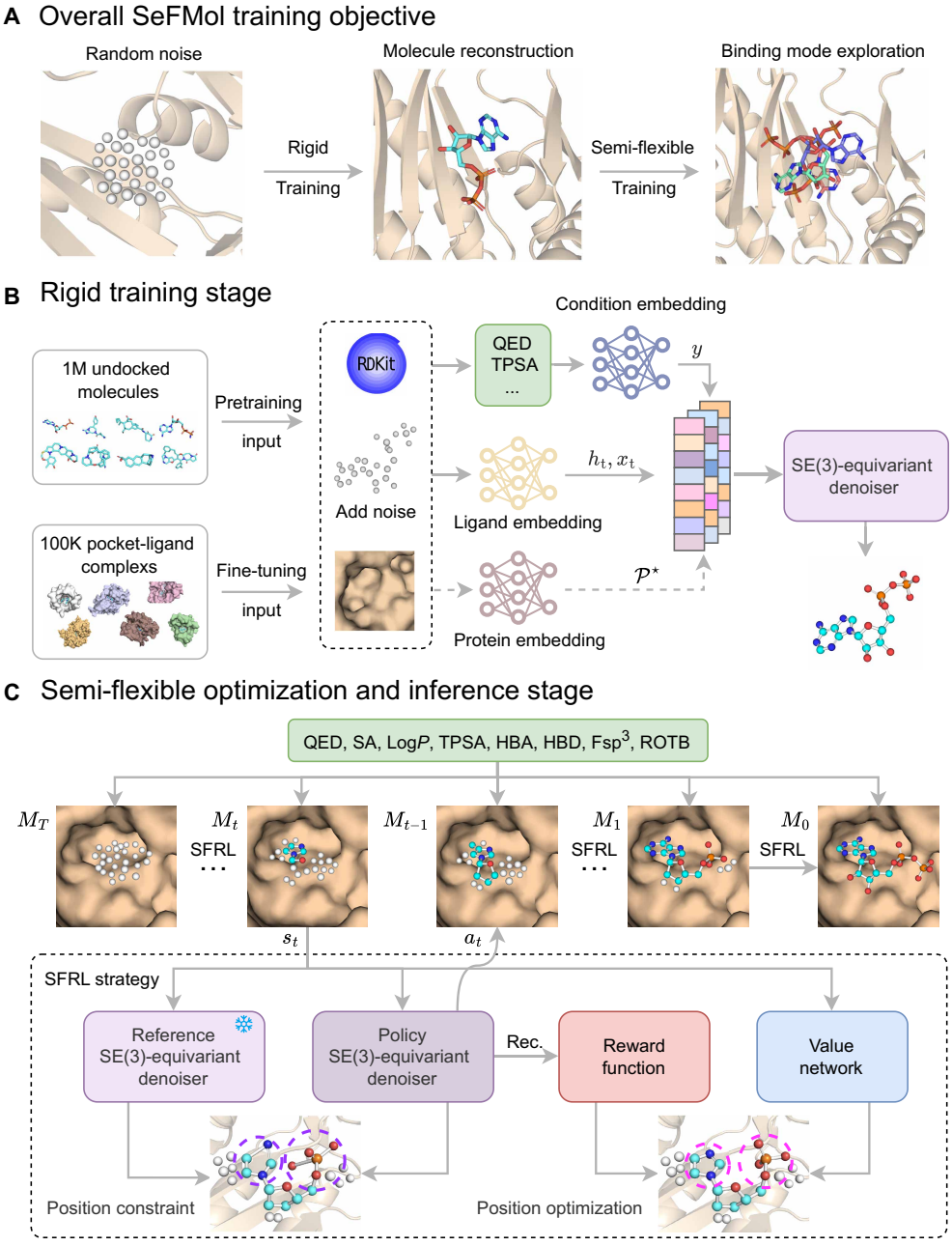

Fig. 1. Overview of the SeFMol framework. (**A**) Illustration of the overall training objectives of SeFMol. The rigid training phase, using static data for molecular reconstruction, focuses solely on static conformations. The semi-flexible phase uses RL to optimize a reward for exploring chemically plausible interaction patterns. (**B**) Two-stage rigid training strategy of SeFMol. At time step t in the diffusion process, small Gaussian and uniform noise are added to atomic coordinates and atom types via a fixed variance schedule. After obtaining embeddings x_t and h_t (and protein embedding $\mathcal{P}^\star$ during fine-tuning), they are fused with y and fed into the denoiser to reconstruct the original coordinates and atom types. (**C**) The semi-flexible optimization training and inference stage. This stage performs a complete denoising process from M_T to M_0. When using the semi-flexible optimization steered by RL (SFRL) strategy, this constitutes optimization training; otherwise, it is the inference phase.

pockets in the test set. Conversely, the other methods failed to generate molecules for some pockets. Regarding statistical significance, the one-sided paired Wilcoxon test (p_W) of SeFMol is 1.16×10^{-9}, the lowest among methods, and is better than MolCRAFT (3.13×10^{-4}) and IPDiff (4.35×10^{-6}). For the probability of superiority relative to the reference (PoS_{ref}), SeFMol attains 0.665, the highest among all approaches. Consistent with these results, fig. S1 indicates that the distribution of SeFMol's PoS_{ref} is strongly right-skewed, with the highest proportion falling within the 0.8 to 1.0 range. Points of SeFMol lie predominantly in the upper-left quadrant and exhibit a monotonic relationship whereby more negative values correspond to a higher superiority probability. For effect magnitude, the mean paired difference (MPD) is −0.700 kcal/mol and Hodges-Lehmann estimator (HLE) is 0.981 kcal/mol (95% confidence interval, −1.316 to −0.603), both

Table 1. Performance comparison on various metrics. ↓ indicates that smaller values are better, whereas ↑ indicates that larger values are better. Best results are highlighted in bold. Avg., average value; Med., median value.

Method	Vina score (↓)		Vina min (↓)		Vina dock (↓)		High affinity (↑)		QED (↑)		SA (↑)		Lipinski (↑)	Diversity (↑)
	Avg.	Med.	Avg.	Med.	Avg.	Med.	Avg.	Med.	Avg.	Med.	Avg.	Med.	Avg.	
Reference	−6.36	−6.46	−6.71	−6.49	−7.45	−7.26	–	–	0.48	0.47	0.73	0.74	4.27	–
AR	−5.75	−5.64	−6.18	−5.88	−6.75	−6.62	37.9%	31.0%	0.51	0.50	0.64	0.63	4.75	0.690
Pocket2Mol	−5.14	−4.70	−6.42	−5.82	−7.15	−6.79	48.4%	51.0%	0.57	0.58	0.76	0.76	4.88	0.685
ResGen	10.50	2.54	−2.94	−4.41	−6.59	−6.45	38.0%	25.0%	0.58	0.59	**0.78**	**0.79**	4.90	0.742
FLAG	45.98	36.62	6.17	−2.91	−5.24	−5.71	22.7%	7.0%	0.61	0.62	0.63	0.62	**4.98**	**0.766**
TargetDiff	−5.47	−6.30	−6.64	−6.83	−7.80	−7.91	58.1%	59.1%	0.48	0.48	0.58	0.58	4.51	0.708
DecompDiff	−5.67	−6.04	−7.04	−7.09	−8.39	−8.43	64.4%	71.0%	0.45	0.43	0.61	0.60	4.31	0.660
MolCRAFT	−6.59	−7.04	−7.27	−7.26	−7.92	−8.01	59.1%	62.6%	0.50	0.51	0.69	0.68	4.46	0.718
IPDiff	−6.66	−7.47	−7.64	−7.69	−8.49	−8.39	68.5%	72.2%	0.50	0.51	0.56	0.56	4.40	0.728
SeFMol	**−7.23**	**−7.70**	**−8.03**	**−8.00**	**−8.72**	**−8.75**	**68.7%**	**76.3%**	**0.63**	**0.64**	0.60	0.60	4.90	0.686

the most negative among methods. Notably, the HLE interval lies entirely below zero, indicating that under a robust location estimate, Vina scores of SeFMol-generated molecules are systematically lower than the reference molecules. Together, these results show that SeFMol outperforms existing methods under statistical tests.

We compared all molecules generated by each model in terms of Vina score and Vina dock (Fig. 2A). Molecules generated by SeFMol tend to exhibit higher affinity. In addition, the Vina score and Vina dock for molecules generated by SeFMol are closely aligned, with a correlation coefficient of 0.95, indicating the ability of SeFMol to directly generate conformations with chemically plausible interaction patterns. To explore the role of the SFRL strategy, we conducted an ablation experiment. Without the SFRL strategy, the correlation coefficient decreases by 4%, and the generated molecules are more concentrated in the low-affinity range. This comparison underscores the positive impact of the SFRL strategy in enhancing molecular docking performance. Figure 2B illustrates the rationality of interaction patterns for molecules generated by SeFMol. Higher ligand-binding efficacy (LBE) indicates greater contributions from each atomic pair interaction. In this metric, the LBE of SeFMol is only lower than Pocket2Mol, ResGen, and AR. Further analysis in Fig. 2C reveals that molecules generated by Pocket2Mol, ResGen, and AR have lower molecular weight, while SeFMol-generated molecules maintain the same weight as reference molecules. This indicates that SeFMol has the potential to enhance Vina energies and improve LBE without altering molecular weight, showcasing its advantages.

SeFMol also demonstrates excellence in evaluations related to drug-like properties. For quantitative estimation of drug-likeness (QED), SeFMol outperforms all other methods. Similarly, in the context of Lipinski's rule, SeFMol exhibits an advantage over other diffusion-based models. For synthetic accessibility (SA), SeFMol shows competitive performance among diffusion models. Notably, molecules generated by autoregressive methods tend to perform better, as they frequently generate molecules with fewer atoms and simpler structures. In addition, the diversity of generated molecules is critical for discovering potential candidates. The diversity of molecules generated by SeFMol is slightly lower than TargetDiff but higher than Pocket2Mol and DecompDiff. This result indicates that SeFMol effectively balances diversity with other performance metrics, meeting the requirements for molecular generation in the context of protein pockets. To further evaluate the rationality of the generated molecules, we selected two targets from the test set and visualized them together with the molecules achieving the highest Vina scores against them. As shown in Fig. 2 (D and E), the high-scoring molecules generated by SeFMol exhibit favorable drug-like properties and reasonable geometries.

