Preprints.org

From Latent Manifolds to Functional Probes: An Interpretable, Kinome-Scale Generative Machine Learning Framework for Family-Targeted Kinase Inhibitor Design

Gennady Verkhivker [*] , Ryan Kassab , Keerthi Krishnan

Article

From Latent Manifolds to Functional Probes: An Interpretable, Kinome-Scale Generative Machine Learning Framework for Family-Targeted Kinase Inhibitor Design

Gennady Verkhivker [1,2,*], Ryan Kassab [1] and Keerthi Krishnan [1]

[1] Keck Center for Science and Engineering, Graduate Program in Computational and Data Sciences, Schmid, College of Science and Technology, Chapman University, Orange, CA 92866, United States of America

[2] Department of Biomedical and Pharmaceutical Sciences, Chapman University School of Pharmacy, Irvine, CA 92618, United States of America

* Correspondence: verkhivk@chapman.edu; Tel.: +1-714-516-4586

Abstract

The design of selective kinase inhibitors remains a formidable challenge due to the high structural conservation of the ATP-binding site across the kinome, and the topological complexity of pharmacophores required for potent inhibition. While modern generative AI has enabled rapid exploration of chemical space, many advanced models operate as black boxes, obscuring the chemical rationale behind design choices and limiting interpretability for medicinal chemists. Here, we present a modular, chemistry-first generative framework for *de novo* design of SRC kinase inhibitors, integrating ChemVAE-based latent space modeling, a chemically interpretable Kinase Inhibition Likelihood scoring function, Bayesian optimization, and cluster-guided local neighborhood sampling. Our generative pipeline employs a hybrid AI framework that integrates deep variational autoencoding, interpretable machine learning–based scoring, and probabilistic optimization to enable targeted exploration of kinase inhibitor chemical space. Our analysis reveals three pivotal findings. We demonstrate that kinase inhibitors—spanning ten families—spontaneously organize into a coherent, low-dimensional manifold in latent space, with SRC acting as a structural "hub" that enables rational scaffold transformation. Our local neighborhood sampling-based approach successfully converts inhibitors from other kinase families (notably LCK) into novel SRC-like chemotypes, with LCK-derived molecules accounting for ~40% of high-similarity outputs. However, both generative strategies reveal a critical limitation: SMILES-based representations systematically fail to recover multi-ring aromatic systems—a hallmark of clinical kinase inhibitors—despite aromatic ring count being a top feature in Kinase Inhibition Likelihood scoring function. This "representation gap" underscores that no amount of scoring refinement can compensate for a generative engine that cannot access topologically complex regions. By diagnosing these constraints within a transparent, interpretable pipeline, our work provides a foundational benchmark for current AI and a blueprint for hybrid systems that blend algorithmic innovation with medicinal chemistry principles.

Keywords: autonomous molecular design; deep learning models; latent space landscapes; protein kinases; kinase inhibitors; local neighborhood sampling chemical modeling; kinase inhibition likelihood classifiers; explainable machine learning

1. Introduction

Discovery of small-molecule inhibitors—especially against high-value but structurally complex targets such as kinases, GPCRs, and protein–protein interfaces—unfolded through a painstaking

cycle of chemical synthesis, high-throughput screening, and iterative structure–activity relationship (SAR) analysis. This process, constrained by experimental throughput and human intuition, often took years to yield a single clinical candidate. The past decade, however, has witnessed a profound transformation: the integration of artificial intelligence (AI) and machine learning (ML) into drug discovery has enabled the *de novo*, property-driven generation of drug-like molecules with unprecedented speed, scale, and chemical novelty [1–8]. Many deep learning approaches have been put forward employing various neural network architectures, molecular representations and analysis metrics for targeted compound design, and their applications [9–16]. This paradigm shift in the drug discovery field has not been monolithic but has evolved through a series of methodologically distinct yet conceptually linked phases, each building on the successes and correcting the shortcomings of the last and collectively steering the field from syntax-aware sequence modeling toward structure- and function-aware molecular design.

The first wave of this transformation emerged between 2017 and 2019, when several studies began treating molecules as textual sequences using the Simplified Molecular Input Line Entry System (SMILES) and applying natural language processing (NLP) techniques to chemical space. Deep neural network (DNN) models, most notably variational autoencoder (VAE) [9] and generative adversarial networks (GAN) [17] have been particularly fruitful in molecular design of novel chemical probes [17–28]. Among the earliest efforts was sequence data generation (SeqGAN) approach [18], and Objective-Reinforced Generative Adversarial Networks (ORGAN) [19] which coupled a recurrent neural network (RNN) generator with a discriminator trained not just to assess chemical validity but to maximize user-defined molecular properties—a pioneering step toward goal-directed generation. LatentGAN combined an autoencoder and a generative adversarial neural network for de novo molecular design [20]. DruGAN approach combined GAN and VAE by training an adversarial autoencoder to efficiently sample molecules from the latent space [23]. Soon after, MolGAN [29] adapted GANs to molecular design, generating adjacent and feature matrices to represent molecular graphs directly. CycleGAN provided unpaired Image-to-Image translation using Cycle-Consistent Adversarial Networks [30]. MolCycleGAN, which extended the CycleGAN framework, can learn transformation rules from the sets of compounds with desired and undesired values of the considered property [31]. The methodological progress in GAN applications to molecular discovery has been catalyzed by the development of several comprehensive benchmarking sets and cheminformatics infrastructure [32–36].

Despite its conceptual elegance, MolGAN and related GAN approaches suffered from severe mode collapse and produced valid molecules less than 30% of the time, highlighting the fragility of adversarial training in discrete spaces. A more robust alternative arrived with ChemVAE [9], a variational autoencoder that encoded SMILES strings into a smooth, continuous 196-dimensional latent space while simultaneously predicting key drug-likeness metrics—quantitative estimate of drug-likeness (QED) [37], synthetic accessibility score (SAS) [38], and logP [39].

The field then pivoted between 2019 and 2021 toward active, reward-guided strategies that could steer generation with greater precision. The most influential of these was REINVENT [20,40,41] a reinforcement learning (RL) framework that used policy gradients to fine-tune an RNN generator toward a customizable reward function. This reward could combine multiple objectives—such as predicted binding affinity, QED, and Tanimoto similarity to a reference scaffold—effectively turning the generative model into a programmable design engine. REINVENT quickly became the industry standard for scaffold hopping and lead optimization [40,41]. GENTRL approach compresses the space of small molecule structures onto a distribution that parameterizes the latent space in a high-dimensional lattice following by exploration and optimization of the latent space by reinforcement learning to discover novel kinase inhibitors [42].

Concurrently, transformer architectures began to reshape the landscape of chemical AI. Attention-based generative models for de novo molecular design offered new architectures that enabled a more accurate sampling from the latent space and exploration of novel chemistry space not present in the training data [43], thus optimizing the tradeoffs between model exploration and

structure of the latent memory. Efficient multi-objective molecular design approaches combine *in silico* prediction of molecular properties defined desirability ranges and substructure constraints with particle swarm optimization for optimal navigation in a continuous latent space [44–46]. A highly efficient and generic query-based molecule optimization framework QMO facilitates molecule optimization by decoupling molecule representation learning and guided search method based on zeroth-order optimization in the molecular property landscape [47]. The performances of DL-based, VAE, GAN and RNN models were evaluated in goal-directed (rediscovery, optimization and scaffold hopping of active compounds) and target-specific (generation of novel compounds for a given target) tasks [48]. Simultaneously, SMILES-BERT [49] and Chemformer [50,51] applied transformer architectures to molecular sequences, leveraging self-supervised pretraining on billions of compounds to improve generation quality and transfer learning. These approaches offered greater controllability and higher validity, but remained constrained by the sequential nature of SMILES, which struggles to represent cyclic and stereochemical complexity. Meanwhile, MolDQN bypassed SMILES by applying deep Q-learning to discrete molecular graph actions, optimizing molecular properties through a Markov decision process [52]. Despite these advances, this phase remained two-dimensional: rewards were often computed using surrogate predictors—such as Random Forests trained on RDKit descriptors rather than direct protein–ligand interactions, yielding molecules that were chemically plausible but often pharmacologically inert.

Some generative models aiming at three-dimensional (3D) molecule generation have also been proposed, gaining attention for their unique advantages and potential to explicitly design drug-like molecules in a target-conditioning manner [53]. A novel molecular deep generative model adopts a recurrent neural network architecture coupled with a ligand-protein interaction fingerprint as constraints [54]. DeepLigBuilder, a deep learning-based method for *de novo* drug design combined Ligand Neural Network (L-Net) graph generative model for design of chemically and conformationally valid 3D molecules with Monte Carlo tree search to optimize structure-based *de novo* drug design parameters such as high predicted affinity, and similar binding features to those of known inhibitors [55]. A comprehensive review of 3D molecular generative models reported current techniques for the molecular structure generation and categorized them into three types, depending on featurization of 3D molecular structures: cubic grid-based, Euclidean distance matrix(EDM)-based, and Cartesian coordinate-based, where each type of featurization requires distinct generative architectures and optimization strategies [56].

Th advent of graph representation learning, a class of machine learning methods that natively operate on graph-structured data, has enabled neural networks to learn directly from molecular topology. Over the past seven years, graph neural networks (GNNs) have redefined the very architecture of *de novo* molecular design, virtual screening, and protein–ligand interaction modeling. The adoption of GNNs in drug discovery began with their ability to natively represent molecules as graphs and learn structure–property relationships directly from topology. Gilmer et al. [57] laid out the conceptual foundation of GNNs with the introduction of Neural Message Passing (MPNN), a unifying framework that cast molecular property prediction as an iterative process of information exchange between atoms and bonds. This insight catalyzed a wave of chemistry-specific GNN architectures. Kearnes et al. demonstrated that Graph Convolutional Networks (GCNs) could predict ADMET properties with high accuracy by aggregating neighborhood features in molecular graphs [58]. Soon after, Veličković et al. introduced Graph Attention Networks (GATs), which learned to weight the importance of neighboring atoms dynamically, capturing subtle electronic effects critical for reactivity and binding [59]. A pivotal advance came with Directed Message Passing Neural Networks (D-MPNN) approach which explicitly modeled bond directionality and chirality—features essential for drug-likeness and target specificity [60] D-MPNN achieved state-of-the-art performance across quantum chemical (QM9) and bioactivity (MUV, Tox21) benchmarks and became the core of the open-source Chemprop framework, now widely adopted in both academia and industry for interpretable molecular property prediction [60].JT-VAE approach widely adopted by 2021, decomposed molecules into hierarchical junction trees of rings and chains, enabling near-perfect

validity (>99%) and precise control over scaffold modification [61]. Junction tree variational autoencoder (JT-VAE) generates molecular graphs in two phases, by first generating a tree-structured scaffold over chemical substructures and then combining them into a molecule with a graph message passing network [61]. Similarly, GraphAF used autoregressive normalizing flows to build molecular graphs atom-by-atom with high fidelity, ensuring that valency and chirality were respected [62]. More recently, GFlowNets [63,64] introduced a probabilistic framework for sampling molecules proportional to a reward function (e.g., binding affinity), mitigating the mode collapse and low diversity that plagued earlier generative approaches.

As the field matured, the limitations of static, 2D graph representations became evident—particularly for tasks requiring 3D conformational awareness, such as protein–ligand docking and allosteric modulation. This spurred the rise of geometric deep learning, where models respect the rotational and translational symmetries of physical space. SchNet, introduced by Schütt et al. [65,66] pioneered the use of continuous-filter convolutions operating directly on atomic coordinates, enabling accurate prediction of molecular energies and interatomic forces with quantum-mechanical fidelity. This was significantly refined in DimeNet and DimeNet++ [67,68] which incorporated directional message passing using interatomic angles, dramatically improving the modeling of torsional strain, steric clashes, and binding pocket complementarity. The culmination of this trend arrived with SE(3)-equivariant GNNs, exemplified by EquiBind which predicted protein–ligand binding poses in seconds—without traditional docking—by learning geometric constraints directly from structural data [69]. EquiBind achieved near-experimental accuracy on the PDBBind benchmark, effectively replacing physics-based scoring in early-stage screening [69].

Critically, graph representation learning also enabled direct modeling of protein–ligand complexes as heterogeneous graphs, where protein residues and ligand atoms form distinct node types connected by cross-edges. TANKBind approach segments the whole protein into functional blocks and predict their interactions with the ligand, creating a protein-ligand interaction energy landscape using a novel trigonometry-aware architecture. In the second stage, TANKBind prioritizes the crystal structures by constrastively ensuring a weaker binding affinity for non-native interactions [70]. Self-supervised, pretrainable geometric GNNs can learn rich representations of molecules and proteins from unlabeled structural data and represent a new class of models designed for molecular property prediction that leverage 3D molecular structure information during pre-training to improve performance on downstream tasks [71–73]. Graph Multi-View Pre-training (GraphMVP) framework addresses limited 3D molecular data by using self-supervised learning with contrastive learning to enforce consistency between 2D and 3D molecular representations, enhancing performance in property prediction tasks [71].

The current frontier integrates generative modeling and massive scale pretraining into end-to-end systems capable of co-designing proteins and ligands from first principles. Diffusion models have emerged as one of the dominant generative frameworks. GeoDiff is the first SE(3)-equivariant diffusion model that operates directly on atomic coordinates and learns to reverse a diffusion process that gradually adds noise to a molecule's 3D structure [74]. This approach enabled structure-aware ligand generation for docking and binding prediction and laid the foundation for protein-conditioned diffusion models. By expanding this work, Tang and his group introduced a pretrainable, SE(3)-equivariant geometric GNN specifically designed for antibody affinity maturation [75]. This work bridged geometric deep learning and self-supervised pretraining with high-throughput experimental biology to create a predictive, generative, and actionable platform for antibody engineering. These studies underscored a broader trend: the shift from sequence-based to structure-based AI in therapeutic discovery, with geometric GNNs at the forefront. TorsionDiff [76] operates in torsion angle space to produce realistic side-chain rotamers. DiffDock predicts binding poses with near-experimental accuracy, effectively replacing classical docking pipelines [77]. DiffDock-L, a latest version of DiffDock provides a significant improvement in performance and generalization capacity [78]. DiffLinker is a new Equivariant 3D-conditional Diffusion Model for

Molecular Linker Design that places missing atoms in between and designs a molecule incorporating all the initial fragments [79].

When conditioned on protein structure, diffusion-based models achieve remarkable biological specificity. RFdiffusion approach developed by the Baker Lab uses a protein backbone diffusion model [80] and when paired with the sequence design tool ProteinMPNN [81] enables *de novo* creation of protein binders, allosteric pockets, and small-molecule scaffolds. Complementing these are foundation models trained on multimodal biological data: ESM3 developed by EvolutionaryScale, integrates sequences, 3D structures, and functional annotations into a single architecture capable of zero-shot ligand generation via in-context learning [82]. Chroma, a generative model for proteins and protein complexes, can directly sample novel protein structures and sequences, and that can be conditioned to steer the generative process towards desired properties and functions [83]. This evolution has been accelerated by open-source ecosystems and standardized benchmarks. PyTorch Geometric (PyG) [84] and Deep Graph Library (DGL) [85] provide modular, scalable implementations of GNN layers. The Therapeutics Data Commons (TDC) [86] offers a unified benchmark with 66 therapeutic tasks—including kinase inhibitor design, antibody escape, and molecular generation—enabling fair comparison and reproducibility across the field.