Sampling efficiency and chemical rationality analysis of generated molecules

In this section, we analyze the sampling efficiency of SeFMol and the molecular properties of the generated molecules. To evaluate the sampling efficiency across different models, we conducted 100 molecular sampling operations per target in the test set, counted the total number of successfully reconstructed molecules, and calculated the average time required to generate each molecule. As shown in Fig. 3A, SeFMol outperforms the other methods, requiring, on average, 0.81 s per molecule with a generation rate of 98.3%, whereas MolCRAFT needs 1.41 s, and TargetDiff and DecompDiff require 34.28 and 61.89 s, respectively. This remarkable sampling performance is primarily attributed to our carefully designed variable fast sampling strategy, which notably enhances sampling efficiency without compromising the completion rate.

To further assess the impact of the proposed fast sampling strategy on model performance, we varied the number of denoising steps during both the SFRL training phase and the sampling phase. As shown in fig. S6, docking performance deteriorated as the number of steps increased. When the property guidance (PG) strategy was removed, Vina energies worsened across all time steps, indicating that excessively long trajectories exacerbate the instability caused by reward sparsity, whereas PG helps mitigate the optimization instability induced by sparse terminal rewards. For molecular properties, fig. S7 presents the joint QED-SA distribution of generated molecules across timestep settings. At 50 steps, the distribution concentrates in the reasonable region with QED $\geq$ 0.5 and SA $\geq$ 0.5, similar to the result at 100 steps. Further increases to 500 and 1000 produce a modest shift of QED and SA toward higher values. Longer sampling allows

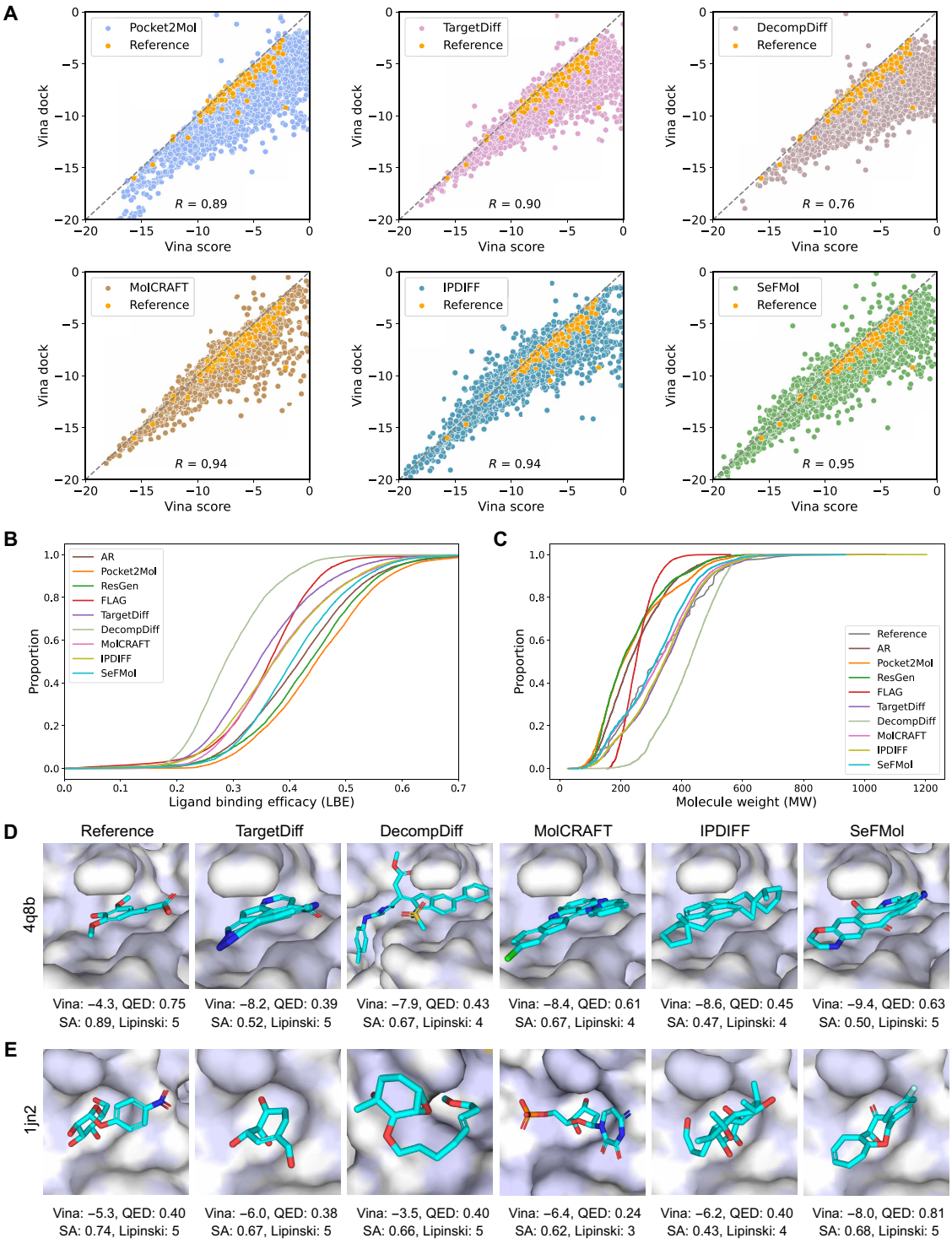

Fig. 2. Evaluation of common properties for generated molecules. (**A**) Distribution and correlation analysis of the Vina score and Vina dock for generated molecules. (**B**) Analysis of the relationship between LBE and the proportion of generated molecules. (**C**) Analysis of the relationship between molecular weight (MW) and proportion of generated molecules. (**D** and **E**) Comparison of the geometric structure, Vina score, and drug-like properties of generated molecules for the two protein pockets in the test set. QED, quantitative estimation of drug-likeness; SA, synthetic accessibility.

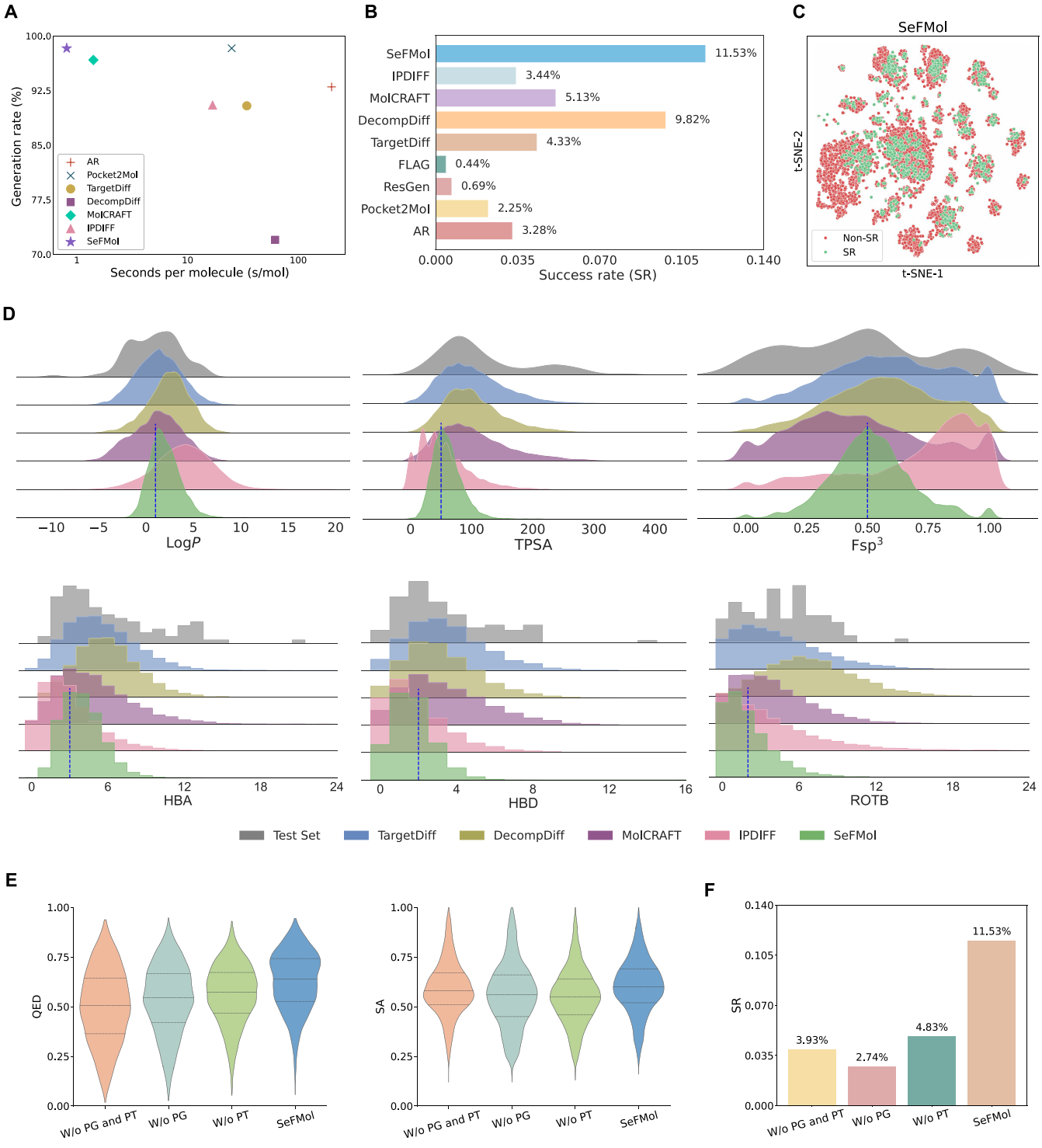

Fig. 3. Chemical rationality evaluation of generated molecules. (**A**) Sampling efficiency analysis including the time required to generate a molecule measured in seconds per molecule and the completion rate of generated molecules. (**B**) Comparison of SR of molecules generated by different methods. (**C**) Comparison of structural distributions between successful (SR) and nonsuccessful (Non-SR) molecules generated by SeFMol. (**D**) Comparison of property distributions for molecules generated by different methods and the reference molecules in the test set, including Log*P*, TPSA, Fsp^3, HBA, HBD, and ROTB. The blue dashed line denotes the properties prespecified by SeFMol. (**E**) Comparison of QED and SA for molecules generated by different settings of pretraining (PT) and PG. (**F**) Comparison of SR for molecules generated by different settings of PT and PG.

PG to remain active in later stages and brings limited gains in properties, but the improvement is small. Overall, 50 steps already yield a QED-SA distribution comparable to that with 500 or 1000 steps while reducing inference cost by 20 times relative to 1000 steps.

Figure 3B displays the success rate (SR) of different methods for generating molecules in the test set. SR is a crucial metric for rigorously evaluating the reasonableness of molecular properties and affinities, defined as the percentage of molecules that satisfy nine physicochemical property constraints, detailed in the "Evaluation" section. SeFMol achieved an SR of 11.53%, which is 8.09% higher than IPDIFF, demonstrating its superiority in generating molecules with reasonable chemical properties. Figure 3C presents a t-distributed stochastic neighbor embedding (t-SNE) visualization of the structural distribution of successful (SR) and nonsuccessful molecules. SR molecules exhibit a highly dispersed distribution, indicating rich diversity, whereas some failed samples cluster in the upper left area, suggesting shared intrinsic characteristics or similar structural defects. These observations imply that the PG strategy effectively directs generated molecules to regions with higher SRs. However, some molecules remain difficult to correct due to potential biases in the pretraining data.

In biology and chemistry, the functional behavior of molecules depends on a series of complex and highly refined mechanisms. Molecules need to achieve geometric complementarity with specific targets to establish efficient interaction patterns while also having specific drug-like properties necessary to trigger and execute their functions. Figure 3D illustrates the distribution of six key properties of molecules generated by SeFMol and baseline models in the test set, with dashed lines representing preset property conditions. The results demonstrate that molecules generated by SeFMol closely align with target values and exhibit low dispersion. This indicates that SeFMol achieves high accuracy in controlling molecular properties, generating molecules that meet the desired criteria. Other baseline methods, lacking property control mechanisms, sample molecules with uncontrollable and widely dispersed property distributions. In the critical physiological process of crossing the blood-brain barrier, this barrier serves as a protective shield for the central nervous system using strict screening mechanisms for substances attempting to pass through. To penetrate the blood-brain barrier and act on receptors within the central nervous system, the topological polar surface area (TPSA) of a molecule typically needs to be less than 90 $Å^2$ (*11*, *37*). Figure 3D shows that a substantial proportion of molecules generated by the baseline models have a TPSA greater than 90 $Å^2$, indicating limitations in controlling this critical property for blood-brain barrier penetration. In contrast, with an expected TPSA of 50 $Å^2$, SeFMol generates molecules with values almost entirely below 90 $Å^2$, concentrated around 50 $Å^2$.

Figure 3E presents the performance of SeFMol and its three settings in terms of the QED and SA metrics: without pretraining (w/o PT), without PG (w/o PG), and without both PT and PG (w/o PT & PG). Overall, SeFMol exhibits the best performance on both metrics, whereas w/o PT & PG yields the lowest QED, indicating that simultaneously removing these two strategies substantially degrades model performance. The QED of w/o PG and w/o PT falls between SeFMol and w/o PG & PT, indicating that PT and PG independently contribute to performance. However, w/o PT & PG performs better on SA than the models lacking only one of the two strategies, implying that PT and PG exhibit a synergistic effect. To further quantify this phenomenon, we conducted a statistical analysis of PG and PT (table S2). For SA, we obtained $\Delta_{PG} = -0.0393$ and $\Delta_{PT} = -0.0323$, indicating that retaining either component alone reduces SA, whereas retaining both strategies together yields a positive gain ($\Delta_{both} = 0.0121$). The resulting synergy magnitude is $\psi = 0.0836$, consistent with a both-or-none effect pattern. For QED, $\Delta_{PG} = 0.0607$ and $\Delta_{PT} = 0.0343$ show that retaining either component alone improves the metric, while retaining both further amplifies the gain ($\Delta_{both} = 0.1259$). We speculate that the chemical priors learned during PT constrain the generative distribution to a more feasible region, after which PG performs fine-grained optimization around the target properties on top of this prior. Accordingly, for SA, the synthesis-safety boundary provided by PT effectively prevents the overoptimization induced by PG. When PG or PT is used in isolation, one typically observes either boundary violations or insufficient optimization, leading to limited improvements in SA or even degradation. By contrast, QED is a weighted composite of multiple factors. PG and PT affect these factors in partly overlapping yet partly complementary ways. Consequently, using either strategy alone improves QED, while combining both yields the best performance. Figure 3F shows the SR of SeFMol under different settings. The SR of w/o PT and w/o PG are both lower than SeFMol but higher than w/o PG & PT, suggesting that PT and PG independently have a positive impact on SR, and their combined effect is superior to using them individually. In summary, both PT and PG play crucial roles in improving the performance of the model, and their combination yields better results.

We then visualized the distributions of common physicochemical properties in the pretraining dataset (Molecule3D) and fine-tuning dataset (CrossDocked2020). As shown in fig. S9, molecules in the pretraining dataset are generally smaller and contain fewer heavy atoms, and therefore often fail to fully occupy receptor pockets. By contrast, the fine-tuning dataset more frequently contains larger, more polar, and less drug-like molecules, resulting in a mismatch between the property prior learned during pretraining and the distribution encountered during finetuning. This mismatch causes the pretrained model, relative to the nonpretrained model, to slightly underestimate high-affinity ligands, thereby inducing a degree of negative transfer in terms of high-affinity metric (table S2). Nevertheless, the pretrained model still outperforms the nonpretrained model on the overall evaluation metrics, indicating that pretraining remains beneficial on balance.

Reliability and conformational analysis of generated molecules

In molecular generation, ideal molecules should have excellent drug-like properties, stable conformations, reasonable structural distribution, and reliable interaction patterns. We evaluated the reliability of SeFMol by comparing root mean square deviation (RMSD) between the molecules generated by each method and their redocked conformations using AutoDock Vina (*38*). As illustrated in Fig. 4A, the RMSD of SeFMol-generated molecules is closer to the reference molecules. Furthermore, Fig. 4B shows that SeFMol generates molecules with fewer clashes during the docking process compared to other methods, achieving highly competitive results. The success of SeFMol in modeling 3D conformations is primarily attributed to its SFRL strategy, which constructs reasonable conformations by dynamically exploring the optimal spatial positions of molecules within the binding pocket, thereby laying the foundation for generating high-quality molecules.

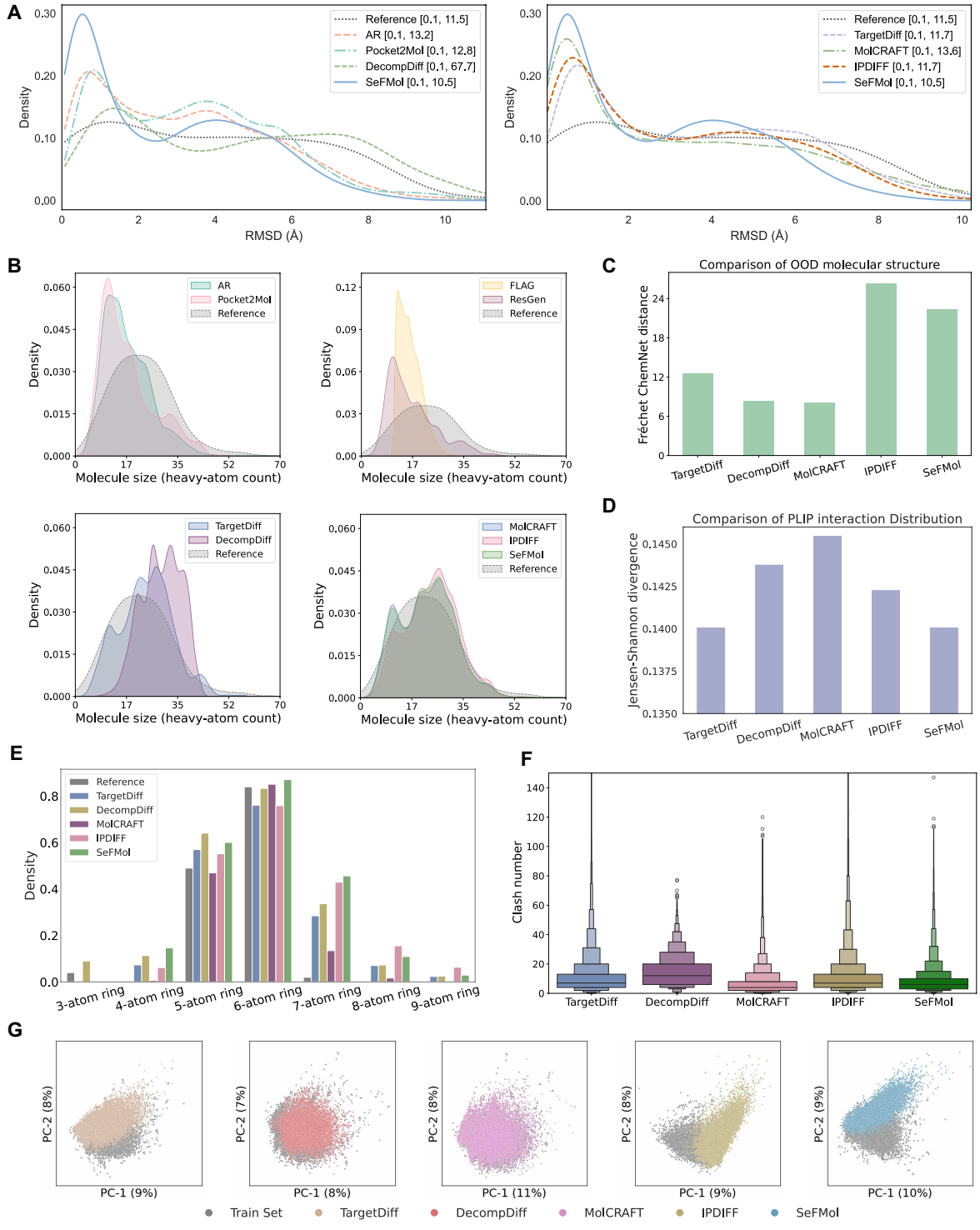

Fig. 4. Structure and conformation evaluation of generated molecules. (**A**) Distribution of RMSD of generated molecules relative to reference molecules. (**B**) Distribution of molecule size (heavy-atom count) of generated molecules and reference molecules in the test set. (**C**) FCD of structural distribution between generated molecules and molecules in the training set. (**D**) Jensen-Shannon divergence (JSD) of interaction type distribution between generated molecules and molecules in the test set. (**E**) Comparison of ring count distributions for molecules generated by different methods. (**F**) Comparison of the distributions of atomic clash counts between generated molecules and protein pockets across different methods. (**G**) Structure visualization via dimensionality reduction using PCA for molecules generated by different methods and the reference molecules in the test set.