Overall, graph representation learning has transformed computational drug design from a descriptor-driven science into topology- and geometry-aware engineering. By respecting the intrinsic structure of molecules and their biological targets, GNNs have not only improved predictive accuracy but have restored chemical realism to generative AI—paving the way for the next generation of rational, mechanism-informed, and human-aligned drug discovery.

This progression reflects a profound paradigm shift: ligands are no longer optimized against scalar activity predictors but directly against 3D protein structures; scoring is embedded within the generative process itself, eliminating reliance on external reward functions; and multi-objective trade-offs—between potency, selectivity, and developability—are handled implicitly through conditional diffusion or multi-task pretraining. Yet significant challenges remain. Data scarcity for mutant or allosteric targets demands robust few-shot and zero-shot capabilities. And as models grow more end-to-end, their black-box nature threatens interpretability—a gap that hybrid approaches, combining the predictive power of graph neural networks with chemically grounded, interpretable features metric, may help bridge. In the development of kinase inhibitory therapeutics, generating novel selective probes to interrogate specific protein kinases is a major challenge and machine learning-enabled targeted transformations and chemical morphing between kinase inhibitors from different families can provide a valuable resource for new indications of existing kinase molecules.

The current study is situated at the inflection point—between the promise of end-to-end deep generative models and the enduring need for chemical interpretability. While modern AI tools offer unprecedented generative power, they often operate as black boxes, making it difficult to diagnose failure modes, ensure pharmacophoric fidelity, or guide iterative refinement. In response, we present a modular, interpretable machine learning platform for the *de novo* design of SRC kinase inhibitors that strategically combines the representational capacity of ChemVAE, the chemical grounding of a feature-based kinase-specific scoring function, the sample efficiency of Bayesian optimization, and the scaffold-aware exploration of local neighborhood sampling-based latent space engineering.

Our generative pipeline employs a hybrid AI framework that integrates deep variational autoencoding, interpretable machine learning–based scoring, and probabilistic optimization to enable targeted exploration of kinase inhibitor chemical space. The pipeline explicitly decomposes the design process into several interconnected stages : (a) deep generative backbone for latent space representation, (b) ML scorer for target-specific guidance, (c) probabilistic optimization engine (Bayesian Optimization) for search, (d) clustering-and-local neighborhood sampling layer for scaffold transformation. This work represents a significant conceptual and methodological extension of our earlier study [87]. While that study introduced the ChemVAE framework and demonstrated initial scaffold transformation via latent space local neighborhood sampling, the present study delivers a more comprehensive, multi-strategy generative pipeline with several critical advances. Most notably,

we now expand the data sets of random molecules and kinase inhibitors, introduce and rigorously benchmark Bayesian Optimization as a complementary global search strategy, revealing its strengths (efficient drug-likeness tuning) and fundamental limitations (systematic failure to recover multi-ring pharmacophores)—a diagnostic insight absent in the prior work. Beyond methodology, the current study provides structural validation through computational docking, confirming that remodeled molecules not only resemble but functionally mimic clinical SRC inhibitors in binding mode and affinity. Furthermore, we offer a mechanistic interpretation of latent space organization, identifying SRC as a structural "hub" and LCK as a uniquely "plastic" scaffold for transformation—findings grounded in statistical analysis of latent distributions across ten kinase families. Critically, we place our results in the context of modern generative AI, diagnosing the "representation gap" of SMILES-based models and articulating a clear path toward hybrid systems that integrate geometric GNNs with interpretable, chemistry-first design.

2. Materials and Methods

2.1. Data Sets of Protein Kinase Inhibitors and Small Molecules

To construct a robust and representative foundation for generative kinase inhibitor design, we assembled a large-scale, multi-source dataset that reflects the current state of kinase-targeted chemical space. Numerous large databases are available that contain molecules in a variety of representations including SMILES, 2D, and 3D. For this study, we explored the databases of generic small molecules and drug-like inhibitors primarily ChEMBL [88], DrugBank [89,90], BindingDB [91], BindingMoad [92], ChEBI [93], ZINC, a free database of commercially available compounds that contains over 230 million purchasable compounds in ready-to-dock, 3D formats [94–96]. Our inhibitor collection integrates high confidence bioactive compounds from ChEMBL v32, DrugBank v5.1 [90], PDBbind v2023 [91] and ZINC20 [96] . To provide a meaningful contrast to kinase-biased chemistry, we sampled drug-like matter from two ultra-large enumerative databases: GDB-17 Lead-Like Set: ~11 million molecules filtered for lead-like properties (MW ≤ 450, logP ≤ 4, ≤4 HBD/HBA) [97,98], FDB-17 subset ~10 million fragment-like compounds derived from GDB-17 using synthetic accessibility and complexity filters [99]. From these, we selected ~ 220,000 diverse molecules satisfying Lipinski's Rule of Five (MW < 700, logP ∈ [–4,6], ≤6 rotatable bonds, ≤12 HBD/HBA) and restricted to biologically relevant atoms (C, N, O, F, S, P, Cl, Br, I). This "random" background set ensures the model learns to distinguish kinase-specific pharmacophores from generic drug-like space.

For generative kinase inhibitor design, we assembled a comprehensive dataset of protein kinase inhibitors (PKIs). In 2023, Bajorath reported a total of 155,579 qualifying unique human PKIs [100]. Our curation strategy is informed by recent systematic analyses of the kinome-wide inhibitor landscape, including the landmark 2025 review by Koch, Kullmann, and Bajorath [101] which reports that over 206,000 protein kinase inhibitors have been disclosed as of 2025—spanning orthosteric, allosteric, and covalent mechanisms across the human kinome. For datasets of PKIs, we used ~60,000 available high-confidence PKIs. The expanded set covered the expanded set of kinase families totaling 37 distinct kinase families across the human kinome, including: SRC (SRC, LCK, FYN, YES), ABL (ABL1, ABL2), EGFR (EGFR, ERBB2/HER2, ERBB4), PDGFR (PDGFRα, PDGFRβ, KIT, CSF1R, FLT3), FGFR (FGFR1–4) , INSR (INSR, IGF1R), TRK (NTRK1/2/3), ROS (ROS1, DDR1, DDR2), MET (MET, RON), RAF (ARAF, BRAF, CRAF), MLK (MAP3K9, MAP3K10, MAP3K11), LRRK (LRRK1, LRRK2), STKR (ALK, LTK, ROS, RYK), TLK (TLK1, TLK2), RIPK (RIPK1–4), WNK (WNK1–4), CLK (CLK1–4), STE20 (PAK1–7, MAP4K1–7) STE11 (MAP3K1–13), STE7 (MAP2K1–7), CAMK (CAMK1–4, DAPK1–3) , DAPK, PHK (PHKG1/2), MLCK (MYLK) , DCAMKL (DCAMKL1–3), MELK, BRSK, PKA (PRKACA/B/C) , PKG (PRKG1/2) , PKC (PRKCA–Z), AKT (AKT1–3) , RSK (RPS6KA1–6) , SGK (SGK1–3) CDK (CDK1–20), MAPK (MAPK1/3/8/9/11/14/p38α–δ),GSK3 (GSK3A/B).

In the earlier study [87[we used the data set of competitive and allosteric protein kinase inhibitors confirmed by X-ray crystallography that contained a total of 2,899 unique inhibitors

including 136 allosteric and 2763 orthosteric compounds with a total of 231 protein kinases [102–104]. In the current study, we included the latest data from the KLIFS website (accessed April 2025) that reported 4,179 unique ligands confirmed by X-ray across 6,738 structures for 326 kinases [105].

We also expanded the list of allosteric kinase ligands based on recent systematic analysis of X-ray structures that identified a total of 262 allosteric PK ligands [106]. For focused generative experiments on SRC, we extracted 3,477 high-confidence SRC inhibitors ($IC_{50} \leq 100$ nM) and 1,883 ABL1 inhibitors as reference scaffolds. All molecules were standardized using RDKit [107,108] with salts removed, tautomers normalized, and stereochemistry preserved. All molecules including both kinase inhibitors and background compounds were converted to canonical SMILES and encoded into a 196-dimensional continuous latent space using the ChemVAE architecture [9]. ChemVAE converts discrete representations of molecules to and from a multidimensional continuous representation, enabling generation of new molecules for efficient exploration and optimization via open ended chemical spaces, enabling Bayesian optimization in latent space and allowing to navigate toward regions enriched for desired properties.

2.2. Guided Remodeling of Latent Neighborhoods via Cluster-Directed Sampling

To enable scaffold-aware transformation of kinase inhibitors across families, we developed a guided latent space remodeling strategy that leverages the intrinsic structural organization of the ChemVAE embedding. Rather than applying global or random modifications, our approach performs targeted local neighborhood sampling—a process that shifts molecular representations toward chemically coherent regions of latent space while preserving scaffold integrity. We began by applying K-means clustering to the 196-dimensional ChemVAE latent space to identify functionally homogeneous neighborhoods. This unsupervised step avoids manual labeling and allows molecular embeddings to self-organize into groups based solely on structural and physicochemical similarity. We evaluated cluster configurations ranging from 2 to 5 partitions and found that a 3-cluster split yielded the highest diversity and validity of generated molecules, as well as the clearest separation of scaffold motifs (e.g., fused heterocycles vs. linear aromatics). This configuration was selected for all subsequent remodeling experiments. Within each cluster, we performed centroid-directed sampling: for every molecule with latent representation: : for every molecule with latent representation x, we computed its displacement toward the cluster centroid c using a controlled interpolation:

$$\vec{x_i^*} = \vec{x_i} + s(\vec{c_i} - \vec{x_i}) \quad (1)$$

where the scaling factor governs the degree of remodeling. Given that the lower bound of $s = 0$ corresponds to the original encoding of a given molecule, while $s = 1$ provides us with the centroid of the cluster, this parameter was initially set to be a threshold of 0.5. By performing local sampling steps and evaluating kinase inhibition likelihood probabilities, we found that with the scaling factor $s < 0.5$ the yield of valid molecules decreased, while a scaling factor $s = 0.8$ remodels the molecule gradually towards the centroid of the cluster yielding valid molecules without losing information of the molecular attributes. To encourage diversity without destabilizing the latent geometry, we introduced low-magnitude isotropic noise (standard deviation = 5.0) to the remodeled vectors. Higher noise levels (≥10) degraded validity, as they pushed samples into sparse, low-decoding-density regions of the latent space. The combination of 3-cluster partitioning, centroid-directed sampling with , and minimal noise consistently produced the highest yield of valid, structurally diverse molecules. After remodeling, each vector was decoded into a SMILES string using the ChemVAE decoder. To ensure chemical plausibility, we implemented a two-stage filtering protocol: For validity screening, the decoder was run 500 times per vector; if at least one valid SMILES (as verified by RDKit) was produced, the molecule advanced. For size filtering, molecules with SMILES length < 10 were discarded to exclude trivial or non-drug-like outputs. The resulting compounds were then evaluated for kinase inhibition likelihood, structural similarity to SRC inhibitors, and drug-like properties to assess the success of scaffold transformation.

All scripts, software and models used in the development and experiments are available in the GitHub site https://github.com/kassabry/Local neighborhood sampling_Experiment. The GitHub repository contains the framework and code for molecular transformations of kinase inhibitors using generative learning methodologies and targeted remodeling. The software written is a combination with the ChemVAE deep learning framework that converts discrete representations of molecules to and from a multidimensional continuous representation, enabling generation of new molecules [9]. The GitHub site provides detailed documentation and guides of the deposited information and software. The deep learning frameworks were supported by the TensorFlow backend [109] and python tools such as NumPy, scipy, pandas, and scikitlearn.

2.3. Kinase Inhibition Likelihood Classifiers

The Random Forest classification method [110] was used to develop and evaluate multiclass and binary kinase inhibition likelihood classifiers in the latent and chemical spaces of small molecules. The model is initiated with the training set of molecules from all kinase families as well as GDB-17 molecules. Each molecule within the training set was processed through RDKit [107p,108] to calculate chemical features. Binary decision trees are created, and the chemical attributes were used as parameters to determine the most key features in determining the target variable. Each decision tree makes a prediction on the value of the target variable and the predictions are then aggregated and averaged to get a value between 0 and 1. If there are more than two classes, the predictions are normalized and then averaged to maintain a predicted value between 0 and 1. This would ensure that a target value would still be between 0 and 1, while allowing for multiple classification variables. For chemical feature-based classifier, 20 chemical features are considered for each molecule during training and testing: the number of rings, the exact molecular weight, the number of rotatable bonds, the fraction of carbon Sp3 atoms, the Hall–Kier alpha value, the Labute ASA value, the number of aliphatic carbocycles, the number of aliphatic heterocycles, the number of aliphatic rings, the number of amide bonds, the number of aromatic carbocycles, the number of aromatic heterocycles, the number of aromatic rings, the number of stereocenters, the number of bridgehead atoms, the number of H-bond acceptors, the number of H-bond donors, the QED value, the SAS value, and the logP value (Figure 1).

The resulting score the Random Forest models output represents the probability or "likelihood" that a molecule can be deemed an SRC Kinase Inhibitor. Values closer to 0 indicate that the molecule has low kinase inhibition likelihood whereas values closer to 1 indicate that the molecules have a high kinase inhibition likelihood. To assess the performance of each model, Accuracy, Recall, Precision and F1 score were calculated to measure the performance of classification models. These parameters are defined as follows :

$$Accuracy = \frac{TP+TN}{all}; \; Precision = \frac{TP}{TP+FP} \quad (2)$$

$$Recall = \frac{TP}{TP+FN}; \; F_1 = 2\frac{Precision*Recall}{Precision+Recall} \quad (3)$$
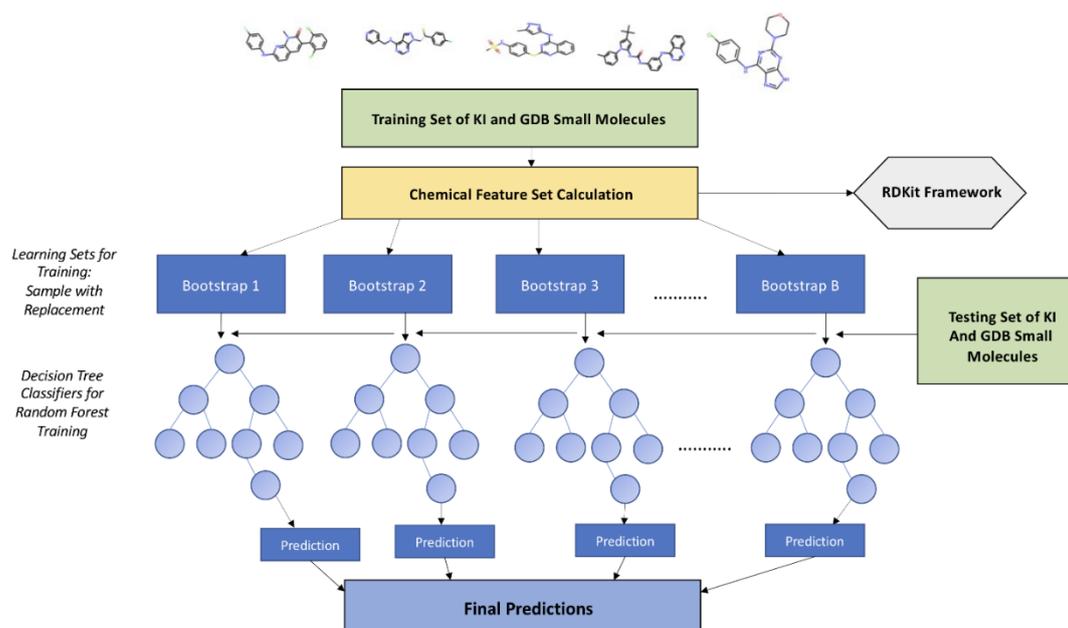
Figure 1. A schematic diagram of the Random Forest Model used for the kinase inhibition likelihood classifier.

An F-score is a measure of precision and recall and is often used in binary classification problems. Precision is defined as the number of positive samples the model predicts correctly (true positives) divided by the true positives plus the false positives. Recall is defined as true positives divided by true positives plus false negatives. The model performance was evaluated using receiver operating characteristic area under the curve. The receiver operating curve (ROC) is a graph where sensitivity is plotted as a function of 1-specificity. The area under the ROC is denoted AUC. A reliable and valid AUC estimate can be interpreted as the probability that the classifier will assign a higher score to a randomly chosen positive example than to a randomly chosen negative example.

3. Results and Discussion

3.1. The Kinase Inhibitor Dataset and Its Embedding Reveals Organized Kinome Manifold in Latent Space

This curated hybrid dataset comprising of ~ 220,000 diverse molecules forming background set and ~60,000 available high-confidence PKIs from 37 distinct kinase families across the human kinome served as the training corpus for all machine learning components of our pipeline. Central to our approach was the ChemVAE architecture trained on SMILES strings that learns a continuous, low-dimensional latent representation of molecular structure. ChemVAE encodes each molecule into a fixed-length vector (here, 196-dimensional) by compressing its SMILES sequence through a bottleneck layer, while simultaneously optimizing for accurate reconstruction and property prediction (e.g., QED, logP, synthetic accessibility). This process effectively translates discrete chemical syntax into a differentiable geometric space, where semantic similarity (e.g., shared scaffolds or functional groups) is reflected in spatial proximity (Figure 2).
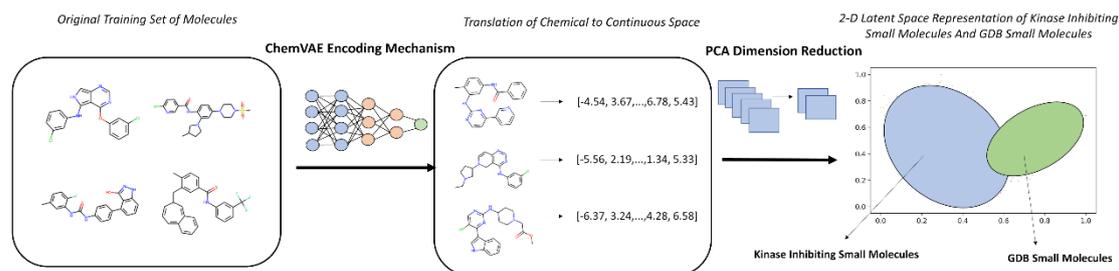
Figure 2. An Overview of Chemical to Continuous Space Translation using ChemVAE Encoding mechanism.

To interrogate the organization of this latent space, we performed principal component analysis (PCA) on the encoded vectors and visualized the results in two dimensions (Figure 3). Embedding our large-scale kinase inhibitor dataset into the ChemVAE latent space revealed a striking and functionally meaningful organization: rather than scattering randomly, 60,000 kinase inhibitors spanning 37 families across the human kinome—collapsed into a dense, low-volume manifold, sharply segregated from the diffuse cloud of 220,000 generic molecules (Figure 3 A,B). The PCA projection revealed that despite their pharmacological diversity, kinase inhibitors collapsed into a dense, spatially contiguous cluster, sharply demarcated from the diffuse, cloud-like distribution of GDB molecules (Figure 3A). This separation was not an artifact of labeling or sampling; it emerged naturally from the model's unsupervised training on SMILES syntax, suggesting that molecular sequence intrinsically encodes functional semantics. This separation persisted even when examining kinase inhibitors in isolation, where sub-clustering by family was evident but incomplete, reflecting shared ATP-binding motifs and overlapping chemotypes (Figure 3B).

Within this kinase-rich region, a hierarchical structure became apparent. At the global level, all ATP-competitive inhibitors clustered together, reflecting the conserved architecture of the kinase catalytic cleft. Yet at a finer scale, family-specific subclusters emerged, highlighted for ABL and SRC kinase inhibitors (Figure 3C) . The SRC family occupied the broadest region of latent space acting as a structural "hub" that overlapped significantly with LCK, ABL1, and EGFR. This proximity suggested that ABL, LCK and EGFR-derived molecules may be amenable to transformation into SRC-like chemotypes—a finding that would prove pivotal in our generative experiments. Visual inspection of the PCA-projected latent space revealed that most kinase inhibitors—regardless of target family—occupied a shared, high-density region that significantly overlapped with the clusters of SRC and ABL1 inhibitors (Figure 3B,C). This spatial co-localization suggests that, despite differences in selectivity and clinical indication, these molecules share a core set of chemical–functional features essential for ATP-competitive binding, such as planar aromatic systems, hydrogen bond acceptors at the hinge region, and moderate molecular weight.
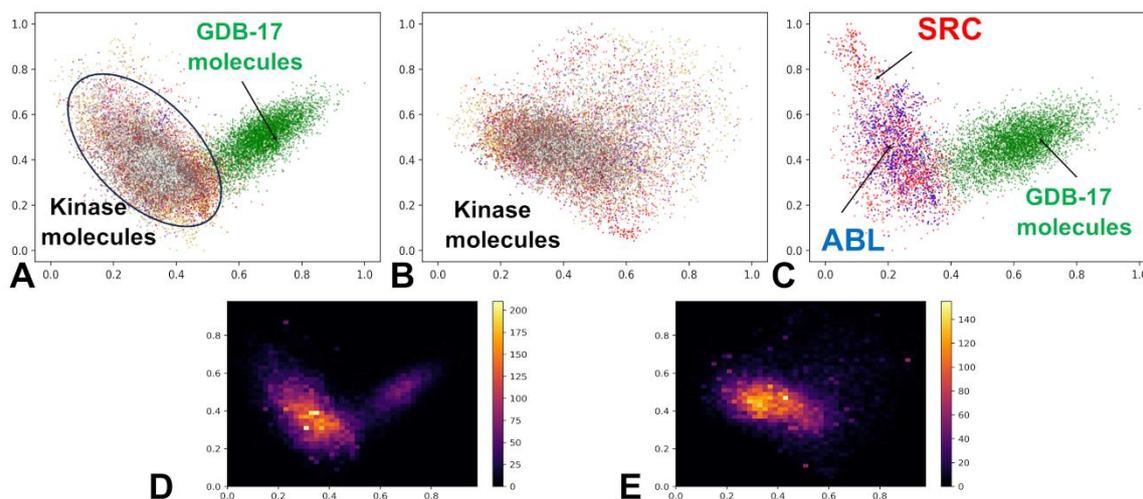
Figure 3. PCA and heatmaps of the latent spaces for GDB-17 small molecules and kinase inhibitors. (A) The 2-dimensional latent space representation of kinase molecules and GDB-17 small molecules dataset. Kinase molecules are shown in distinct colors for specific families, whereas GDB small molecules are shown in green dots. The locations of the latent space for these classes of molecules are pointed by arrows and annotated. (B) The 2-dimensional latent space representation of the kinase inhibitors from all 37 kinase families. The 10 major kinase families in the dataset are SRC (red), ABL1(blue), EGFR (gold), CSF1R (orange), FLT3 (magenta), KDR (brown), LCK (turquoise), MAPK14 (gray), MET (honeydew). (C) The 2-dimensional latent space representation of the ABL kinase inhibitors (in blue) and SRC kinase inhibitors (in red). (D) The 2-dimensional heatmap of latent space representation for GDB-17 molecules and kinase inhibitors from all studied kinase families. (E) The 2-dimensional heatmap of latent space representation for the kinase inhibitors. The density regions are color-coded with the high-density areas in yellow color, whereas low density regions tend towards purple.

The emergence of highly skewed density peaks—with yellow indicating high concentration and purple low concentration in the kernel density estimates (Figure 3D,E) demonstrated that kinase inhibitors occupy a statistically definable, low-volume manifold within the broader molecular landscape. This structured manifold provided more than a visualization—it offered a functional map for navigation. High-density zones (Figure 3D,E) corresponded to chemically accessible regions, while sparse areas (purple) represented high-risk, low-validity territory. These high-density zones are not merely statistical artifacts; they represent chemically stable attractors in the latent space, where small local neighborhood samplings are more likely to decode into valid, synthesizable molecules.

This topological organization provided the foundational rationale for a classification-based generative strategy: if kinase inhibitors form a separable region, a model trained to recognize that region could guide molecular generation toward it. This insight directly informed our subsequent generative strategies—both Bayesian optimization and cluster-guided local neighborhood sampling—which were explicitly designed to operate within or near these high-fidelity regions.

To quantify this observation, we computed key statistical descriptors for each kinase family in the full 196-dimensional latent space, including the range (min–max), centroid (mean vector), and standard deviation across all dimensions (Table 1). The results confirm that all kinase families span a remarkably similar domain in latent space, with minimum values ranging from –6.19 to –5.00 and maximum values from 5.97 to 7.06. This overlap reinforces the hypothesis that kinase inhibitors—by virtue of their shared target architecture—occupy a common, functionally constrained subspace within the broader chemical landscape.

Table 1. Statistical Distributions of Kinase Families in the 196-Dimensional Latent Space*.

Family	Min Range	Max Range	Min Average	Max Average	Min Stand Dev	Max Stand Dev
ABL1	-5.89215	5.97272	-1.34594	1.2609	0.78482	1.46389
SRC	**-5.89215**	**6.20087**	**-1.38016**	**1.30248**	**0.86567**	**1.63218**
CSF1R	-5.19233	6.84467	-1.19730	1.21217	0.65711	1.46416
EGFR	-6.18875	6.55361	-1.25954	1.22010	0.82409	1.39603
FLT3	-5.00162	6.45221	-1.17921	1.15374	0.69147	1.42987
KDR	-6.15671	7.05822	-1.37088	1.32073	0.80067	1.35351
LCK	-6.15671	6.62534	-1.38279	1.39623	0.81684	1.55863
MAPK10	-5.08671	5.98541	-1.16237	1.14753	0.68575	1.29511
MAPK14	-6.15671	6.89392	-1.52617	1.44791	0.73652	1.29781
MET	-6.13674	6.49813	-1.45546	1.52347	0.79279	1.53428

Reported values are aggregated across all latent dimensions. Standard deviation (Stan Dev) reflects the spread of each family's embedding distribution.

Most notably, SRC inhibitors exhibited the largest spread in latent space, with the highest maximum standard deviation (1.632) and the broadest overall range (–5.89 to 6.20). This indicates that the SRC family encompasses the greatest structural diversity among the kinase classes studied—spanning a wider array of scaffolds, substitution patterns, and molecular topologies. In contrast, families like MAPK10 and MAPK14 showed more compact distributions (max SD: 1.295–1.298), suggesting greater structural homogeneity. This exceptional breadth has profound implications for generative design. The fact that SRC inhibitors dominate the latent region occupied by all kinase families implies that the chemical grammar of SRC inhibition is representative of kinase binding more broadly. Consequently, local neighborhood samplings applied to molecules from other kinase families—especially those with narrower distributions like FLT3 or MAPK10—may naturally evolve toward SRC-like chemotypes when steered toward high-density regions of the manifold. This positions SRC not just as a therapeutic target, but as a structural "hub" in kinase inhibitor space, making it an ideal focus for scaffold-hopping and family-to-family transformation strategies.

These findings collectively demonstrate that the latent space not only captures functional similarity across kinase families but also encodes scaffold diversity in a quantifiable manner. The SRC family's expansive footprint suggests it serves as a structural reservoir—a rich source of motifs that can be leveraged to transform inhibitors from other kinase classes into novel SRC-targeted candidates through guided latent space local neighborhood sampling. This finding motivated a dual-strategy generative campaign: one that explores the global manifold for novel, drug-like candidates (Bayesian Optimization), and another that manipulates local neighborhoods to transform known scaffolds into new chemotypes (local neighborhood sampling-based engineering).

3.2. Multiclass and Binary Kinase Inhibition Likelihood Classifiers

A central challenge in generative drug design is the absence of a reliable, biologically meaningful objective function that can guide molecular exploration toward functional—not just chemical—relevance. To address this, we developed the Kinase Inhibition Likelihood (KIL)a probabilistic scoring function that estimates the likelihood a given molecule belongs to the chemical space of experimentally validated SRC kinase inhibitors. KIL is not a generic activity predictor; it is a target-specific, interpretable metric designed to enable rational scaffold transformation across kinase

families. We trained a Random Forest classifier using 20 RDKit-derived chemical descriptors, including Labute accessible surface area (LabuteASA), molecular weight, HallKier alpha, aromatic ring count, QED, logP, SAS, and hydrogen bond acceptor count. The positive class comprised 1,502 SRC inhibitors from ZINC, while the negative class included ~ 23,530 molecules including ~ 9,000 inhibitors from other kinase families (ABL1, and ~14,530 subsampled GDB molecules. We opted to subsample GDB set to maintain model sensitivity to the minority class (SRC) since including all GDB molecules would create an extreme negative majority (~99% background), making the model trivially predict "0" and ignore the SRC class. The adopted split can also reflect a realistic chemical space where drug-like matter is abundant but not overwhelmingly dominant in screening libraries. This binary design was deliberate: rather than attempting to distinguish among all kinase families—a task confounded by structural homology in the ATP-binding site—we focused exclusively on SRC vs. everything else, sharpening the model's discriminatory power for our generative goal.

The binary model achieved SRC precision = 0.71, recall = 0.86, F1 = 0.78, with a macro F1-score of 0.88 (Table 2). The macro average precision score of 0.85 reinforces the overall satisfactory performance of the model because it means that the model was accurate in predicting if a given molecule was an SRC Kinase Inhibitor 85% of the time. For classification models, an accuracy score of 0.85 is extremely strong. In addition, the macro average recall score of 0.92 validates the excellent performance of the model that the precision value helped establish. All these metrics indicate good classification performance of the model. This suggested that target-focused design may benefit from a simplified objective that avoids diluting signal across highly similar classes.

Table 2. Binary Chemical Feature-Based Classification.