To quantitatively assess the binding rationality of generated molecules, we compared the ring size and molecule size distributions of molecules in the test set with those of molecules generated by different methods. Regarding ring size distribution, Fig. 4C shows that molecules generated by SeFMol tend to form more five-atom and six-atom rings. These structures frequently form hydrogen bonds, making them crucial substructures in drug design. For larger rings, SeFMol generates fewer nine-atom and eight-atom rings compared to IPDIFF, but relatively more seven-atom rings due to their geometric similarity to six-atom rings, which the model struggles to distinguish. Overall, SeFMol produces a higher proportion of reasonable five-atom and six-atom rings, with fewer large rings. However, the PT and PG strategies introduce some less frequent ring structures from the training set. The molecule size distribution in Fig. 4D indicates that the sizes of the molecules generated by SeFMol closely match those in the training set, suggesting that their high affinity is not achieved by increasing molecular size but instead arises from a rational generative mechanism in the molecular design process.

In the SBDD task, it is crucial to preserve the interaction patterns between generated molecules and their protein targets (*23*). Accordingly, we examined whether SeFMol can learn hidden microscopic interaction patterns from 3D conformations. Using PLIP (*39*, *40*), we characterized seven types of interactions and computed the Jensen-Shannon divergence (JSD) for each pocket to quantify differences in the interaction-type distributions between the generated molecules and the reference molecules (*41*). As shown in Fig. 4E, SeFMol and TargetDiff both achieve the best performance with a JSD value of 0.1401. These results indicate that, although SeFMol optimizes binding affinity by adjusting molecular structures, the interaction types of its generated molecules remain closely matched to the reference molecules, thereby supporting the effectiveness of SeFMol in learning microscopic interaction patterns.

A common issue with distribution-based learning models is their exploration being limited to the training data distribution, resulting in generated molecules that resemble those in the training set (*42*). However, in practical applications, the design of lead compounds or scaffold molecules often requires structurally novel molecules to more effectively address key challenges in drug discovery, including enhancing drug activity, reducing side effects, and circumventing patented molecular scaffolds. As shown in Fig. 4F, SeFMol ranks second in Fréchet ChemNet distance (FCD), slightly behind IPDIFF. This result indicates that both SeFMol and IPDIFF effectively overcome the limitations of the training data distribution and generate molecules whose structures extend beyond the original distribution. To obtain a more intuitive understanding of structural distributional differences, we applied dimensionality reduction and principal components analysis (PCA) (*43*) to visualize the generated molecules. As shown in Fig. 4G, the molecules generated by SeFMol and IPDIFF exhibit distributions that differ markedly from the training set. Furthermore, SeFMol surpasses IPDIFF in PCA scores. We speculate that this is primarily because both models incorporate interaction priors that bias them toward exploring high-affinity structures, thereby inducing a shift in the structural distribution.

We further analyzed the relationship between the diversity and FCD of the generated molecules. As shown in fig. S3, removing the PT, PG, and SFRL strategies from the model increases molecular diversity, while novelty measured by FCD decreases at the same time. This indicates that when PG and SFRL are used to improve affinity and molecular properties, the generation policy tends to exploit a small neighborhood of high-scoring structures, thereby sacrificing exploration of a broader chemical space and directly reducing overall diversity. Meanwhile, this strategy helps the model overcome the distributional constraints of the training data and enhances the novelty of the generated molecules.

SeFMol customizes molecules for real-world therapeutic targets

We used SeFMol for molecular generation targeting real-world therapeutic proteins, assessing its capability to capture interaction patterns and effectively control the properties of the generated molecules. Two real-world therapeutic targets, cyclin-dependent kinase 2 (CDK2; PDB ID: 1H00) and rho-associated protein kinase 1 (ROCK1; PDB ID: 6E9W), were selected from the Protein Data Bank (PDB) (*44*). These proteins are pivotal in processes such as cell-cycle regulation, cytoskeletal remodeling, and tumorigenesis, making them highly relevant in drug development. We used PLIP to characterize the protein-ligand interactions between these targets and the generated molecules. Specifically, we used the original version of SeFMol (with its prespecified properties detailed in Materials and Methods), as well as versions with prespecified TPSA of 110 and Fsp^3 of 1, to generate 100 ligands for each target. These ligands were then screened on the basis of SR to identify qualified molecules.

As shown in Fig. 5 (A and B), the results demonstrate that the molecules generated by SeFMol exhibit an improvement in Vina score, indicating stronger binding to the targets and the potential to enhance drug efficacy. Notably, these generated molecules reproduce the interaction patterns of some experimentally active molecules, ensuring consistency with known effective patterns while also exploring entirely new interaction patterns. These novel interactions, such as halogen bonds with Ala^{215} and salt bridges with Asp^{216} in ROCK1, have not appeared in the active molecule but are rational. In terms of drug property evaluation, SeFMol performs remarkably well. The generated molecules have reasonable properties, effectively balancing key metrics of drug developability while enhancing molecular affinity for the target, achieving a balance between affinity and properties.

We then analyzed the ability of SeFMol to control the properties of generated molecules. Figure 5 (C to F) visually depicts the property distributions of molecules generated using different pre-specified Fsp^3 and TPSA values. The property distributions of SeFMol-generated molecules fluctuate around the specified values, exhibiting an approximately normal distribution trend. Last, we validated the effectiveness of the SFRL strategy on real-world targets. From the affinity correlation analysis in Fig. 5 (G and H), it is clear that molecules generated using SFRL have a high-affinity correlation. These observations demonstrate that the SFRL strategy successfully learns interaction principles based on geometric structures, which is crucial for enhancing the affinity between generated molecules and targets. For more case studies on real-world targets, please refer to figs. S11 to S14.

In addition, we used SeFMol to generate molecules for protein structures predicted by AlphaFold (*3*). These structures differ from pockets derived from crystal structures in both conformational accuracy and pocket geometry. Figure S10 presents the distribution of Vina scores obtained by SeFMol across multiple protein targets. The Vina score distribution of SeFMol is consistently shifted toward more negative values, with a more favorable central tendency, demonstrating the ability of SeFMol to generate well-performing ligands for the predicted protein structures.

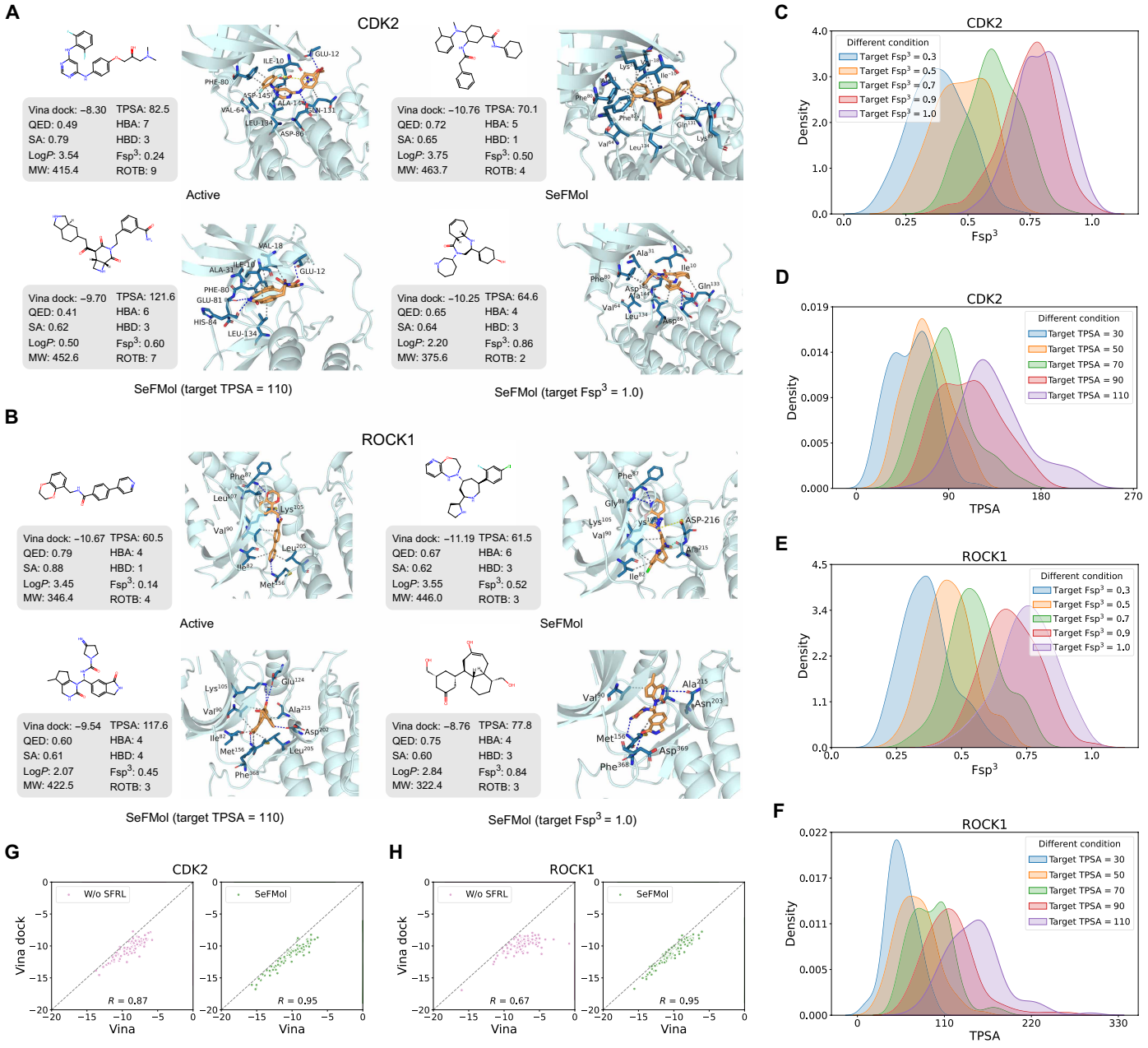

Fig. 5. Evaluation of generated molecules targeting CDK2 and ROCK1. (**A** and **B**) Analysis of protein-ligand interactions and molecular properties for active molecules and SeFMol-generated molecules under different conditions targeting CDK2 and ROCK1. (**C** to **F**) Comparison of the Fsp^3 and TPSA distributions of 100 SeFMol-generated molecules targeting CDK2 and ROCK1, with all other properties held constant and only Fsp^3 or TPSA specified to vary. (**G** and **H**) Vina energy correlation analysis for 100 molecules generated with and without the SFRL strategy targeting CDK2 and ROCK1.