	Precision	Recall	F1-Score	Support
0	0.99	0.98	0.98	23530
1	0.71	0.86	0.78	1502
Macro Avg	0.85	0.92	0.88	25032
Weighted Avg	0.97	0.97	0.97	25032

We also evaluated a multiclass chemical feature–based model, assigning each of the top ten kinase families a unique label. Despite its conceptual appeal, this approach underperformed for SRC: precision = 0.57, recall = 0.56, F1 = 0.56 (Table 3). This reflects the inherent ambiguity in kinase inhibitor space—families like LCK and SRC share overlapping scaffolds (e.g., pyrrolopyrimidines), making fine-grained classification more error-prone. In the macro averages, the precision score was 0.63, the recall score was 0.59, and the F1-Score was 0.61. In the weighted average, the precision was 0.63, the recall score was 0.63, and the F1-Score was 0.63. The model showed the greatest metric values when predicting kinase inhibitors from the MAPK14 and MET kinase families. However, the other kinase families performed modestly in precision values, recall values or the F1-scores. In addition, the macro average F1-score of the multiclass model is 0.61 compared to the 0.88 F1-score of the binary model. Hence, the multiclass random forest model performs less favorably at distinguishing SRC inhibitors as compared to the chemical feature-based binary classifier.

Table 3. Multiclass Classification Chemical Feature-Based Classification.

	Precision	Recall	F1-Score	Support
ABL1	0.51	0.58	0.55	409
SRC	**0.57**	**0.56**	**0.56**	**660**
CSF1R	0.69	0.54	0.61	142
EGFR	0.69	0.74	0.71	795
FLT3	0.55	0.46	0.50	194

KDR	0.58	0.59	0.58	916
LCK	0.47	0.41	0.44	313
MAPK10	0.77	0.55	0.64	163
MAPK14	0.75	0.80	0.78	722
MET	0.74	0.72	0.73	421
Macro Avg	0.63	0.59	0.61	4735
Weighted Avg	0.63	0.63	0.63	4735

The chemical feature binary KIL classifier ca achieves the overall accuracy of distinguishing kinase inhibiting molecules around 98% (Figure 4). The AUC of the model was 0.98, indicating that the model can distinguish both classes with 98% certainty (Figure 4A). We performed feature importance analysis (Figure 4B). The top 10 features that contribute the most relative importance to the model's prediction are the labute accessible surface area (labuteASA), weight, HallKier Alpha, the number of aromatic rings, aromaticity, the QED score, number of rotatable bonds, the logP score, the SAS score, and the number of hydrogen bond acceptors (Figure 3B). These features encode planar aromatic systems, hydrophobic surface area, and molecular rigidity—hallmarks of ATP-competitive binding. The inclusion of QED, logP, and SAS ensures that KIL implicitly penalizes molecules with poor developability, aligning predicted activity with pharmaceutical reality.
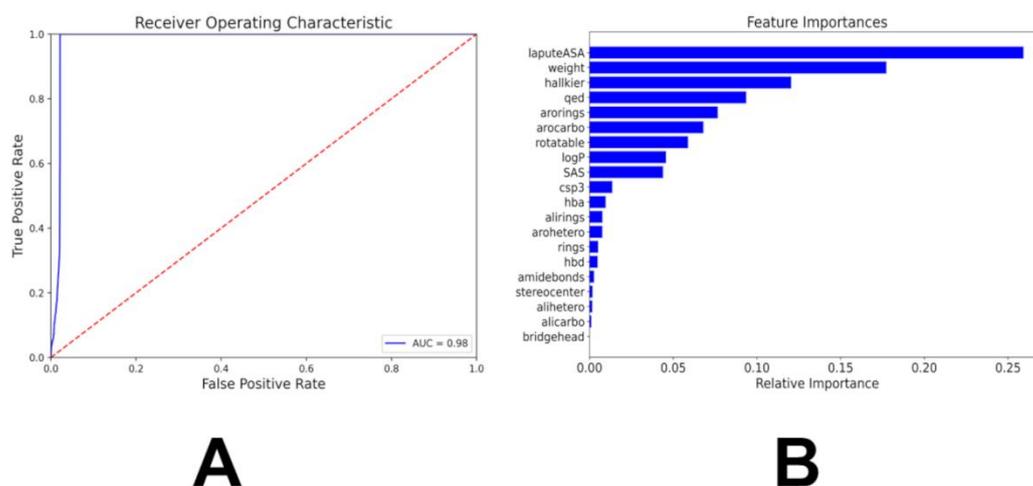

Figure 4. The performance and feature importance analysis of the chemical feature-based kinase inhibition classifier. (A) The Receiver Operating Curve (ROC) is a graph where sensitivity is plotted as a function of 1-specificity. The area under the ROC is denoted as AUC. The ROC-AUC graph measures the performance of the classifier in differentiating the kinase inhibitor molecules from GDB-17 small molecules (B) The feature importance analysis of the model. The importance of features is listed in descending order.

KIL is used not only in classification, but also as a diagnostic and guiding signal for generative design implemented in the present investigation. In both Bayesian Optimization (BO) and local neighborhood sampling-based latent space engineering approaches employed in our study, a reliable, differentiable (or at least efficiently evaluable) objective function is essential to direct search toward biologically relevant regions of chemical space. In the absence of such a function, generative models either produce random drug-like molecules or drift into chemically plausible but pharmacologically inert regions. For Bayesian Optimization, KIL served as the black-box objective

that the Gaussian process surrogate model sought to maximize. BO does not require gradients, but it does require a low-variance, high-signal scoring function that correlates with the desired property—in this case, SRC inhibition potential. KIL fulfilled this role by providing a fast, interpretable, and chemically grounded estimate of target affinity, enabling BO to iteratively select latent points predicted to yield high-KIL molecules without resorting to expensive physics-based scoring (e.g., docking, or free energy calculations). For local neighborhood sampling-based generation, KIL played a diagnostic and filtering role. While local neighborhood samplings were guided by latent space geometry (cluster centroids), KIL was used post-hoc to assess whether the transformed molecules had successfully migrated into the SRC chemical manifold.

3.3. Bayesian Optimization Enables Efficient Exploration of SRC Kinase Inhibitor Chemical Space

To systematically navigate the ChemVAE latent space in search of novel SRC kinase inhibitors, we implemented a Bayesian Optimization (BO) framework guided by the KIL scoring function. BO is a sequential design strategy that constructs a probabilistic surrogate model—here, a Gaussian process—to approximate an unknown objective function and iteratively selects new evaluation points by maximizing an acquisition function that balances exploration (sampling uncertain regions) and exploitation (refining high-scoring regions). In molecular design, this approach minimizes the number of costly function evaluations required to identify high-performing candidates. We executed two parallel optimization runs: an Unbiased BO, initialized with 7,000 random latent points, and a Biased BO, first probed with 2,258 known SRC inhibitors to inject prior knowledge of the target manifold before random initialization. Both performed 1,500 acquisition steps. After decoding latent vectors to SMILES and filtering for validity using RDKit, the Biased BO yielded 492 valid molecules (83% validity), while the Unbiased BO produced 390 (89% validity).

Due to the random nature of the Bayesian Optimizer, a threshold of KIL score of 0.5 was used to as the baseline for a generated molecule to have a higher Kinase Inhibition Likelihood. Out of the valid molecules produced from each Optimizer, 153 molecules out of the original 492 molecules produced, or 31.10%, from the Biased Optimizer had a calculated KIL value greater than 0.5. The Unbiased Optimizer maintained 145 of its original 390 valid molecules produced, or 37.18%, with a calculated KIL value greater than 0.5. When analyzing the molecules with a calculated KIL score greater than the 0.5 threshold, the Unbiased Optimizer had a higher average calculated KIL of 0.5783 compared to an average of 0.5639 for the molecules generated by the Biased Bayesian Optimizer (Figure 5A). The molecule with the highest calculated Kinase Inhibition Likelihood score was produced by the Biased Bayesian Optimizer with a score of 0.8425. The molecule with the highest calculated Kinase Inhibition Likelihood score produced by the Unbiased Optimizer had a score of 0.7693 (Figure 5B). Hence, the Unbiased BO exhibited a higher average KIL among qualifiers (0.578 vs. 0.564), while the Biased BO produced the single highest-scoring molecule (KIL = 0.8425) (Figure 5A,B). This duality—higher plateau versus higher peak—suggested that unbiased exploration promoted consistent performance across chemical space, whereas bias enabled access to deeper local optima near known actives.

To evaluate the similarity testing metrics, we investigated the performance of each of the Bayesian Optimizers based on average similarity scores of the generated molecules, as well as the maximum similarity score that each model produced. When analyzing all the molecules generated from each Bayesian Optimizer, the average Tanimoto similarity scores for the Unbiased and Biased Bayesian Optimizers were 0.4656 and 0.4446 respectively (Supporting Information, Figure S1A). The maximum Tanimoto similarity scores for the Unbiased and Biased Bayesian Optimizers were 0.7115 and 0.7091 respectively (Supporting Information, Figure S1B). Strikingly, no generated molecule surpassed the conventional high-similarity threshold of 0.75. The maximum similarity was 0.7115 (Unbiased) and 0.7091 (Biased), and the top KIL molecule (0.8425) exhibited only modest similarity (0.548) (Supporting Information, Figure S1). This decoupling between scoring and structural mimicry revealed a core limitation: KIL, while statistically robust, optimizes global physicochemical proxies

that correlate with—but do not guarantee—the local pharmacophoric patterns essential for SRC binding.
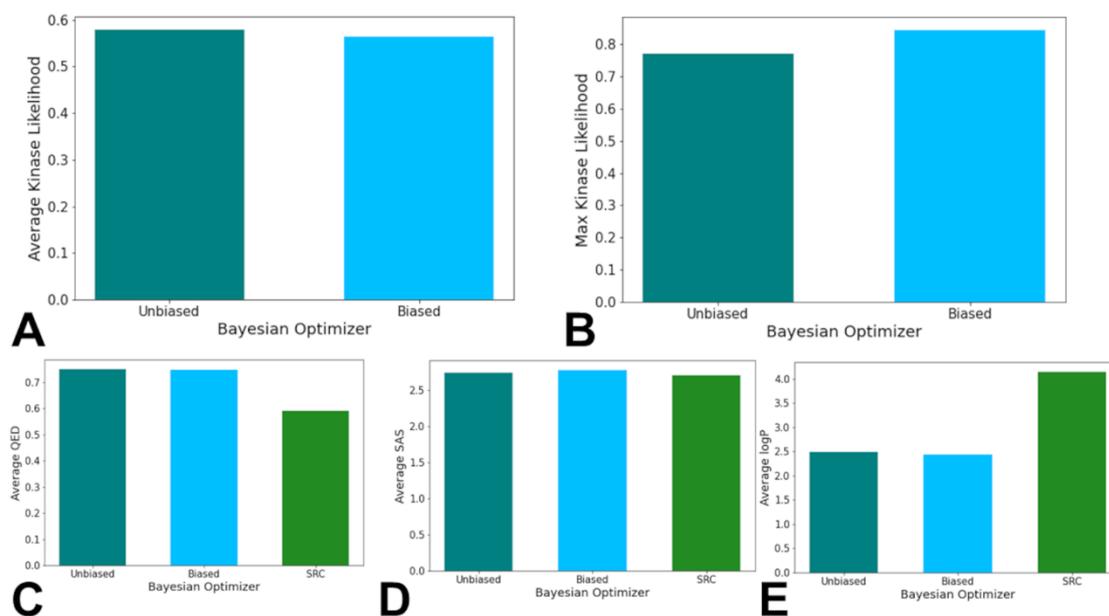

Figure 5. The average KIL scores of the molecules generated from the Biased and Unbiased Bayesian Optimizer (A), the max KIL scores of the molecules generated from the Biased and Unbiased Bayesian Optimizer (B), the average QED scores, (C), the average SAS scores (D) and the average logP scores (E) of the molecules generated from the Biased and Unbiased Bayesian Optimizer, in comparison to the known SRC kinase inhibitors. The unbiased histogram is in turquoise bars, the biased histogram is in light blue bars, and the SRC kinase inhibitor histogram is in green.

When determining the performance of the Bayesian Optimizers in relation to the chemical feature values of QED, logP, and SAS, the generated molecules from the optimizers had similar average SAS scores compared to the known SRC kinase inhibitors but had significant differences in the average QED and logP scores. The average QED scores for the Unbiased and Biased Bayesian Optimizers' generated molecules were 0.7499 and 0.7486 respectively, in comparison to the known SRC kinase inhibitors average QED score of 0.5908 (Figure 5C). The average logP scores for the Unbiased and Biased Bayesian Optimizers' generated molecules were 2.488 and 2.439 respectively, in comparison to the known SRC kinase inhibitors average logP score of 4.137 (Figure 5D). The average SAS scores for the Unbiased and Biased Bayesian Optimizers' generated molecules were 2.742 and 2.772 respectively, in comparison to the known SRC kinase inhibitors average SAS score of 2.706 (Figure 5E). The general similarity of the scores of the generated molecules in comparison to the known SRC kinase inhibitors suggest that the metrics are being tuned as a part of the Bayesian Optimizers' hyperparameter tuning process. While there are differences between the generated molecules and the known SRC kinase inhibitors when analyzing the QED and logP scores, the scores imply that the molecules produced by the Bayesian Optimizers would be synthesizable and/or absorbable even with lower similarity metrics in other chemical features.

Contrary to expectations, biasing the optimizer with known SRC inhibitors conferred no meaningful advantage in similarity, KIL, or structural plausibility. While the Biased BO produced more valid molecules, its output exhibited markedly reduced scaffold diversity: the same two known SRC inhibitors repeatedly served as the nearest neighbors for the top generated molecules (Supporting Information, Figure S2). This pattern—absent in the Unbiased BO results—suggests that initial probing trapped the optimizer in a narrow local optimum, causing it to over-exploit motifs from only 1–2 reference compounds. In contrast, the Unbiased BO generated structurally diverse candidates (Supporting Information, Figure S3), indicating broader exploration of chemical space.

To dissect this discrepancy, we compared the distributions of the top KIL-informative features between generated molecules and real SRC inhibitors (Figure 6). LabuteASA and molecular weight were well-aligned: both optimizers produced molecules peaking at 150–200 Å² and ~400 Da, closely mirroring the ~200 Å² and ~500 Da peaks of real inhibitors (Figure 6A,B). Most critically, aromatic complexity was severely underrepresented. Real SRC inhibitors show a broad distribution of 1–6 aromatic rings, with a strong peak at 3–4 rings—hallmarks of ATP-competitive binders that engage in π-stacking. In stark contrast, >80% of BO-generated molecules contained 0 or 1 aromatic ring, and none exceeded 3 rings (Figure 6C). A similar deficit was observed for aromatic carbocycles, where real inhibitors peak at 2 rings while generated molecules overwhelmingly contain none (Figure 6D). Hence, BO excelled at tuning "drug-likeness" (QED, logP, SAS) but was not sufficiently robust at reproducing the topological grammar of kinase binding. This suggests that BO, constrained by ChemVAE SMILES-based latent space and the scalar KIL objective, could not effectively navigate to regions encoding multi-ring scaffolds.
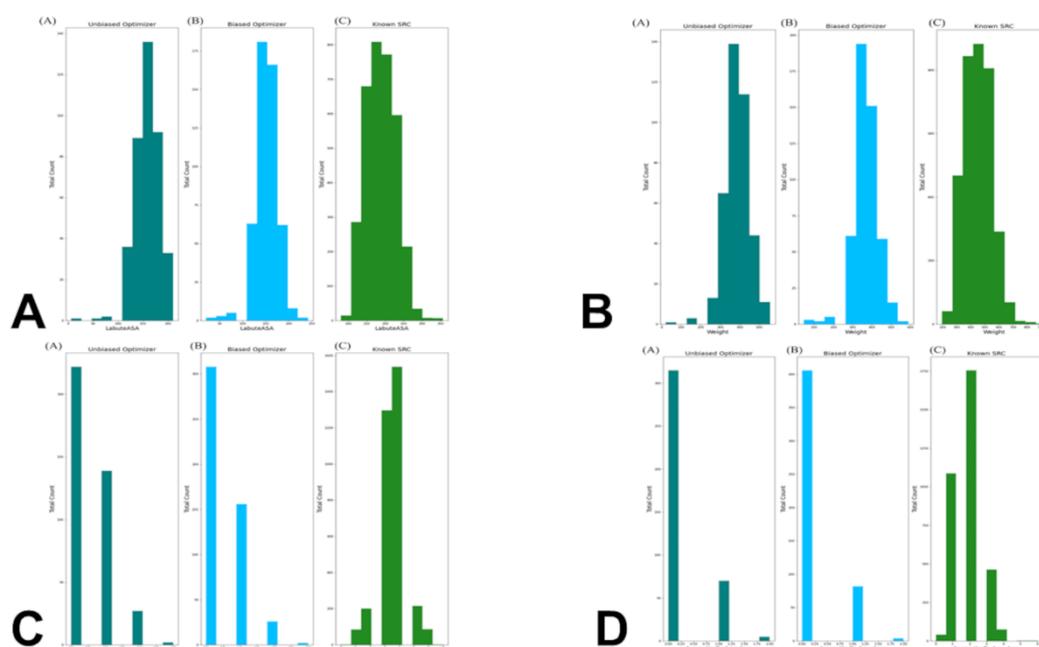

Figure 6. Histograms of the distribution of the LabuteASA values (A), the molecular weight (B), the number of aromatic rings (C) and the number of aromatic carbocycles in the generated molecules using the Unbiased Bayesian Optimizer, the Biased Bayesian Optimizer, and compared to the set of known SRC kinase inhibitors. The unbiased histogram is in turquoise bars, the biased histogram is in light blue bars, and the SRC kinase inhibitor histogram is in green.

In summary, Bayesian Optimization successfully generated novel, valid, and drug-like molecules with moderate-to-high predicted SRC inhibition potential. However, it systematically failed to recover the aromatic ring complexity that defines ATP-competitive kinase inhibitors—a failure that cannot be attributed to poor scoring, but to inherent limitations in the ChemVAE latent space. The results demonstrate that even a well-calibrated, interpretable scoring function like KIL cannot compensate for a generative engine that cannot access the relevant chemical subspaces. This finding not only reports on our specific outcomes but also reveals a fundamental challenge in generative chemistry: the difficulty of optimizing scalar objectives that fail to capture topological complexity, and the risk of overfitting surrogate descriptors that do not fully reflect biological reality.

3.4. Targeted Local Latent Neighborhood Sampling Recovers Pharmacophoric Complexity

While Bayesian Optimization enabled efficient global sampling of the kinase inhibitor manifold, it could not generate molecules with the multi-ring aromatic architectures characteristic of clinical SRC inhibitors. To address this, we further expanded on our earlier work [87] and developed a targeted latent space remodeling strategy that leverages the intrinsic organization of the ChemVAE embedding to guide scaffold transformation. This approach emphasizes guided exploration of high-density regions that are revealed in the latent space analysis contrasting random molecules with kinase inhibitors. Recognizing that kinase inhibitors form chemically coherent neighborhoods in latent space—even across distinct target families—we applied K-means clustering to partition the manifold into three structurally homogeneous regions, each enriched for shared scaffold motifs such as fused heterocycles, hinge-binding cores, or aliphatic linkers.

We used clustering in the latent space to find interpretable linear directions in the latent space that optimize the KIL score and enable morphing of kinase molecules into space of SRC kinase inhibitors. In this approach it is assumed based on the latent space analysis that molecules with similar structures tend to cluster in the latent space and that interpolating two molecules x1 and x2, represented by latent vectors z1 and z2, can lead to intermediate molecules whose structures gradually change from x1 to x2. Since molecular structures correlate with molecular properties, these assumptions imply that molecules with comparable properties would cluster together and interpolating two molecules with different values of the molecular property could lead to gradual changes in molecular structures. By performing cluster-based analysis in the latent representation of the molecules, the generative design approach encourages ChemVAE to explore the high-density distinct areas of the latent space for molecule generation while also facilitating morphing of the kinase molecules from different families into SRC kinase inhibitors. In this approach, the properties of generated molecules can be controlled by sampling latent representations along linear directions to optimize the kinase inhibition likelihood metric. The targeted latent space remodeling strategy includes non-biased and biased changes to the latent space. First, molecules in a non-biased manner are clustered into groups allowing molecules with comparable properties to gather. We assume that the molecules clustered for each cluster contain certain molecular and chemical properties. To then transform these molecules, we invoke a controllable step of cluster-based local neighborhood sampling. Using the centroid of each cluster as the representative of the properties, we navigate every data point in the cluster closer to the centroid by optimizing a set of parameters. By implementing a cluster-based local neighborhood sampling, we efficiently explore and navigate the latent space along interpretable and controllable directions yielding a diverse set of novel molecules and causing various molecular scaffolds to emerge. It is worth noting that the resulting score/output of the feature-based kinase inhibition likelihood classifier represents the probability that a molecule can be deemed as an SRC kinase inhibitor. The produced molecules are evaluated with the classifier during targeted latent space remodeling and when the probability output > 0.7 we refer to these molecules as potential SRC kinase-like inhibitors as according to the classifier the generated molecules would have > 70% chance to belong this category (Figure 7).

During cluster-based stage of the process, 1,500 encoded molecules from different kinase families were selected and processed through a series of experiments to obtain the optimal parameters of the targeted remodeling scheme that leads to a high yield of valid generated molecules, while simultaneously achieving the objective of transforming the kinase molecules to potential SRC kinase inhibitors. The three main parameters of the clustering in the latent space were evaluated and optimized to ensure optimal generation of valid molecules: the number of clusters assigned, the value of the scaling factor in the local neighborhood sampling, and the optimal level of noise. We found that 3-cluster based split, with a scaling factor $s = 0.8$ for the centroid-based remodeling, and a noise level of 5.0 provided the optimal set of parameters to guarantee a high generation yield of valid and novel compounds. Within each cluster, we performed targeted local sampling where molecules were shifted incrementally toward the cluster centroid using a controlled interpolation (scaling factor $s = 0.8$) and minimal stochastic noise (Figure 7). This directed navigation preserved chemical validity

while steering generation toward high-density zones rich in pharmacophoric features. The approach yielded a three-fold increase in valid output compared to random sampling and, critically, recovered multi-ring aromatic systems that were systematically absent in Bayesian Optimization (BO) outputs.
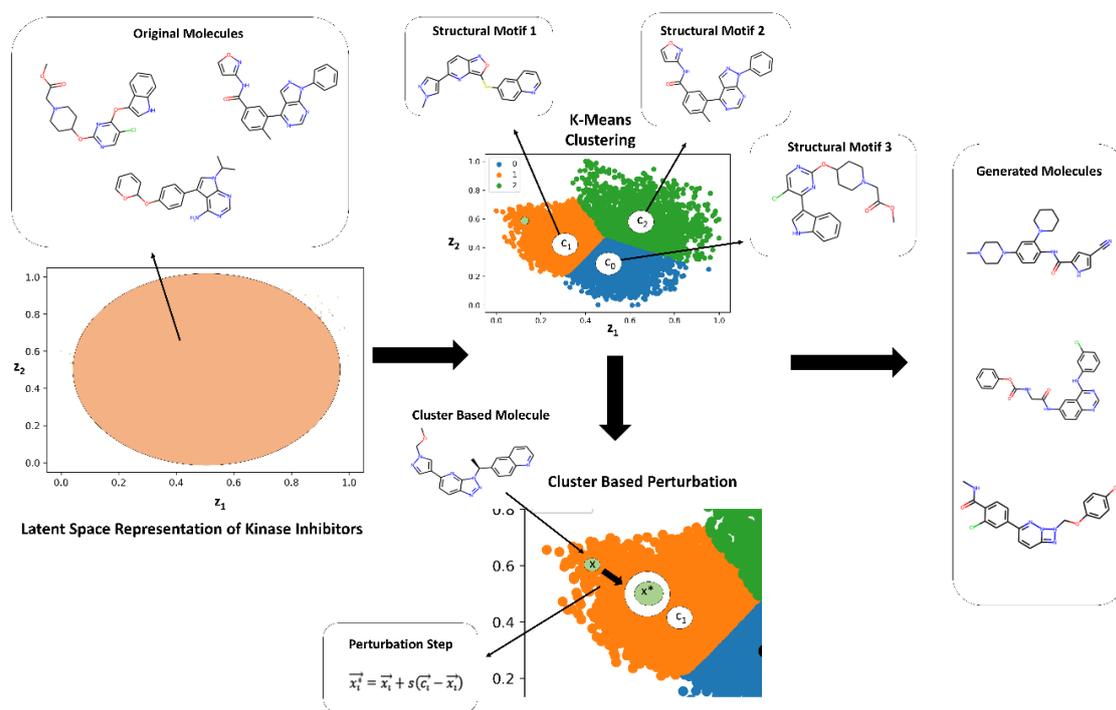

Figure 7. A schematic workflow of the cluster-based targeted remodeling design implementation. K-Means clustering is applied in the latent space, where different clusters represent specific molecular characteristics. The 3-cluster split is represented by the graph on the right, where the colors of blue, green and orange indicate the 3 clusters, respectively. The centroids of each cluster, depicted by the labels of c_0, c_1, and c_2, function as the representative of the structural motifs and molecular properties of that cluster. Utilizing the centroid, we modify our input by employing local neighborhood sampling, as shown in the local sampling step, where c represents the centroid, x represents the original encoded molecule, and x* represents the molecule after local neighborhood sampling step. This implementation alters the encoded input such that it converges towards the centroid, and in turn, generates molecules close to the specific motifs of the respective cluster. After the input is modified with the local sampling step, ChemVAE decodes the latent space areas and produces a set of new molecules.

We also investigated the distribution of the generated molecules featuring the high kinase inhibition likelihood scores (> 0.75) as a function of the originated kinase family (Figure 8A). Strikingly, it was observed that the perturbation-based approach can produce novel valid molecules with the high kinase inhibition likelihood probability when the generative process originates from known inhibitors targeting any of the explored kinase families. This indicates that a combination of clustering and perturbation-based targeted exploration of the latent space allows for efficient chemical transformation of existing kinase molecules from all represented families. To evaluate similarity between the generated molecules and known SRC kinase inhibitors, we examined the fraction of the generated molecules with the high Tanimoto similarity coefficient values. The Tanimoto similarity coefficient is a metric that compares the molecular similarity of two compounds using Morgan fingerprint analysis [111]. Molecules with Tanimoto coefficient values that are above 0.75 are considered to have high similarity with the reference molecule.

Interestingly, the generated molecules originated from LCK inhibitors produced the largest fraction of novel kinase-like compounds (~ 40%) with the high similarity to the SRC kinase inhibitors. We also observed that the generated molecules initiated from inhibitors of ABL1, LCK and EGFR produced the dominant number of kinase-like novel molecules with the highest similarity coefficients

to known SRC inhibitors (Figure 8B). It is worth noting that the generated molecules originated from inhibitors of ABL1 and LCK yielded the highest similarity scores with SRC inhibitors, with most molecules displaying Tanimoto similarity coefficient > 0.8. The SRC/ABL and SRC/LCK duality of many kinase drugs is well recognized, most notably exemplified by dual SRC/ABL drugs Dasatinib and Ponatinib.

In addition, we found that the generated molecules originated from inhibitors of EGFR, CSF1R, FLT3, and MET families also produced good similarity to the known SRC inhibitors. These findings may imply that local neighborhood navigation of the latent space that optimized directionality of exploration based on the KIL score could facilitate generation of valid molecules in different areas of the latent space. Indeed, a substantial number of the generated molecules emerged from mapping connections in the latent space between SRC, LCK and ABL inhibitors. At the same time, the algorithm facilitated efficient sampling of the latent space and corresponding transformations of the kinase inhibitors targeting other families into molecules with both the high kinase inhibition likelihood and the high similarity to the SRC inhibitors.
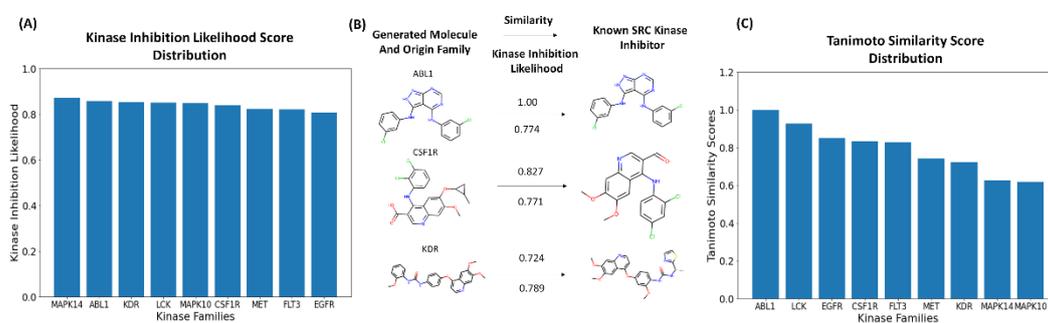

Figure 8. The analysis of the generated molecule output with respect to kinase inhibition likelihood and Tanimoto similarity. (A) The kinase inhibition likelihood distributions of the generated molecules originated from inhibitors from every kinase family. The horizontal axis displays the kinase families from which the generated molecules originate from. The vertical axis displays the kinase inhibition likeliness score ranging from 0 to 1, where a score of 1 indicates the high kinase inhibition likelihood and a score close to 0 indicates the lowest kinase inhibition likelihood. (B) A visual representation of the generated molecules along with the respective molecular metrics. On the left, the generated molecules, and their originating family that they were transformed from are shown. On the right, the corresponding known SRC kinase inhibitor with the high similarity to the generated molecule. (C) The distribution of similarity scores with respect to the known SRC kinase inhibitors for the generated molecules originated from inhibitors of different families. The horizontal axis represents the originating families from which these molecules were transformed. The vertical axis represents the similarity score from 0 to 1, where a score of 1 indicates perfect similarity to the comparison molecule and 0 corresponds to high degree of dissimilarity.