DISCUSSION

In structure-based drug design, understanding the dynamic interaction patterns resulting from ligand conformational changes is crucial for elucidating the nonstatic interactions between proteins and ligands. This understanding can lead to optimized affinity and selectivity of molecules. In this study, we introduced SeFMol, a semi-flexible molecular diffusion model designed to efficiently generate molecules for protein pockets. We conceptualized the denoising process of the pretrained diffusion model as a Markov decision process using gradient descent to dynamically optimize molecular denoising. Notably, at each step of the denoising process, guidance through property conditions mitigates the generation of unreasonable molecules due to sparse rewards. This ultimately results in molecules with both high affinity and desirable drug-like properties during the exploration phase. In addition, we implemented a fast sampling strategy that notably reduces the number of time steps in the diffusion model, thereby enhancing the efficiency of both training and sampling.

SeFMol outperforms baseline methods in generating molecules with high affinity, desirable drug-like properties, and structural novelty. It demonstrates excellent docking performance, achieving superior Vina scores through a RL-steered semi-flexible optimization strategy. SeFMol enhances binding energies without altering molecular size, ensuring that the molecular weight of generated molecules remains similar to reference molecules. SeFMol also samples molecules with excellent drug-like properties, achieving the highest QED, performing well in Lipinski, and maintaining competitive SA. With a rapid sampling strategy, the inference speed of SeFMol exceeds that of other methods, achieving a completion rate of 98.3% and ranking first. SeFMol effectively controls molecular properties, ensuring that generated molecules meet expected values. The conformations generated by SeFMol are stable, exhibiting reliable interaction patterns, superior RMSD performance, fewer atomic clashes, and the ability to form reasonable cyclic structures. SeFMol effectively learns micro-interaction patterns, retaining interaction types while optimizing binding affinity. It is capable of exploring novel structures that deviate from training data, which is beneficial for developing lead molecules with enhanced activity and reduced side effects. In case studies targeting CDK2 and ROCK1, molecules generated by SeFMol demonstrated improved affinity and enhanced drug efficacy. SeFMol reproduced known interaction patterns and explored new interactions, confirming its ability to balance affinity and drug-likeness properties. Although the proposed strategies exhibit strong exploratory behavior, SeFMol trades some diversity for higher binding energies and improved physicochemical properties. This diversity loss stems largely from distributional shift and exploitation bias introduced by the PG, PT, and SFRL strategies. In the future, we aim to better balance the exploration-exploitation trade-off through multiobjective reward design and adaptive exploration.

In summary, as a novel molecular generation tool, SeFMol demonstrates substantial advantages. It achieves notable improvements in affinity, drug-likeness properties, and stable conformations while effectively addressing challenges related to sampling efficiency and chemical rationality. Its capability to explore beyond training data and generate novel structures positions it as a highly promising tool in drug discovery and development.

MATERIALS AND METHODS

Dataset usage

We used the Molecule3D dataset (*35*) in the pretraining stage. This dataset was curated from ground-state 3D geometries in PubChemQC (*45*) and reformatted for deep-learning models. We sampled 1,000,000 molecules with valid 3D conformations from Molecule3D and computed eight molecular properties with RDKit for PG, including QED, SA, Log*P*, TPSA, HBA, HBD, Fsp^3, and ROTB.

In the fine-tuning and optimization stages, we constructed the training and test sets using the CrossDocked2020 dataset (*18*). The same data preprocessing and splitting as (*21*) are used to refine 22.5 million docking complexes into high-quality docking poses (with RMSD < 1 Å between docking poses and ground truth) and diverse proteins (sequence identity < 30%). Ultimately, 100,000 protein-ligand pairs are employed for training and 100 proteins for testing.

Problem formulation

In the SBDD task, a protein pocket is typically represented as a set of N_P atoms, denoted as $\mathcal{P} = \left\{\left(x_P^{(i)}, h_P^{(i)}\right)\right\}_{i=1}^{N_P}$, where $x_P^{(i)} \in \mathbb{R}^3$ represents the atom position and $h_P^{(i)} \in \mathbb{R}^d$ denotes the atom type. Similarly, a ligand molecule is represented as a set of N_M atoms, $\mathcal{M} = \left\{\left(x_M^{(i)}, h_M^{(i)}\right)\right\}_{i=1}^{N_M}$. The features h are invariant under group transformations. The positions x in 3D space are SE(3)-equivariant, meaning that if the molecule undergoes a rotation or translation, the output will transform correspondingly. For simplicity, we represent the ligand molecule as $M = [x, h]$, where $x \in \mathbb{R}^{N_M \times 3}$ and $h \in \mathbb{R}^{N_M \times d}$. The objective is to generate ligand molecules that can effectively bind to this protein. With the advancement of diffusion models in generative tasks, these models have also demonstrated strong performance in the SBDD task. In molecular diffusion models, the molecular atom types and coordinates are represented by Gaussian and discrete distributions, respectively. This models the molecule as a product of atom coordinate distribution and atom type distribution while maintaining SE(3)-equivariance of the generation process.

SeFMol architechture

In this section, we will elaborate on our proposed model SeFMol: a semi-flexible molecular diffusion model. Unlike other diffusion-based methods in SBDD (*29*, *30*, *46*), our objective is to model $p(M \mid \mathcal{P}, y)$, treating the denoising process as an MDP. Within this framework, we incorporate RL to dynamically adjust molecular positions during the denoising process, guided by conditioning signals on molecular properties. Specifically, we first construct a molecular diffusion model conditioned on multiple molecular properties and pretrain it using a large dataset of undocked molecules. Subsequently, we introduce protein targets as conditions, resulting in a diffusion model conditioned on both molecular properties and protein targets, and fine-tune it on a docked molecular dataset. Last, we use RL to simulate the protein-ligand interaction binding process, optimizing the denoising process of the trained diffusion model to generate molecules with both high affinity and desirable drug-like properties. The subsequent sections will detail the components, training strategies, and sampling process of SeFMol.

Property-guided molecular diffusion process

We select multiple physicochemical properties as conditions, including QED (y_1), SA (y_2), Log*P* (y_3), TPSA (y_4), HBA (y_5), HBD (y_6), Fsp^3 (y_7), and ROTB (y_8). These properties relate to the intrinsic properties and binding affinity of molecules. For a detailed introduction to these properties, please refer to the "Evaluation" section. These properties are embedded by a multilayer perceptron (MLP), defined as follows

$$y = \mathrm{MLP}([y_1, y_2, \ldots, y_8]) \tag{1}$$

The default value of y is set as [1.0, 1.0, 1.0, 50.0, 3.0, 2.0, 0.5, 2.0].

At each time step t in the diffusion process of pretraining and fine-tuning phases, small Gaussian noise is added to the atom coordinates and uniform noise to the atom types based on a Markov chain with a fixed variance sequence $\beta_1, \ldots, \beta_T$

$$q(M_t \mid M_{t-1}, \mathcal{P}^\star) = \prod_{i=1}^{N_M} \mathcal{N}\left(x_{t,i}^M; \sqrt{1-\beta_t}\, x_{t-1,i}^M, \beta_t I\right) \cdot C\left[h_{t,i}^M \mid (1-\beta_t) h_{t-1,i}^M + \beta_t / K\right] \tag{2}$$

where $\mathcal{P}^\star$ represents proteins as optional input, which are not used during the pretraining phase and are incorporated during the fine-tuning phase.

In the forward process, the noise is injected into the molecules without property information, thereby preserving their integrity. The posterior distribution incorporates the property embedding y, ensuring that the generated molecules align with the noise characteristics while achieving the desired drug-like properties. The posterior is defined as follows

$$q\left(M_{t-1} \mid M_t, M_0, y, \mathcal{P}^{\star}\right) = \prod_{i=1}^{N_M} \mathcal{N}\left[x_{t-1,i}^{M}; \tilde{\mu}\left(x_{t,i}^{M}, x_{0,i}^{M}\right), \tilde{\beta}_t I\right] \cdot C\left[h_{t-1,i}^{M} \mid \tilde{c}\left(\left[h_{t,i}^{M}, y\right], \left[h_{0,i}^{M}, y\right]\right)\right] \quad (3)$$

where $\tilde{\mu}\left(x_{t,i}^{M}, x_{0,i}^{M}\right) = \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t} x_{0,i}^{M} + \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t} x_{t,i}^{M}$, $\tilde{\beta}_t = \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\beta_t$, $\alpha_t = 1-\beta_t$, $\bar{\alpha}_t = \prod_{i=1}^{t} \alpha_i$, $\tilde{c}\left(h_{t,i}, h_{0,i}\right) = \frac{c^*}{\sum_{k=1}^{K} c_k^*}$, and $c^*\left(h_{t,i}, h_{0,i}\right) = \left[\alpha_t\left[h_{t,i}, y\right] + \left(1-\alpha_t\right)/K\right] \odot \left[\bar{\alpha}_{t-1}\left[h_{0,i}, y\right] + \left(1-\bar{\alpha}_{t-1}\right)/K\right]$. For brevity, let $[\cdot, \cdot]$ denote the concatenation operator, and this definition is also used subsequently.

Fast generation process

In diffusion models, the noise schedule is a predefined sequence that defines the rate at which information is destroyed and reconstructed, affecting the training and sampling performance of the model. Taking atom positions as an example, we follow previous molecular diffusion models (*29*, *30*) and apply a sigmoid schedule to add noise to the coordinates. This provides smoother changes at the beginning and end of the process and steeper transitions in the middle, which is particularly useful for preserving important features in the early and late stages. In contrast, we use a variable beta sequence between the training and sampling processes to improve sampling efficiency. This is achieved by the SE(3)-equivariant GNNs, which effectively capture the symmetry and geometric features of 3D molecular structures, allowing the model to denoise effectively even with a small number of time steps. β_t is defined as follows

$$\beta_t = \frac{\beta_{T^*} - \beta_1}{1 + e^{-\left(a + \frac{bt}{T^*-1}\right)}} + \beta_1 \quad (4)$$

where a is the starting point of the sigmoid function, and b defines its range. β_1 and β_{T^*} are the initial and final β values, respectively. T^* represents the variable number of time steps. During the training process, $T = 1000$ allows noise to be smoothly added to the data, helping the model learn gradual changes in the data distribution. During sampling, setting $T = 50$ to reconstruct β_t means each denoising step is more intense, avoiding potential detail loss or error accumulation with more steps. This increases sampling efficiency by 20 times without compromising the quality of generated molecules. t denotes the current timestep, ranging from 0 to $T^* - 1$.