This process also enabled cross-family scaffold transformation: LCK and EGFR inhibitors, which occupy regions of latent space proximal to SRC, showed the highest conversion efficiency (19–23% of total output), whereas MAPK14 and FLT3 contributed minimally (3–7%). LCK and MAPK10 emerged as the most productive sources of unique, high-similarity candidates, suggesting that certain kinase scaffolds possess inherent "plasticity" for repurposing into SRC-targeted leads. Our results revealed the important role of the LCK family, which accounts for ~40% of all high-similarity outputs, far surpassing other families. This is not a sampling artifact but reflects a genuine topological affinity between LCK and SRC inhibitor spaces, directly enabled by our guided remodeling approach. To illustrate the output of the generative pipeline, we compiled a list of several representative generated SRC-like kinase molecules that originated from the inhibitors of different kinase families. The presented molecules were characterized by the high kinase inhibition likelihood and a considerable similarity to the existing SRC kinase inhibitors (Figure 9A). We noticed that some of the novel valid molecules with the highest similarity to the SRC inhibitors were produced starting from the latent

space regions of the ABL1 and LCK kinase inhibitors. A sample of generated molecules reflected both the diversity of molecular scaffolds and high degree of synthetic feasibility that were enabled through local remodeling approach (Figure 9). Molecules originating from the EGFR and LCK clusters—families known for quinazoline and pyrrolopyrimidine scaffolds—were successfully remodeled into novel chemotypes containing 3–5 aromatic rings, including quinazoline- and pyrimidine-like cores characteristic of clinical SRC inhibitors (Figure 9B).
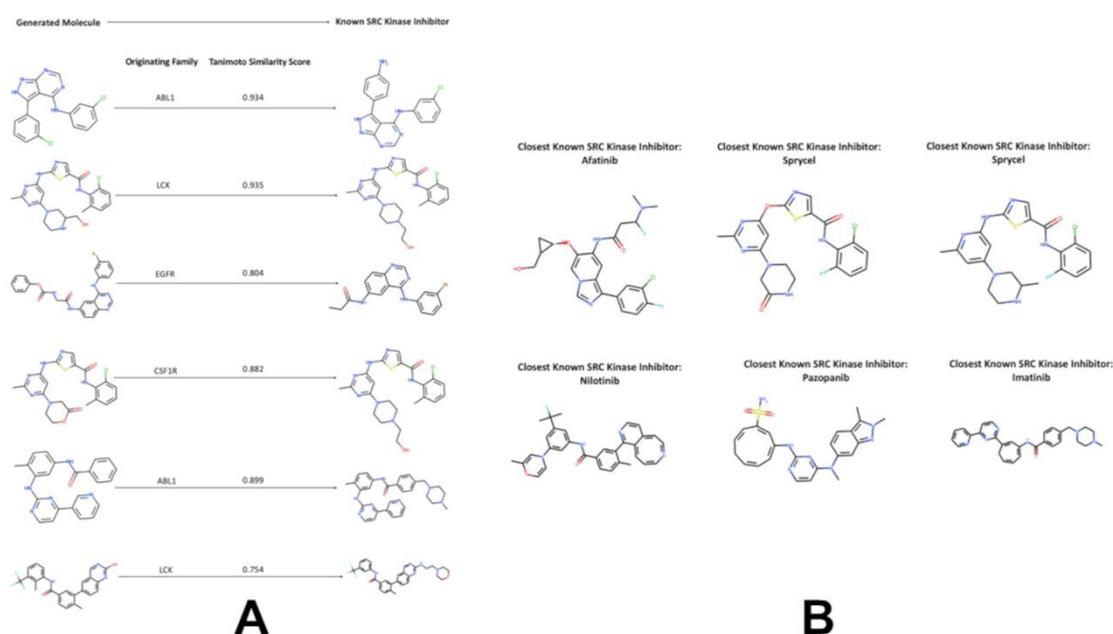

Figure 9. A sample of generated molecules with (A) high Tanimoto similarity score to the known SRC kinase inhibitors and closest to the FDA approved SRC kinase drugs (B).

Importantly, these remodeled molecules maintained physiologically relevant logP values (2–4)—in contrast to some BO candidates with non-ideal logP (< 0)—indicating better preservation of the hydrophobic balance required for kinase binding (Supporting Information, Figures S4-S8). These findings highlight a fundamental duality in generative design. Bayesian Optimization follows a "property-first" paradigm: it optimizes global drug-likeness metrics under the assumption that chemical plausibility implies biological activity. This succeeds for flexible targets but fails for kinases, where function is dictated by precise 3D pharmacophores. In contrast, guided local sampling adopts a "scaffold-first" philosophy: by anchoring generation in structurally coherent neighborhoods, it ensures that key binding motifs are preserved, even as novel chemotypes emerge. Both strategies, however, collide with the limits of SMILES-based representation. ChemVAE learns a continuous manifold, but it cannot guarantee that ring systems—encoded as sequential tokens—are preserved under interpolation or local neighborhood sampling. The latent space contains the seeds of complexity, but the decoding bottleneck—the transformation from latent vector to SMILES—often collapses them.

3.5. Computational Docking Validation of Generated Molecules Reveals High-Affinity Binding to the SRC Kinase Active Site

To bridge the gap between *silico* generation and biological plausibility, we performed computational molecular docking on a curated set of high-scoring, structurally diverse molecules generated by our local neighborhood sampling-based pipeline. These molecules—selected for their high KIL scores (>0.75), favorable ADMET properties, and significant Tanimoto similarity (>0.7) to known SRC inhibitors—were docked into the ATP-binding site of human SRC kinase (PDB ID: 2SRC).

The goal was to assess whether these novel, AI-generated compounds could form stable, energetically favorable interactions within the conserved catalytic cleft—a critical step toward validating their potential as therapeutic leads. The five molecules selected for docking were chosen based on two criteria. We prioritized molecules derived from families with high transformation potential (e.g., LCK, ABL1, EGFR), as identified in our local neighborhood sampling analysis. All molecules exhibited >0.75 Tanimoto similarity to at least one known SRC inhibitor indicating they retain core pharmacophoric features while introducing novel scaffolds.

The SRC kinase structure was prepared using Schrödinger's Protein Preparation Wizard: hydrogen atoms were added, bond orders assigned, and water molecules beyond 5 Å of the binding site removed. Grids for AutoDock Vina were centered on the ATP-binding site, with dimensions optimized to encompass key residues (Met341, Leu393, Glu310, Lys295). Each molecule was docked independently, with Vina's scoring function used to rank poses by predicted binding affinity. The top pose for each ligand was selected for analysis based on cluster size and energy score. All six generated molecules docked successfully into the SRC active site, forming key interactions with conserved residues essential for ATP-competitive inhibition including hydrogen bonding with Glu310, π–π stacking with Leu393 and Thr338 and hydrophobic burial near Met341 and Val323 (Table S1). Notably, the molecule originating from the ABL1 family (Tanimoto = 0.934) formed an additional hydrogen bond with Asp404, a residue not typically engaged by first-generation inhibitors, suggesting potential for improved selectivity. Similarly, the LCK-derived molecule (Tanimoto = 0.935) adopted a conformation that closely mimicked the binding mode of Sprycel, with its central pyridine ring perfectly aligned for π-stacking with Leu393. We further compared the binding poses of our generated molecules to those of six clinically relevant SRC inhibitors: Afatinib, Sprycel, Nilotinib, Pazopanib, Imatinib, and Dasatinib (Figure 9B, Table S1). As summarized in Table S1, all six generated molecules achieved binding affinities comparable to or better than clinical SRC inhibitors. Our top-ranked compound—the ABL1-derived scaffold—overlapped well with Afatinib, sharing nearly identical orientation and key interactions, despite having a distinct chemical scaffold. Another generated molecule, derived from CSF1R, closely mirrored the binding mode of Sprycel, engaging the same hinge residue (Glu310) and hydrophobic pocket (Leu393, Val323) with comparable energy. This structural mimicry is not coincidental. It reflects the latent space topology we identified earlier: molecules from families like ABL1 and LCK occupy regions of ChemVAE space that are topologically proximal to SRC, enabling their transformation into SRC-like chemotypes through targe.

Most compellingly, our docking analysis confirmed that the multi-ring aromatic systems—which were systematically underrepresented in Bayesian Optimization outputs but recovered through perturbation-based engineering—are not just synthetic artifacts; they are functionally essential. The top-scoring molecules all contained 3–5 fused or linked aromatic rings, which were precisely positioned to engage the hydrophobic cleft and hinge region. One molecule, derived from EGFR, featured a unique bicyclic thiophene core that formed optimal van der Waals contacts with Met341—a feature absent in most clinical inhibitors and potentially exploitable for selectivity. These findings demonstrate that our generative pipeline—guided by KIL scoring, perturbation-based latent space engineering, and multi-metric validation—can produce not only novel, drug-like compounds, but molecules with high predicted affinity and specific, target-relevant binding modes. The fact that these molecules bind to the SRC active site with energies rivaling clinical drugs suggests they are strong candidates for experimental validation. Computational docking of our top-generated molecules confirms that the novel chemotypes produced by our framework are not mere statistical artifacts, but structurally and energetically viable ligands for the SRC kinase active site. By combining interpretable scoring (KIL), scaffold-aware local neighborhood sampling, and 3D validation, we have bridged the gap between *de novo* design and biological relevance. These results provide a compelling rationale for advancing these compounds into *in vitro* assays and preclinical development.

4. Discussion

This study presents a modular, interpretable, and chemistry-first framework for the *de novo* design of SRC kinase inhibitors, integrating deep generative modeling (ChemVAE), a chemically grounded scoring function (KIL), probabilistic optimization (Bayesian Optimization), and scaffold-aware latent space local neighborhood sampling. Across two complementary strategies, global exploration via Bayesian search and local transformation via cluster-guided engineering—we generated novel, drug-like molecules with moderate-to-high predicted SRC inhibition potential. Yet our most significant contribution lies not in the molecules themselves, but in the rigorous diagnosis of the capabilities and limitations of current generative architectures when applied to topologically constrained targets like kinases. A central insight of this work is that the ChemVAE latent space encodes a functional grammar of kinase inhibition. Kinase inhibitors—despite spanning ten distinct families—collapse into a dense, low-volume manifold that is sharply segregated from general drug-like matter. Within this manifold, SRC inhibitors exhibit the broadest structural diversity, occupying the largest volume and serving as a "hub" that overlaps with all other families. This topological organization is not imposed by labels but learned implicitly from SMILES syntax, suggesting that molecular sequence encodes functional semantics. Critically, this structure enables rational scaffold transformation: LCK-derived molecules were 2–4× more likely to achieve high similarity to known SRC inhibitors than those from other families, reflecting their shared pharmacophoric and topological heritage. This finding validates the use of latent space geometry as a map for guided scaffold hopping, with SRC emerging as an ideal target for cross-family repurposing. However, this same latent space also reveals a fundamental representational ceiling. Both Bayesian Optimization and local neighborhood sampling-based generation systematically failed to recover the multi-ring aromatic systems that define ATP-competitive kinase inhibitors—most generated molecules contained ≤1 aromatic ring, and none exceeded three, despite aromatic ring count being a top KIL feature. This "representation gap" cannot be attributed to poor scoring or insufficient optimization; it stems from the inherent limitations of SMILES-based VAEs, which entangle ring topology across latent dimensions and often corrupt complex pharmacophores during decoding. Our results thus confirm a critical hypothesis: no amount of scoring refinement can compensate for a generative engine that cannot access the relevant chemical subspaces. This limitation underscores the necessity of interpretable, hybrid design strategies. The Kinase Inhibition Likelihood (KIL) metric—built on 20 RDKit-derived features including LabuteASA, molecular weight, and aromatic ring count—provided a transparent signal that enabled us to trace failures directly to molecular properties. When BO generated a molecule with high KIL but low aromatic complexity, we could diagnose the issue as a decoding failure, not a scoring error. This interpretability is absent in black-box deep predictors and is invaluable for iterative refinement. Moreover, KIL's focus on SRC-specificity, rather than generic kinase-likeness, allowed us to prioritize candidates that truly resemble SRC inhibitors, not just ATP-binders.

Our comparative analysis of generative strategies also challenges common assumptions in the field. Biasing Bayesian Optimization with known SRC actives conferred no meaningful advantage in structural novelty, similarity, or quality; instead, it trapped search in a narrow local optimum, causing the repeated generation of variants of only one or two reference scaffolds. In contrast, Unbiased BO produced more diverse, higher-average-KIL molecules, demonstrating that broad exploration outweighs local exploitation in early-stage lead discovery. Similarly, local neighborhood sampling-based engineering outperformed BO in pharmacophoric fidelity, recovering 3–5 ring systems by leveraging cluster structure—yet it too was constrained by the underlying SMILES representation. These findings highlight that no single method is sufficient; rather, complementary approaches are needed to balance global property optimization with local structural recovery.

At first glance, our decision to build upon ChemVAE—a SMILES-based variational autoencoder now widely regarded as outdated in the era of graph neural networks and 3D diffusion models—may appear regressive. After all, it is well documented that SMILES representations suffer from non-uniqueness, poor handling of ring systems, and decoding failures that corrupt pharmacophoric

complexity—limitations we further confirm in this work. So why look backward? The answer lies in a fundamental principle: to advance AI-driven drug discovery, we must first rigorously understand *why* and *how* current tools fail. While modern end-to-end models like RFdiffusion or GeoDiff offer astonishing generative power, they operate as black boxes—making it difficult to isolate whether a failure stems from representation, scoring, search strategy, or data bias. In contrast, ChemVAE provides a transparent, modular, and interpretable scaffold in which each component—encoding, scoring, optimization, local neighborhood sampling—can be independently probed, validated, and debugged using medicinal chemistry principles. We chose ChemVAE not because it is the most powerful generative model, but because it is the ideal diagnostic platform. By coupling it with a chemically grounded, feature-based scoring function (KIL), Bayesian optimization, and cluster-aware local neighborhood sampling, we created a controlled experimental system in which the representation gap could be cleanly isolated and quantified. Our results—such as the systematic under-generation of aromatic rings despite their high importance in KIL, or the privileged transformability of LCK into SRC-like chemotypes—reveal structural and functional truths about kinase inhibitor space that would be obscured in a monolithic diffusion pipeline.

Moreover, this work responds to a growing concern in the field: the decline of interpretability in pursuit of performance. As generative models grow more complex, they risk becoming "alchemy engines"—producing novel molecules without explaining *why* they work. Our pipeline reaffirms that chemistry must guide AI, not the reverse. The fact that a "legacy" framework, when augmented with domain knowledge, can yield actionable insights into scaffold hopping, latent space organization, and generative failure modes underscores a critical message: algorithmic novelty alone is insufficient; chemical fidelity is paramount. Indeed, our work affirms that the future of drug discovery lies not in fully autonomous AI, but in AI–chemist collaboration. The KIL framework, though modular, embodies this philosophy: it aligns AI with domain knowledge, prioritizes pharmacophoric fidelity over mere chemical plausibility, and validates candidates through integrated assessment (KIL + Tanimoto + ADMET). As the field adopts increasingly powerful foundation models, this chemistry-first paradigm will remain essential to ensure that generated molecules are not just novel, but biologically relevant. In conclusion, this study provides both a caution and a compass. It cautions that SMILES-based generative models, even when guided by robust scorers, are fundamentally limited in their ability to capture the topological complexity of kinase inhibitors. Our study also offers a compass: by diagnosing these limits, showcasing the power of interpretable design, and revealing how latent space encodes functional relationships, we provide a blueprint for next-generation hybrid systems.

5. Conclusions

This study presents a modular, chemistry-first generative framework for *de novo* SRC kinase inhibitor design, integrating deep generative modeling (ChemVAE), interpretable machine learning (KIL), probabilistic optimization (Bayesian Optimization), and structure-aware latent space remodeling. Across two complementary strategies—global exploration via Bayesian search and local transformation via cluster-guided targeted local neighborhood sampling, we successfully generated novel, drug-like molecules with moderate-to-high predicted SRC inhibition potential. However, our most significant contributions lie not in the molecules themselves, but in the rigorous diagnosis of what works, what does not, and why—providing a foundational benchmark for the next generation of AI-driven drug discovery. Our pipeline demonstrates three practical strengths that remain relevant even in the era of end-to-end generative AI: Interpretability enables debugging and refinement: The KIL scoring function—built on chemically meaningful RDKit features—allowed us to trace failures directly to molecular properties (e.g., underrepresentation of aromatic rings). This transparency is absent in black-box deep scorers and is invaluable for iterative design. Latent space structure encodes functional relationships: We showed that kinase inhibitors form a coherent manifold in ChemVAE space, with LCK and SRC families exhibiting exceptional proximity. This enabled successful scaffold transformation—LCK-derived molecules were 2–4× more likely to

achieve high similarity to known SRC inhibitors than those from other families—demonstrating that latent geometry can guide rational scaffold hopping. Hybrid strategies outperform single-method approaches: While Bayesian Optimization excelled at global drug-likeness tuning, targeted local neighborhood sampling engineering recovered critical pharmacophoric complexity (e.g., multi-ring systems) that BO missed. This synergy underscores the value of multiple generative lenses in lead discovery. Despite its strengths, our framework exposed a fundamental limitation: SMILES-based VAEs cannot reliably represent or manipulate complex ring topologies. Both BO and targeted local neighborhood sampling generated molecules with ≤3 aromatic rings—far fewer than the 3–6 rings typical of clinical SRC inhibitors—even when aromatic ring count was a top KIL feature. This "representation gap" means that no amount of scoring refinement can compensate for a generative engine that cannot access the relevant chemical subspaces. Moreover, biasing optimization with known actives reduced scaffold diversity without improving quality, trapping search in narrow local optima. This challenges the common assumption that seeding with reference compounds universally enhances generative outcomes.

Our work exemplifies a chemistry-first paradigm in AI-driven drug design: rather than treating molecules as abstract tokens, we grounded every component—scoring, generation, validation—in medicinal chemistry principles. This approach prioritizes pharmacophoric fidelity over mere chemical plausibility, uses interpretable features to align AI with domain knowledge, Validates candidates through multi-metric assessment (KIL + Tanimoto + ADMET). This approach yielded three critical lessons. First, we confirmed that molecular syntax encodes functional semantics: the ChemVAE latent space—learned without target labels—spontaneously organizes kinase inhibitors into a coherent, low-dimensional manifold, with SRC acting as a structural "hub" that overlaps all other families. This organization enabled rational scaffold transformation, most strikingly in the LCK → SRC conversion, where 40% of high-similarity output originated from a single family. This demonstrates that latent geometry can serve as a map for guided chemical innovation principles that transcend the underlying generative architecture. Second, we exposed a representation gap that no amount of scoring refinement can overcome. Despite aromatic ring count being a top feature in our interpretable Kinase Inhibition Likelihood (KIL) metric, both Bayesian Optimization and targeted local neighborhood sampling failed to generate molecules with >3 aromatic rings—a hallmark of clinical SRC inhibitors. This failure is not a flaw in optimization, but a fundamental limitation of SMILES-based decoding, which entangles ring topology and corrupts pharmacophoric complexity. By isolating this bottleneck in a transparent pipeline, we provide a diagnostic benchmark that even state-of-the-art models must confront. Third, we reaffirmed that chemistry must guide AI, not the reverse. The KIL framework—built on medicinal chemistry principles rather than abstract embeddings—enabled us to trace generative failures to specific molecular properties, prioritize candidates through multi-metric validation (KIL + Tanimoto + ADMET), and reject the illusion that "drug-likeness" alone ensures biological relevance. Our finding that biasing with known actives reduces diversity without improving quality further challenges heuristic assumptions in generative design and underscores the value of unbiased exploration in lead discovery.

By demonstrating how latent space structure encodes functional relationships between kinase families (e.g., the privileged transformability of LCK into SRC-like chemotypes), we provide both a critical benchmark for current AI and a blueprint for hybrid systems that blend algorithmic sophistication with domain knowledge. By formulating our work as a benchmark of limitations, we provide an empirical framework for the next generation of AI-driven kinase inhibitor design. While the frontier of generative AI is moving toward end-to-end, structure-aware models that reduce reliance on handcrafted scorers, grounding AI in medicinal chemistry principles remains central to successful drug design.

Supplementary Materials: The following supporting information can be downloaded at the website of this paper posted on Preprints.org.

Author Contributions: Conceptualization, G.V.; methodology, G.V.; K.K.; R.K.; software, G.V.; K.K.; R.K. validation, G.V.; K.K.; R.K.; formal analysis, G.V.; K.K.; R.K. ; investigation, G.V.; K.K.; R.K. ; resources, G.V.; K.K.; R.K.; data curation, G.V.; K.K.; R.K. writing—original draft preparation, G.V.; K.K.; R.K. writing—review and editing, G.V.; K.K.; R.K.; visualization, G.V.; K.K.; R.K. supervision, G.V.; project administration, G.V.; funding acquisition, G.V. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: Data is fully contained within the article, Supplementary Materials. The data presented in this study are available in the article and Supplementary Materials and the Github website All scripts, software and models used in the development and experiments are available in the GitHub site https://github.com/kassabry/Perturbation_Experiment. The GitHub site provides detailed documentation of the deposited information and software. The deep learning frameworks were supported by the TensorFlow backend and python tools such as NumPy, scipy, pandas, and scikitlearn (Supplementary Materials Information). Crystal structures were obtained and downloaded from the Protein Data Bank (http://www.rcsb.org). The rendering of protein structures was done with UCSF ChimeraX package (https://www.rbvi.ucsf.edu/chimerax/) and Pymol (https://pymol.org/2/).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. White, D.; Wilson, R. C. Generative models for chemical structures. *J. Chem. Inf. Model.* **2010**, *50*, 1257–1274. doi: 10.1021/ci9004089.
2. Goh, G. B.; Hodas, N. O.; Vishnu, A. Deep learning for computational chemistry. *J Comput Chem* **2017**, *38*, 1291-1307. doi: 10.1002/jcc.24764.
3. Mater, A. C.; Coote, M. L., Deep Learning in Chemistry. *J Chem Inf Model* **2019**, *59*, 2545-2559. doi: 10.1021/acs.jcim.9b00266.
4. Chen, H.; Engkvist, O.; Wang, Y.; Olivecrona, M.; Blaschke, T. The rise of deep learning in drug discovery. *Drug Discov Today* **2018**, *23*, 1241-1250. doi: 10.1016/j.drudis.2018.01.039.
5. Popova, M.; Isayev, O.; Tropsha, A. Deep reinforcement learning for de novo drug design. *Sci Adv* **2018**, *4*, eaap7885. doi: 10.1126/sciadv.aap7885
6. Dimitrov, T.; Kreisbeck, C.; Becker, J. S.; Aspuru-Guzik, A.; Saikin, S. K. Autonomous Molecular Design: Then and Now. *ACS Appl Mater Interfaces* **2019**, *11*, 24825-24836. doi: 10.1021/acsami.9b01226.
7. Korotcov, A.; Tkachenko, V.; Russo, D. P.; Ekins, S. Comparison of Deep Learning With Multiple Machine Learning Methods and Metrics Using Diverse Drug Discovery Data Sets. *Mol Pharm* **2017**, *14*, 4462-4475. doi: 10.1021/acs.molpharmaceut.7b00578.
8. Sanchez-Lengeling, B.; Aspuru-Guzik, A. Inverse molecular design using machine learning: Generative models for matter engineering. *Science* **2018**, *361*, 360–365. doi: 10.1126/science.aat2663.
9. Gomez-Bombarelli, R.; Wei, J. N.; Duvenaud, D.; Hernández- Lobato, J. M.; Sánchez- Lengeling, B.; Sheberla, D.; Aguilera- Iparraguirre, J.; Hirzel, T. D.; Adams, R. P.; Aspuru-Guzik, A. Automatic Chemical Design Using a Data-Driven Continuous Representation of Molecules. *ACS Cent. Sci.* **2018**, *4*, 268–276. doi: 10.1021/acscentsci.7b00572.

10. Yuan, W.; Jiang, D.; Nambiar, D. K.; Liew, L. P.; Hay, M. P.; Bloomstein, J.; Lu, P.; Turner, B.; Le, Q.-T.; Tibshirani, R.; Khatri, P.; Moloney, M. G.; Koong, A. C. Chemical Space Mimicry for Drug Discovery. *J. Chem. Inf. Model.* **2017**, *57*, 875−882. doi: 10.1021/acs.jcim.6b00754.

11. Segler, M. H.; Kogej, T.; Tyrchan, C.; Waller, M. P. Generating focused molecule libraries for drug discovery with recurrent neural networks. *ACS Cent. Sci.* **2018**, *4*, 120−131. doi: 10.1021/acscentsci.7b00512.

12. Elton, D. C.; Boukouvalas, Z.; Fuge, M. D.; Chung, P. W. Deep learning for molecular design - a review of the state of the art. *Mol. Syst. Des. Eng.* **2019**, *4*, 828−849. https://doi.org/10.1039/C9ME00039A.

13. Bian, Y.; Xie, X.-Q. Generative chemistry: drug discovery with deep learning generative models. *J. Mol. Model.* **2021**, *27*, 71. doi: 10.1007/s00894-021-04674-8.

14. Chen, H.; Engkvist, O.; Wang, Y.; Olivecrona, M.; Blaschke, T. The rise of deep learning in drug discovery. *Drug Discov. Today* **2018**, *23*, 1241−1250. doi: 10.1016/j.drudis.2018.01.039.

15. Vamathevan, J.; Clark, D.; Czodrowski, P.; Dunham, I.; Ferran, E.; Lee, G.; Li, B.; Madabhushi, A.; Shah, P.; Spitzer, M.; Zhao, S. Applications of machine learning in drug discovery and development. *Nat. Rev. Drug Discov.* **2019**, *18*, 463−477. doi: 10.1038/s41573-019-0024-5.

16. Sousa, T.; Correia, J.; Pereira, V.; Rocha, M. Generative Deep Learning for Targeted Compound Design. *J. Chem. Inf. Model.* **2021**, *61*, 5343-5361. doi: 10.1021/acs.jcim.0c01496.

17. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative Adversarial Nets. *Adv. Neural Inf. Proc. Syst.* **2014**, 2672−2680.

18. Yu, L.; Zhang, W.; Wang, J.; Yu, Y. SeqGAN: Sequence Generative Adversarial Nets With Policy Gradient. Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI-17), **2017**, pp 2852−2858.

19. Guimaraes G. L.; Sanchez-Lengeling B.; Outeiral C.; Farias P. L. C.; Aspuru-Guzik A. Objective-reinforced generative adversarial networks (ORGAN) for sequence generation models. *arXiv*, **2017**, 1705.10843.

20. Olivecrona, M.; Blaschke, T.; Engkvist, O.; Chen, H. Molecular de-novo design through deep reinforcement learning. *J Cheminform* **2017,** *9*, 48. doi: 10.1186/s13321-017-0235-x.

21. Sanchez-Lengeling, B.; Outeiral, C.; L, G.; Aspuru-Guzik, A. Optimizing distributions over molecular space. An Objective-Reinforced Generative Adversarial Network for Inverse-design Chemistry (ORGANIC). *ChemRxiv.org* **2017**, 5309668. DOI: 10.26434/chemrxiv.5309668.v2

22. Prykhodko, O.; Johansson, S.V.; Kotsias, P.C.; Arús-Pous, J.; Bjerrum, E.J.; Engkvist, O.; Chen, H. A de novo molecular generation method using latent vector based generative adversarial network. J. Cheminform**. 2019**, *11*, 74. doi: 10.1186/s13321-019-0397-9.

23. Kadurin, A.; Nikolenko, S.; Khrabrov, K.; Aliper, A.; Zhavoronkov, A. druGAN: An Advanced Generative Adversarial Autoencoder Model for de Novo Generation of New Molecules with Desired Molecular Properties in Silico. *Mol. Pharm.* **2017,** *14*, 3098-3104. doi: 10.1021/acs.molpharmaceut.7b00346.

24. Putin, E.; Asadulaev, A.; Ivanenkov, Y.; Aladinskiy, V.; Sanchez-Lengeling, B.; Aspuru-Guzik, A.; Zhavoronkov, A. Reinforced Adversarial Neural Computer for de Novo Molecular Design. *J. Chem. Inf. Model.* **2018,** *58*, 1194-1204. doi: 10.1021/acs.jcim.7b00690.

25. Putin, E.; Asadulaev, A.; Vanhaelen, Q.; Ivanenkov, Y.; Aladinskaya, A. V.; Aliper, A.; Zhavoronkov, A. Adversarial Threshold Neural Computer for Molecular de Novo Design. *Mol. Pharm.* **2018,** *15*, 4386-4397. doi: 10.1021/acs.molpharmaceut.7b01137.

26. Gupta, A.; Muller, A. T.; Huisman, B. J. H.; Fuchs, J. A.; Schneider, P.; Schneider, G. Generative Recurrent Networks for De Novo Drug Design. *Mol. Inform.* **2018,** *37* (1-2). doi: 10.1002/minf.201700111

27. Kadurin, A.; Aliper, A.; Kazennov, A.; Mamoshina, P.; Vanhaelen, Q.; Khrabrov, K.; Zhavoronkov, A. The cornucopia of meaningful leads: Applying deep adversarial autoencoders for new molecule development in oncology. *Oncotarget* **2017,** *8*, 10883-10890. doi: 10.18632/oncotarget.14073.

28. Polykovskiy, D.; Zhebrak, A.; Vetrov, D.; Ivanenkov, Y.; Aladinskiy, V.; Mamoshina, P.; Bozdaganyan, M.; Aliper, A.; Zhavoronkov, A.; Kadurin, A. Entangled Conditional Adversarial Autoencoder for de Novo Drug Discovery. *Mol. Pharm.* **2018,** *15* , 4398-4405. doi: 10.1021/acs.molpharmaceut.8b00839.

29. Dr Cao, N.; Kipf, T. MolGAN: An implicit generative model for small molecular graphs. *arXiv.org*, **2018**, 1805.11973. https://doi.org/10.48550/arXiv.1805.11973

30. Zhu, J-Y.; Park,T.; Isola, P.; Efros. A.A. Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. *arXiv.org*, **2018**, 1703.10593v6. https://doi.org/10.48550/arXiv.1703.10593

31. Maziarka, L.; Pocha, A.; Kaczmarczyk, J.; Rataj, K.; Warchol, M. Mol-CycleGAN - a generative mode, for molecular optimization. *J. Cheminf.* **2020**, *12*, 2. doi: 10.1186/s13321-019-0404-1.

32. Racz, A.; Bajusz, D.; Heberger, K. Multi-Level Comparison of Machine Learning Classifiers and Their Performance Metrics. *Molecules* **2019,** *24*, 2811. doi: 10.3390/molecules24152811

33. Olson, R. S.; La Cava, W.; Orzechowski, P.; Urbanowicz, R. J.; Moore, J. H. PMLB: a large benchmark suite for machine learning evaluation and comparison. *BioData Min* **2017,** *10*, 36. doi: 10.1186/s13040-017-0154-4.