In the reverse process, atom types are first embedded and then concatenated with the property embedding y to obtain molecular invariant features that contain property information. We then recover the ground truth molecule M_0 from the initial noise M_T and y, approximating the distribution through a neural network

$$p_\theta\left(M_{t-1} \mid M_t, y, \mathcal{P}\right) = \mathcal{N}\left[x_{t-1}; \mu_\theta\left(\left[x_t, y, h_t\right], t, \mathcal{P}\right), \sigma_t^2 I\right] \cdot C\left[h_{t-1} \mid c_\theta\left(\left[x_t, y, h_t\right], t, \mathcal{P}\right)\right] \quad (5)$$

The molecular features and protein features are then input into a denoiser ϕ_θ to predict $\left[x_0, a_0\right]$, defined as follows

$$\left[\hat{x}_0, \hat{h}_0\right] = \phi_\theta\left(M_t, y, t, \mathcal{P}^{\star}\right) = \phi_\theta\left(x_t, \left[h_t, y\right], t, \mathcal{P}^{\star}\right) \quad (6)$$

where ϕ_θ is a denoiser constructed with an SE(3)-equivariant GNN. During the pretraining phase, proteins are not used, allowing the denoiser to learn latent associations between molecules and properties. In the fine-tuning phase, proteins are included, enabling the denoiser to model interactions between ligand atoms and protein atoms.

Semi-flexible denoising steered by RL

We introduce RL to simulate the dynamic changes of ligands during the binding process and to optimize the denoising process of the rigid diffusion model, thereby generating molecules with both high affinity and favorable drug-like properties. Specifically, we formulate the denoising process as an MDP, in which the spatial position of ligand is progressively adjusted at each timestep of generation to better match plausible binding modes, while the diffusion model parameters are iteratively updated via policy gradients. Furthermore, by specifying target properties for generation, we impose an explicit bias on the generated intermediate molecules throughout the process, thereby reducing invalid molecule generation and blind exploration caused by sparse rewards. The overall optimization pipeline consists of PG, embedding modules, policy network (denoiser), value network, reward function, and a KL regularization term that constrains the magnitude of parameter updates.

The denoising process of the molecular diffusion model can be formulated as an MDP

$$s_t = \left(M_{T-t}, y, \mathcal{P}\right), \quad a_t = M_{T-t-1}$$

$$\rho_0\left(s_0\right) = \left(\left[\varepsilon_x, \varepsilon_h\right], y, \mathcal{P}\right)$$

$$\rho\left(s_{t+1} \mid s_t, a_t\right) = \left(\delta_{M_{T-1-t}}, \delta_y, \delta_{\mathcal{P}}\right)$$

$$\pi_\theta\left(a_t \mid s_t\right) = p_\theta\left(M_{T-t-1} \mid M_{T-t}, y, \mathcal{P}\right)$$

$$R\left(s_t, a_t\right) = \begin{cases} r\left(s_{t+1}\right) = r\left(M_0, \mathcal{P}\right) & \text{if } t = T-1, \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

where s_t represents the state, a_t is the action, ρ_0 and ρ are the initial state distribution and the dynamics, $\left[\varepsilon_x, \varepsilon_h\right]$ are the initial noise of atom positions and types, δ denotes the Dirac delta distribution, π_θ is the parameterized policy network, and $R\left(s_t, a_t\right)$ is the reward function.

The state s_t is defined as the triplet $\left(M_{T-t}, y, \mathcal{P}\right)$ comprising the current molecular state, protein pocket $\mathcal{P}$, and target property y. This design ensures that the parameterized policy network π_θ can account simultaneously for binding modes and property requirements, generating molecules that fit binding pockets and satisfy drug-like property criteria. The representation is naturally compatible with the denoising logic of diffusion models. The denoising process of diffusion models is progressive in nature, and s_t corresponds precisely to an intermediate step in that process, perfectly aligning with the diffusion paradigm of progressively denoising from an initially

noisy state. The initial state s_0 is sampled from atomic position noise and atomic type noise $[\varepsilon_x, \varepsilon_h]$, corresponding to the starting point of the diffusion model.

To align with the state design, we define the action a_t to coincide with a single denoising step. Given state s_t, the policy π_θ outputs a_t, which directly yields the next molecular state along the denoising trajectory (M_{T-t-1}). The action is the direct output of a denoising step, enabling tight integration of RL with the diffusion framework without auxiliary adapter layers and allowing RL to optimize the denoising process directly. At the level of state distributions and transitions, since a_t explicitly specifies the next molecular state, the transition is strictly deterministic. Once a_t is produced, the next state s_{t+1} is uniquely determined (M_{T-t-1} becomes the new current molecular state). Consequently, $\rho(s_{t+1} \mid s_t, a_t)$ is characterized by a Dirac delta distribution δ. This property simplifies the complexity of the MDP and makes the solution process more tractable.

We adopt a standard terminal reward (*47*, *48*) for $R(s_t, a_t)$, computing the reward only at the final denoising step ($t = T - 1$) from the generated molecule M_0. For all intermediate steps ($t < T - 1$), the reward is set to zero. However, this choice induces a sparse reward problem, meaning the model receives little informative feedback at most steps. We therefore introduce additional molecular property priors through specifically designed pretraining conditions to guide the policy network. This narrows the exploration space during generation and biases sampling toward high-affinity structures, thereby alleviating sparsity. In addition, we introduce a KL divergence to constrain the discrepancy between the optimized model and the pretrained model, which not only controls the magnitude of policy updates but also helps prevent the model from overfitting to the reward and forgetting the knowledge acquired during pretraining (*48*, *49*).

Building on the MDP defined above, we present a principled integration of diffusion models and RL. Given a specified protein pocket and properties, the system maximizes the reward associated with the generated molecules, thereby optimizing the molecular diffusion model

$$\min_\theta \mathbb{E}_{p(y,\mathcal{P})} \mathbb{E}_{p_\theta(M_{t-1}|M_t,y,\mathcal{P})} \left[-r(M_{t-1}, \mathcal{P}) \right] \tag{8}$$

where $\mathbb{E}_{p(y,\mathcal{P})}$ represents the expectation over the joint distribution of the desired properties and proteins, $\mathbb{E}_{p_\theta(M_{t-1}|M_t,y,\mathcal{P})}$ represents the expectation over the distribution of generated molecules under given conditions.

Training objective

The training objectives of SeFMol are divided into two parts: rigid training and SFRL optimization. In the rigid training stage, the model is expected to learn the mapping between molecular properties, proteins, and molecular structures. Thus, the training objective is to maximize the reconstruction of input molecules conditioned on molecular properties and proteins

$$\mathcal{L}_{t-1}^{(x)} = \|x_0 - \hat{x}_0\|^2 + C \tag{9}$$

$$\mathcal{L}_{t-1}^{(h)} = \sum_k c(h_t, h_0)_k \log \frac{c(h_t, h_0)_k}{c(h_t, \hat{h}_0)_k} \tag{10}$$

where $\mathcal{L}_{t-1}^{(x)}$ is the atomic coordinates loss, c is a constant term, $\mathcal{L}_{t-1}^{(h)}$ is the reconstruction loss for the atomic types, and $\hat{x}_0$ and $\hat{h}_0$ are predicted by the denoiser. The final molecular reconstruction loss is a weighted sum of two losses: $\mathcal{L}^{\text{rec}} = \mathcal{L}_{t-1}^{(x)} + \lambda \mathcal{L}_{t-1}^{(h)}$.

The training objective of the SFRL stage is to maximize the reward. We use a method similar to proximal policy optimization (*50*) for gradient clipping in policy gradient updates to constrain p_θ and $p_{\theta_{\text{old}}}$. KL regularization is introduced to prevent overfitting the reward and forgetting the pretrained knowledge. In addition, a value function is incorporated to reduce the variance of the gradient estimation. The training loss in the SFRL stage is as follows

$$\mathcal{L}_{t-1}^{\text{adv}} = r(M_0, \mathcal{P}) - \hat{V}(M_{t-1}, \mathcal{P}) \tag{11}$$

$$\mathcal{L}_{t-1}^{\text{clip}} = \nabla_\theta \text{clip}(\tau_t, 1-\epsilon, 1+\epsilon) \tag{12}$$

$$\mathcal{L}_{t-1}^{\text{KL}} = \text{KL}\left[p_\theta(M_{t-1} \mid M_t, y, \mathcal{P}) \,\|\, p_{\text{pre}}(M_{t-1} \mid M_t, y, \mathcal{P}) \right] \tag{13}$$

$$\mathcal{L}^{\text{RL}} = -\gamma \mathcal{L}^{\text{adv}} \cdot \mathcal{L}^{\text{clip}} + \eta \mathcal{L}^{\text{KL}} \tag{14}$$

where γ and η are the weights of the reward term and the KL regularization term, $\tau_t = \frac{p_\theta(M_{t-1}|M_t,y,\mathcal{P})}{p_{\theta_{\text{old}}}(M_{t-1}|M_t,y,\mathcal{P})}$ represents the importance sampling ratio, $\text{clip}(\tau_t, 1-\epsilon, 1+\epsilon)$ constrains the range of the ratio to ensure stability, and ϵ is the clip hyperparameter.

Baselines

We compared our model with eight advanced SBMG models. AR (*21*), Pocket2Mol (*22*), ResGen (*23*), and FLAG (*24*) are autoregressive generative models, while TargetDiff (*29*), DecompDiff (*30*), MolCRAFT (*32*), and IPDiff (*31*) are one-shot generative models.

Evaluation

Binding affinity

The mean and median of binding affinity-related metrics are calculated using AutoDock Vina (*38*). Vina score directly estimates the binding affinity of the generated molecules with their respective proteins; Vina min estimates the binding affinity after local structure minimization; Vina dock reflects the best possible binding affinity through a redocking process. If the Vina score is close to Vina min and Vina dock, then the generated poses are highly favorable and effectively capture the 3D interactions. High affinity measures the ratio at which generated molecules exhibit better Vina dock than reference molecules for each test protein. LBE represents the contribution of individual atoms in a molecule to Vina dock, eliminating the influence of molecular size (*23*).

Molecular properties

We evaluate the molecular properties from multiple perspectives. QED is the quantitative estimation of drug-likeness; SA is the synthetic accessibility score; Lipinski measures how many rules the drug follows the Lipinski's rule (*51*, *52*); Log*P* represents the octanol-water partition coefficient, and the Log*P* values of drug candidates typically range from −0.4 to 5.6 (*53*); TPSA is the topological polar surface area, and its value is typically less than 90 Å^2 (*11*, *37*). HBA and

HBD represent the number of hydrogen bond acceptors (no more than 10) and donors (no more than 5) (*51*). ROTB means the number of rotatable bonds (no more than 10). Fsp^3 is the average saturation calculated as the number of sp3 hybridized carbons divided by the total number of carbons, and its value is typically greater than 0.47 (*54*). Molecular weight: The sum of the atomic weights of all atoms in a molecule. To rigorously evaluate the reasonableness of the properties and affinities of generated molecules, we introduce the SR metric. This metric is defined as the proportion of molecules that meet the following nine conditions: Vina dock < -8.18, QED > 0.25, SA > 0.59, $-0.4 \leq \text{Log}P \leq -5.6$, TPSA ≤ 140, $Fsp^3 \geq 0.42$, HBA ≤ 10, HBD ≤ 5, and ROTB ≤ 10.

Statistical test

We conduct statistical tests using five metrics to assess the generative capability. Hit pocket is the number of protein targets in the test set for which the model successfully generates molecules. The test set contains 100 protein pockets in total. One-sided paired Wilcoxon signed-rank test (p_W): A nonparametric test suitable for paired comparisons without normality assumptions (*55*). For each pocket, we compute the Vina score difference between generated and reference molecules and then test whether these differences are less than zero. Smaller values of (p_W) indicate that the observed improvements are unlikely to have arisen by chance. In probability of superiority (PoS_{ref}), analogous to the common language effect size (*56*), we compute, for each pocket, the proportion of generated molecules that achieve a lower Vina score than the reference molecule and then average this proportion over all pockets. The metric ranges from 0 to 1, and a value above 0.5 indicates that the generated molecules outperform the standard molecule on average. In MPD, following general principles of effect size analysis (*57*), we compute, for each pocket, the difference between the mean Vina score of the generated molecules and that of the reference molecule and then average these differences over all pockets to obtain the mean effect size (with more negative values indicating improvement). In HLE, we take the median of all paired differences across pockets, which provides a robust effect-size estimate that is less sensitive to outliers (*58*). For HLE, 95% confidence interval is obtained by bootstrapping to quantify estimation uncertainty.

Overall

The RMSD between the generated ligand atoms and the Vina redocking conformations measures binding mode consistency, where conformations with an RMSD below 2 Å are generally considered to be chemically meaningful (*59*–*61*). FCD is the distance between the training set and the generated molecule set, calculated based on the penultimate layer activations of ChemNet, used to measure the structural differences between the generated molecules and the training set (*42*, *62*).

Sample efficiency

The average sampling time and generation success are used to evaluate the sampling efficiency. Time refers to the average time taken to generate 100 molecules for each protein in the test set. Generation success is the ratio of valid and complete molecules to the predetermined sample number.

Supplementary Materials

This PDF file includes:

Sections S1 to S14
Tables S1 to S4
Figs. S1 to S14

REFERENCES

1. E. Nogales, The development of cryo-em into a mainstream structural biology technique. *Nat. Methods* **13**, 24–27 (2016).
2. X. Li, P. Mooney, S. Zheng, C. R. Booth, M. B. Braunfeld, S. Gubbens, D. A. Agard, Y. Cheng, Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-em. *Nat. Methods* **10**, 584–590 (2013).
3. J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, Highly accurate protein structure prediction with alphafold. *Nature* **596**, 583–589 (2021).
4. B. K. Shoichet, Virtual screening of chemical libraries. *Nature* **432**, 862–865 (2004).
5. A. W. Thorman, J. Reigle, S. Chutipongtanate, J. Yang, B. Shamsaei, M. Pilarczyk, M. Fazel-Najafabadi, R. Adamczak, M. Kouril, S. Bhatnagar, S. Hummel, W. Niu, A. L. Morrow, M. F. Czyzyk-Krzeska, R. McCullumsmith, W. Seibel, N. Nassar, Y. Zheng, D. A. Hildeman, M. Medvedovic, A. B. Herr, J. Meller, Accelerating drug discovery and repurposing by combining transcriptional signature connectivity with docking. *Sci. Adv.* **10**, eadj3010 (2024).
6. J. Lyu, N. Kapolka, R. Gumpper, A. Alon, L. Wang, M. K. Jain, X. Barros-Álvarez, K. Sakamoto, Y. Kim, J. DiBerto, K. Kim, I. S. Glenn, T. A. Tummino, S. Huang, J. J. Irwin, O. O. Tarkhanova, Y. Moroz, G. Skiniotis, A. C. Kruse, B. K. Shoichet, B. L. Roth, Alphafold2 structures guide prospective ligand discovery. *Science* **384**, eadn6354 (2024).
7. M. Goel, R. Aggarwal, B. Sridharan, P. K. Pal, U. D. Priyakumar, Efficient and enhanced sampling of drug-like chemical space for virtual screening and molecular design using modern machine learning methods. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **13**, e1637 (2023).
8. Y. Luo, K. Yan, S. Ji, Graphdf: A discrete flow model for molecular graph generation. *International conference on machine learning* (PMLR, 2021), pp. 7192–7203.
9. B. Qiang, Y. Zhou, Y. Ding, N. Liu, S. Song, L. Zhang, B. Huang, Z. Liu, Bridging the gap between chemical reaction pretraining and conditional molecule generation with a unified model. *Nat. Mach. Intell.* **5**, 1476–1485 (2023).
10. Z. Wang, Y. Chen, P. Ma, Z. Yu, J. Wang, Y. Liu, X. Ye, T. Sakurai, X. Zeng, Image-based generation for molecule design with sketchmol. *Nat. Mach. Intell.* **7**, 244–255 (2025).
11. V. Bagal, R. Aggarwal, P. Vinod, U. D. Priyakumar, Molgpt: Molecular generation using a transformer-decoder model. *J. Chem. Inf. Model.* **62**, 2064–2076 (2021).
12. M. Mokaya, F. Imrie, W. P. van Hoorn, A. Kalisz, A. R. Bradley, C. M. Deane, Testing the limits of smiles-based de novo molecular generation with curriculum and deep reinforcement learning. *Nat. Mach. Intell.* **5**, 386–394 (2023).
13. J. You, B. Liu, Z. Ying, V. Pande, J. Leskovec, Graph convolutional policy network for goal-directed molecular graph generation. *Adv. Neural Inf. Process. Syst.* **31**, 6410–6421 (2018).
14. W. Jin, R. Barzilay, T. Jaakkola, Hierarchical generation of molecular graphs using structural motifs. *International conference on machine learning* (PMLR, 2020), pp. 4839–4848.
15. O. Mahmood, E. Mansimov, R. Bonneau, K. Cho, Masked graph modeling for molecule generation. *Nat. Commun.* **12**, 3156 (2021).
16. M. Moret, I. Pachon Angona, L. Cotos, S. Yan, K. Atz, C. Brunner, M. Baumgartner, F. Grisoni, G. Schneider, Leveraging molecular structure and bioactivity with chemical language models for de novo drug design. *Nat. Commun.* **14**, 114 (2023).
17. N. W. Gebauer, M. Gastegger, S. S. Hessmann, K.-R. Müller, K. T. Schütt, Inverse design of 3D molecular structures with conditional generative neural networks. *Nat. Commun.* **13**, 973 (2022).
18. P. G. Francoeur, T. Masuda, J. Sunseri, A. Jia, R. B. Iovanisci, I. Snyder, D. R. Koes, Three-dimensional convolutional neural networks and a cross-docked data set for structure-based drug design. *J. Chem. Inf. Model.* **60**, 4200–4215 (2020).
19. Y. Li, J. Pei, L. Lai, Structure-based de novo drug design using 3D deep generative models. *Chem. Sci.* **12**, 13664–13675 (2021).
20. O. Zhang, T. Wang, G. Weng, D. Jiang, N. Wang, X. Wang, H. Zhao, J. Wu, E. Wang, G. Chen, Y. Deng, P. Pan, Y. Kang, C.-Y. Hsieh, T. Hou, Learning on topological surface and geometric structure for 3D molecular generation. *Nat. Comput. Sci.* **3**, 849–859 (2023).
21. S. Luo, J. Guan, J. Ma, J. Peng, A 3D generative model for structure-based drug design. *Adv. Neural Inf. Process. Syst.* **34**, 6229–6239 (2021).
22. X. Peng, S. Luo, J. Guan, Q. Xie, J. Peng, J. Ma, Pocket2mol: efficient molecular sampling based on 3d protein pockets. *International Conference on Machine Learning* (PMLR, 2022), pp. 17644–17655.
23. O. Zhang, J. Zhang, J. Jin, X. Zhang, R. Hu, C. Shen, H. Cao, H. Du, Y. Kang, Y. Deng, F. Liu, G. Chen, C.-Y. Hsieh, T. Hou, Resgen is a pocket-aware 3D molecular generation model based on parallel multiscale modelling. *Nat. Mach. Intell.* **5**, 1020–1030 (2023).
24. Z. Zhang, Y. Min, S. Zheng, Q. Liu, Molecule generation for target protein binding with structural motifs. *International Conference on Learning Representations* (ICLR, 2023).
25. J. Ho, A. Jain, P. Abbeel, Denoising diffusion probabilistic models. *Adv. Neural Inf. Process. Syst.* **33**, 6840–6851 (2020).

26. T. Weiss, E. Mayo Yanes, S. Chakraborty, L. Cosmo, A. M. Bronstein, R. Gershoni-Poranne, Guided diffusion for inverse molecular design. *Nat. Comput. Sci.* **3**, 873–882 (2023).
27. Y. Wang, M. Song, F. Liu, Z. Liang, R. Hong, Y. Dong, H. Luan, X. Fu, W. Yuan, W. Fang, G. Li, H. Lou, W. Chang, Artificial intelligence using a latent diffusion model enables the generation of diverse and potent antimicrobial peptides. *Sci. Adv.* **11**, eadp7171 (2025).
28. L. Chen, Y. Li, Y. Ma, L. Gao, L. Yu, Multiscale graph equivariant diffusion model for 3D molecule design. *Sci. Adv.* **11**, eadv0778 (2025).
29. J. Guan, W. W. Qian, X. Peng, Y. Su, J. Peng, J. Ma, 3D equivariant diffusion for target-aware molecule generation and affinity prediction. *International Conference on Learning Representations* (ICLR, 2023).
30. J. Guan, X. Zhou, Y. Yang, Y. Bao, J. Peng, J. Ma, Q. Liu, L. Wang, Q. Gu, Decompdiff: Diffusion models with decomposed priors for structure-based drug design. arXiv:2403.07902 (2024).
31. Z. Huang, L. Yang, X. Zhou, Z. Zhang, W. Zhang, X. Zheng, J. Chen, Y. Wang, C. Bin, W. Yang, Protein-ligand interaction prior for binding-aware 3d molecule diffusion models. *International Conference on Learning Representations* (ICLR, 2024).
32. Y. Qu, K. Qiu, Y. Song, J. Gong, J. Han, M. Zheng, H. Zhou, W.-Y. Ma, MolCRAFT: structure-based drug design in continuous parameter space. *International Conference on Machine Learning* (PMLR, 2024), pp. 41749–41768.
33. S. J. De Vries, M. Van Dijk, A. M. Bonvin, The haddock web server for data-driven biomolecular docking. *Nat. Protoc.* **5**, 883–897 (2010).
34. X. Zhang, S. Zhao, X. Su, L. Xu, From docking to dynamics: Unveiling the potential non-peptide and non-covalent inhibitors of mpro from natural products. *Comput. Biol. Med.* **181**, 108963 (2024).
35. Z. Xu, Y. Luo, X. Zhang, X. Xu, Y. Xie, M. Liu, K. Dickerson, C. Deng, M. Nakata, S. Ji, Molecule3D: A benchmark for predicting 3D geometries from molecular graphs. arXiv:2110.01717 [cs.LG] (2021).
36. A. P. Bento, A. Hersey, E. Félix, G. Landrum, A. Gaulton, F. Atkinson, L. J. Bellis, M. De Veij, A. R. Leach, An open source chemical structure curation pipeline using rdkit. *J. Chem.* **12**, 51 (2020).
37. D. E. Clark, Rapid calculation of polar molecular surface area and its application to the prediction of transport phenomena. 1. prediction of intestinal absorption. *J. Pharm. Sci.* **88**, 807–814 (1999).
38. O. Trott, A. J. Olson, Autodock vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **31**, 455–461 (2010).
39. M. F. Adasme, K. L. Linnemann, S. N. Bolz, F. Kaiser, S. Salentin, V. J. Haupt, M. Schroeder, Plip 2021: Expanding the scope of the protein–ligand interaction profiler to dna and rna. *Nucleic Acids Res.* **49**, W530–W534 (2021).
40. S. Salentin, S. Schreiber, V. J. Haupt, M. F. Adasme, M. Schroeder, Plip: Fully automated protein–ligand interaction profiler. *Nucleic Acids Res.* **43**, W443–W447 (2015).
41. H. Lin, G. Zhao, O. Zhang, Y. Huang, L. Wu, C. Tan, Z. Liu, Z. Gao, S. Z. Li, CBGBench: fill in the blank of protein-molecule complex binding graph. *International Conference on Learning Representations* (ICLR, 2025).
42. S. Lee, J. Jo, S. J. Hwang, Exploring chemical space with score-based out-of-distribution generation. *International Conference on Machine Learning* (PMLR, 2023), pp. 18872–18892.
43. H. Abdi, L. J. Williams, Principal component analysis. *Wiley Interdiscip. Rev. Comput. Stat.* **2**, 433–459 (2010).
44. S. K. Burley, C. Bhikadiya, C. Bi, S. Bittrich, H. Chao, L. Chen, P. A. Craig, G. V. Crichlow, K. Dalenberg, J. M. Duarte, S. Dutta, M. Fayazi, Z. Feng, J. W. Flatt, S. Ganesan, S. Ghosh, D. S. Goodsell, R. K. Green, V. Guranovic, J. Henry, B. P. Hudson, I. Khokhriakov, C. L. Lawson, Y. Liang, R. Lowe, E. Peisach, I. Persikova, D. W. Piehl, Y. Rose, A. Sali, J. Segura, M. Sekharan, C. Shao, B. Vallat, M. Voigt, B. Webb, J. D. Westbrook, S. Whetstone, J. Y. Young, A. Zalevsky, C. Zardecki, Rcsb protein data bank (rcsb.org): Delivery of experimentally-determined pdb structures alongside one million computed structure models of proteins from artificial intelligence/machine learning. *Nucleic Acids Res.* **51**, D488–D508 (2023).
45. M. Nakata, T. Shimazaki, Pubchemqc project: A large-scale first-principles electronic structure database for data-driven chemistry. *J. Chem. Inf. Model.* **57**, 1300–1308 (2017).
46. L. Huang, T. Xu, Y. Yu, P. Zhao, X. Chen, J. Han, Z. Xie, H. Li, W. Zhong, K.-C. Wong, H. Zhang, A dual diffusion model enables 3D molecule generation and lead optimization based on target pockets. *Nat. Commun.* **15**, 2657 (2024).
47. M. Goel, S. Raghunathan, S. Laghuvarapu, U. D. Priyakumar, Molegular: Molecule generation using reinforcement learning with alternating rewards. *J. Chem. Inf. Model.* **61**, 5815–5826 (2021).
48. Y. Fan, O. Watkins, Y. Du, H. Liu, M. Ryu, C. Boutilier, P. Abbeel, M. Ghavamzadeh, K. Lee, K. Lee, Dpok: Reinforcement learning for fine-tuning text-to-image diffusion models. *Adv. Neural Inf. Process. Syst.* **36**, 79858–79885 (2023).
49. X. Hu, G. Liu, Y. Zhao, H. Zhang, De novo drug design using reinforcement learning with multiple gpt agents. *Adv. Neural Inf. Process. Syst.* **36**, 7405–7418 (2023).
50. J. Schulman, F. Wolski, P. Dhariwal, A. Radford, O. Klimov, Proximal policy optimization algorithms. arXiv:1707.06347 [cs.LG] (2017).
51. C. A. Lipinski, F. Lombardo, B. W. Dominy, P. J. Feeney, Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* **23**, 3–25 (1997).
52. D. F. Veber, S. R. Johnson, H.-Y. Cheng, B. R. Smith, K. W. Ward, K. D. Kopple, Molecular properties that influence the oral bioavailability of drug candidates. *J. Med. Chem.* **45**, 2615–2623 (2002).
53. A. K. Ghose, V. N. Viswanadhan, J. J. Wendoloski, A knowledge-based approach in designing combinatorial or medicinal chemistry libraries for drug discovery. 1. A qualitative and quantitative characterization of known drug databases. *J. Comb. Chem.* **1**, 55–68 (1999).
54. W. Wei, S. Cherukupalli, L. Jing, X. Liu, P. Zhan, Fsp3: A new parameter for drug-likeness. *Drug Discov. Today* **25**, 1839–1845 (2020).
55. F. Wilcoxon, Individual comparisons by ranking methods. *Biometrics Bull.* **1**, 80–83 (1945).
56. K. O. McGraw, S. P. Wong, A common language effect size statistic. *Psychol. Bull.* **111**, 361–365 (1992).
57. J. Cohen, *Statistical power analysis for the behavioral sciences* (Taylor & Francis, 2013); 10.4324/9780203771587.
58. J. L. Hodges Jr, E. L. Lehmann, in *Selected works of EL Lehmann* (Springer, 2012), pp. 287–300.
59. A. Alhossary, S. D. Handoko, Y. Mu, C.-K. Kwoh, Fast, accurate, and reliable molecular docking with quickvina 2. *Bioinformatics* **31**, 2214–2216 (2015).
60. N. M. Hassan, A. A. Alhossary, Y. Mu, C.-K. Kwoh, Protein-ligand blind docking using quickvina-w with inter-process spatio-temporal integration. *Sci. Rep.* **7**, 15451 (2017).
61. A. T. McNutt, P. Francoeur, R. Aggarwal, T. Masuda, R. Meli, M. Ragoza, J. Sunseri, D. R. Koes, Gnina 1.0: Molecular docking with deep learning. *J. Chem.* **13**, 43 (2021).
62. K. Preuer, P. Renz, T. Unterthiner, S. Hochreiter, G. Klambauer, Fréchet chemnet distance: A metric for generative models for molecules in drug discovery. *J. Chem. Inf. Model.* **58**, 1736–1741 (2018).

Acknowledgments: We thank all those who maintain excellent databases and to all experimentalists who enabled this work by making their data publicly available. G. Chen would like to acknowledge M.J. Chen for the lifelong support and for influencing the exploration of AI-driven drug discovery. **Funding:** This work was supported by the National Key Research and Development Program of China (2022YFA1103102 to Y.Z.), the National Natural Science Foundation of China (32170589 to Y.Z., 32370616 to Y.Z., 32488201 to S. Gao, 32330030 to S. Gu, 22074105 to Z.T., and 62506263 to S.Q.), the China Postdoctoral Science Foundation (2025M771539 to S.Q. and GZB20250385 to S.Q.), the Xiaomi Young Talents Program (to G.C.), and the European Union's Horizon 2020 Framework Programme for Research and Innovation under the specific grant agreement no. 945539 (Human Brain Project SGA3; to A.K.). **Author contributions:** Conceptualization: G.C., X.Z., Y.Z., C.J., and A.K. Investigation: X.Z., G.C., Y.Z., S. Gu, and Z.T. Methodology: X.Z., Y.Z., and G.C. Resources: G.C., Y.Z., S.G., Z.T., S. Gao, C.J., and A.K. Data curation: X.Z. Software: X.Z. and J.W. Validation: X.Z., S.Q., and J.W. Formal analysis: X.Z., S.Q., and F.L. Visualization: X.Z. and G.C. Writing—original draft: X.Z., G.C., S.Q., F.L., and J.W. Writing—review and editing: X.Z., G.C., Y.Z., S. Gu, Z.T., S.Q., F.L., C.J., and J.W. Funding acquisition: G.C., Y.Z., S.Q., C.J., and A.K. Supervision: G.C., Y.Z., C.J., and A.K. Project administration: G.C., Y.Z., X.Z., and C.J. **Competing interests:** The authors declare that they have no competing interests. **Data, code, and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. This study did not generate any new physical or biological materials. The original Molecule3D dataset is obtained from https://github.com/divelab/MoleculeX, and the CrossDocked2020 dataset is accessed at https://bits.csb.pitt.edu/files/crossdock2020. The processed Molecule3D and CrossDocked2020 datasets used in this study, along with all source code, are available at Zenodo (https://zenodo.org/records/17790890) or GitHub (https://github.com/ispc-lab/SeFMol).

Submitted 16 May 2025
Accepted 18 March 2026
Published 15 April 2026
10.1126/sciadv.ady9955

ScienceAdvances

Steering semi-flexible molecular diffusion model for structure-based drug design with reinforcement learning

Xudong Zhang, Sanqing Qu, Fan Lu, Jianmin Wang, Zhixin Tian, Shangding Gu, Yanping Zhang, Alois Knoll, Shaorong Gao, Guang Chen, and Changjun Jiang

Sci. Adv. **12** (16), eady9955. DOI: 10.1126/sciadv.ady9955