34. Polykovskiy, D.; Zhebrak, A.; Sanchez-Lengeling, B.; Golovanov, S.; Tatanov, O.; Belyaev, S.; Kurbanov, R.; Artamonov, A.; Aladinskiy, V.; Veselov, M.; Kadurin, A.; Nikolenko, S.; Aspuru-Guzik, A.; Zhavoronkov, A. Molecular Sets (MOSES): A Benchmarking Platform for Molecular Generation Models. *Front. Pharmacol.* **2020**, 11, 565644. doi: 10.3389/fphar.2020.565644.

35. Preuer, K.; Renz, P.; Unterthiner, T.; Hochreiter, S.; Klambauer, G. Fréchet ChemNet Distance: A Metric for Generative Models for Molecules in Drug Discovery. *J. Chem. Inf. Model.* **2018**, *58*, 1736-1741. doi: 10.1021/acs.jcim.8b00234.

36. Brown, N.; Fiscato, M.; Segler, M.H.S.; Vaucher, A.C. GuacaMol: Benchmarking Models for de Novo Molecular Design. *J. Chem. Inf. Model.* **2019**, *59*, 1096-1108. doi: 10.1021/acs.jcim.8b00839.

37. Bickerton, G.R.; Paolini, G.V.; Besnard, J.; Muresan, S.; Hopkins, A.L. Quantifying the chemical beauty of drugs. *Nat. Chem.* **2012**, *4*, 90-98. doi: 10.1038/nchem.1243.

38. Ertl, P.; Schuffenhauer, A. Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions. *J. Cheminform.* **2009**, *1*, 8. doi: 10.1186/1758-2946-1-8.

39. Buchwald, P.; Bodor, N. Octanol-water partition: searching for predictive models. *Curr. Med. Chem.* **1998**, *5*, 353-380.

40. Blaschke, T.; Arús-Pous, J.; Chen, H.; Margreitter, C.; Tyrchan, C.; Engkvist, O.; Papadopoulos, K.; Patronov, A. REINVENT 2.0: An AI Tool for De Novo Drug Design. *J. Chem. Inf. Model.* **2020**, *60*, 5918–5922. doi: 10.1021/acs.jcim.0c00915.

41. Loeffler, H. H.; He, J.; Tibo, A.; Janet, J. P.; Voronov, A.; Mervin, L. H.; Engkvist, O. Reinvent 4: Modern AI–Driven Generative Molecule Design. *J. Cheminform.* **2024**, *16*, 20. doi: 10.1186/s13321-024-00812-5.

42. Zhavoronkov, A.; Ivanenkov, Y.A.; Aliper, A.; Veselov, M.S.; Aladinskiy, V.A.; Aladinskaya, A.V.; Terentiev, V.A.; Polykovskiy, D.A.; Kuznetsov, M.D.; Asadulaev, A.; Volkov, Y.; Zholus, A.; Shayakhmetov, R.R.; Zhebrak, A.; Minaeva, L.I.; Zagribelnyy, B.A.; Lee, L.H.; Soll, R.; Madge, D.; Xing, L.; Guo, T.; Aspuru-Guzik, A. Deep learning enables rapid identification of potent DDR1 kinase inhibitors. *Nat. Biotechnol.* **2019**, *37*, 1038-1040. doi: 10.1038/s41587-019-0224-x.

43. Dollar, O.; Joshi, N.; Beck, D.A.C.; Pfaendtner, J. Attention-based generative models for *de novo* molecular design. *Chem. Sci.* **2021**, *12*, 8362-8372. doi: 10.1039/d1sc01050f.

44. Winter, R.; Montanari, F.; Noé, F.; Clevert, D.A. Learning continuous and data-driven molecular descriptors by translating equivalent chemical representations. *Chem. Sci.* **2018**, *10*, 1692-1701. doi: 10.1039/c8sc04175j.

45. Winter, R.; Montanari, F.; Steffen, A.; Briem, H.; Noé, F.; Clevert, D.A. Efficient multi-objective molecular optimization in a continuous latent space. *Chem. Sci.* **2019**, *10*, 8016-8024. doi: 10.1039/c9sc01928f.

46. Winter, R.; Retel, J.; Noé, F.; Clevert, D.A.; Steffen, A. grünifai: interactive multiparameter optimization of molecules in a continuous vector space. *Bioinformatics* **2020**, *36*, 4093-4094. doi: 10.1093/bioinformatics/btaa271.

47. Hoffman, S.C.; Chenthamarakshan, V.; Wadhawan, K.; Cen, P-Y., Das, P. Optimizing molecules using efficient queries from property evaluations. *Nat. Mach. Intell.* **2022**, *4***,** 21–31 https://doi.org/10.1038/s42256-021-00422-y.

48. Wang, M.; Sun, H.; Wang, J.; Pang, J.; Chai, X.; Xu, L.; Li, H.; Cao, D.; Hou, T. Comprehensive assessment of deep generative architectures for de novo drug design. *Brief. Bioinform.* **2022**, *23*, bbab544. doi: 10.1093/bib/bbab544.

49. Wang, S.; Guo, Y.; Wang, Y.; Sun, H.; Huang, J. Smiles-Bert: large scale unsupervised pre-training for molecular property prediction. *Proceedings of the 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics,* **2019**, 429–436. https://doi.org/10.1145/3307339.3342186.

50. Irwin, R.; Dimitriadis, S.; He, J.; Bjerrum, E. J. Chemformer: A Pre-Trained Transformer for Computational Chemistry. *Mach. Learn.: Sci. Technol.* **2022**, *3* (1), 015022. https://doi.org/10.1088/2632-2153/ac3ffb.

51. Westerlund, A. M.; Manohar Koki, S.; Kancharla, S.; Tibo, A.; Saigiridharan, L.; Kabeshov, M.; Mercado, R.; Genheden, S. Do Chemformers Dream of Organic Matter? Evaluating a Transformer Model for Multistep Retrosynthesis. *J. Chem. Inf. Model.* **2024**, *64*, 3021–3033. https://doi.org/10.1021/acs.jcim.3c01685.

52. Zhou, Z.; Kearnes, S.; Li, L.; Zare, R. N.; Riley, P. Author Correction: Optimization of Molecules via Deep Reinforcement Learning. *Sci Rep* **2020**, 10478. doi: 10.1038/s41598-020-66840-x.

53. Xu, Y.; Lin, K.; Wang, S.; Wang, L.; Cai, C.; Song, C.; Lai, L.; Pei, J. Deep learning for molecular generation. *Future Med Chem.* **2019,** *11*, 567-597. doi: 10.4155/fmc-2018-0358.

54. Zhang, J.; Chen, H. *De Novo* Molecule Design Using Molecular Generative Models Constrained by Ligand-Protein Interactions. *J. Chem. Inf. Model.* **2022**, doi: 10.1021/acs.jcim.2c00177

55. Li, Y.; Pei, J.; Lai, L. Structure-based *de novo* drug design using 3D deep generative models. *Chem. Sci.* **2021**, *12*, 13664-13675. doi: 10.1039/d1sc04444c.

56. Xie, W.; Wang, F.; Li, Y.; Lai, L.; Pei, J. Advances and Challenges in De Novo Drug Design Using Three-Dimensional Deep Generative Models. *J. Chem. Inf. Model.* **2022,** *62*, 2269-2279. doi: 10.1021/acs.jcim.2c00042.

57. Gilmer, J.; Schoenholz, S. S.; Riley, P. F. ; Vinyals, O.; Dahl, G. E. Neural Message Passing for Quantum Chemistry. *arXiv* **2017**. https://doi.org/10.48550/ARXIV.1704.01212.

58. Kearnes, S.; McCloskey, K.; Berndl, M.; Pande, V.; Riley, P. Molecular Graph Convolutions: Moving beyond Fingerprints. *J Comput Aided Mol. Des.* **2016**, *30*, 595–608. doi: 10.1007/s10822-016-9938-8.

59. Veličković, P.; Cucurull, G.; Casanova, A.; Romero, A.; Liò, P.; Bengio, Y. Graph Attention Networks. *arXiv* **2017**, 1710.10903 https://doi.org/10.48550/ARXIV.1710.10903.

60. Yang, K.; Swanson, K.; Jin, W.; Coley, C.; Eiden, P.; Gao, H.; Guzman-Perez, A.; Hopper, T.; Kelley, B.; Mathea, M.; Palmer, A.; Settels, V.; Jaakkola, T.; Jensen, K.; Barzilay, R. Analyzing Learned Molecular Representations for Property Prediction. *J. Chem. Inf. Model.* **2019**, *59*, 3370–3388. doi: 10.1021/acs.jcim.9b00237.

61. Jin, W.; Barzilay, R.; Jaakkola, T. Junction Tree Variational Autoencoder for Molecular Graph Generation. *arXiv* **2018**. 1802.04364. https://doi.org/10.48550/ARXIV.1802.04364.

62. Shi, C.; Xu, M.; Zhu, Z.; Zhang, W.; Zhang, M.; Tang, J. GraphAF: A Flow-Based Autoregressive Model for Molecular Graph Generation. *arXiv* **2020**. https://doi.org/10.48550/ARXIV.2001.09382.

63. Bengio, E.; Jain, M.; Korablyov, M.; Precup, D.; Bengio, Y. Flow network based generative models for non-iterative diverse candidate generation. In *Advances in Neural Information Processing Systems 34: Proceedings of the Neural Information Processing Systems Conference (NeurIPS 2021)*, **2021** , 7924-7936.

64. Jain, M.; Deleu, T.; Hartford, J.; Liu, C.-H.; Hernandez-Garcia, A.; Bengio, Y. GFlowNets for AI-Driven Scientific Discovery. *Digital Discovery* **2023**, *2*, 557–577. https://doi.org/10.1039/d3dd00002h.

65. Schütt, K. T.; Kindermans, P-J.; Sauceda, H.E.; Chmiela, S.; Tkatchenko, A.; Müller, K. R. SchNet: A continuous-filter convolutional neural network for modeling quantum interactions. *NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems*, **2017**, 992 – 1002.

66. Schütt, K. T.; Arbabzadah, F.; Chmiela, S.; Müller, K. R.; Tkatchenko, A. Quantum-Chemical Insights from Deep Tensor Neural Networks. *Nat Commun* **2017**, *8*, 13890. https://doi.org/10.1038/ncomms13890.

67. Gasteiger, J.; Groß, J.; Günnemann, S. Directional Message Passing for Molecular Graphs. *arXiv* **2020**. https://doi.org/10.48550/ARXIV.2003.03123.

68. Gasteiger, J.; Giri, S.; Margraf, J. T.; Günnemann, S. Fast and Uncertainty-Aware Directional Message Passing for Non-Equilibrium Molecules. *arXiv* **2020**. https://doi.org/10.48550/ARXIV.2011.14115.

69. Stärk, H.; Ganea, O.-E.; Pattanaik, L.; Barzilay, R.; Jaakkola, T. EquiBind: Geometric Deep Learning for Drug Binding Structure Prediction. *arXiv* **2022**. https://doi.org/10.48550/ARXIV.2202.05146.

70. Lu, W.; Wu, Q.; Zhang, J.; Rao, J.; Li, C.; Zheng, S. TANKBind: Trigonometry-Aware Neural NetworKs for Drug-Protein Binding Structure Prediction. *bioRxiv* **2022**, 2022.06.06.495043; doi: https://doi.org/10.1101/2022.06.06.495043.

71. Liu, S.; Wang, H.; Liu, W.; Lasenby, J.; Guo, H.; Tang, J. Pre-Training Molecular Graph Representation with 3D Geometry. *arXiv* **2021**. https://doi.org/10.48550/ARXIV.2110.07728.

72. Stärk, H.; Beaini, D.; Corso, G.; Tossou, P.; Dallago, C.; Günnemann, S.; Liò, P. 3D Infomax Improves GNNs for Molecular Property Prediction. *arXiv* **2021**. https://doi.org/10.48550/ARXIV.2110.04126.

73. Wang, X.; Zhao, H.; Tu, W.; Yao, Q. Automated 3D Pre-Training for Molecular Property Prediction. *arXiv* **2023**. https://doi.org/10.48550/ARXIV.2306.07812.

74. Xu, M.; Yu, L.; Song, Y.; Shi, C.; Ermon, S.; Tang, J. GeoDiff: A Geometric Diffusion Model for Molecular Conformation Generation. *arXiv* **2022**. https://doi.org/10.48550/ARXIV.2203.02923.

75. Cai, H.; Zhang, Z.; Wang, M.; Zhong, B.; Li, Q.; Zhong, Y.; Wu, Y.; Ying, T.; Tang, J. Pretrainable Geometric Graph Neural Network for Antibody Affinity Maturation. *Nat Commun* **2024**, *15*, 7785. doi: 10.1038/s41467-024-51563-8.

76. Jing, B.; Corso, G.; Chang, J.; Barzilay, R.; Jaakkola, T. Torsional Diffusion for Molecular Conformer Generation. *arXiv* **2022**. https://doi.org/10.48550/ARXIV.2206.01729.

77. Corso, G.; Stärk, H.; Jing, B.; Barzilay, R.; Jaakkola, T. DiffDock: Diffusion Steps, Twists, and Turns for Molecular Docking. *arXiv* **2022**. https://doi.org/10.48550/ARXIV.2210.01776.

78. Corso, G.; Deng, A.; Fry, B.; Polizzi, N.; Barzilay, R.; Jaakkola, T. Deep Confident Steps to New Pockets: Strategies for Docking Generalization. *arXiv* **2024**. https://doi.org/10.48550/ARXIV.2402.18396.

79. Zheng, J.; Yi, H.-C.; You, Z.-H. Equivariant 3D-Conditional Diffusion Model for De Novo Drug Design. *IEEE J. Biomed. Health Inform.* **2025**, *29*, 1805–1816. doi: 10.1109/JBHI.2024.3491318.

80. Watson, J. L.; Juergens, D.; Bennett, N. R.; Trippe, B. L.; Yim, J.; Eisenach, H. E.; Ahern, W.; Borst, A. J.; Ragotte, R. J.; Milles, L. F.; Wicky, B. I. M.; Hanikel, N.; Pellock, S. J.; Courbet, A.; Sheffler, W.; Wang, J.; Venkatesh, P.; Sappington, I.; Torres, S. V.; Lauko, A.; De Bortoli, V.; Mathieu, E.; Ovchinnikov, S.; Barzilay, R.; Jaakkola, T. S.; DiMaio, F.; Baek, M.; Baker, D. De Novo Design of Protein Structure and Function with RFdiffusion. *Nature* **2023**, *620,* 1089–1100. doi: 10.1038/s41586-023-06415-8.

81. Dauparas, J.; Anishchenko, I.; Bennett, N.; Bai, H.; Ragotte, R. J.; Milles, L. F.; Wicky, B. I. M.; Courbet, A.; de Haas, R. J.; Bethel, N.; Leung, P. J. Y.; Huddy, T. F.; Pellock, S.; Tischer, D.; Chan, F.; Koepnick, B.; Nguyen, H.; Kang, A.; Sankaran, B.; Bera, A. K.; King, N. P.; Baker, D. Robust Deep Learning–Based Protein Sequence Design Using ProteinMPNN. *Science* **2022**, *378*, 49–56. doi: 10.1126/science.add2187.

82. Hayes, T.; Rao, R.; Akin, H.; Sofroniew, N. J.; Oktay, D.; Lin, Z.; Verkuil, R.; Tran, V. Q.; Deaton, J.; Wiggert, M.; Badkundri, R.; Shafkat, I.; Gong, J.; Derry, A.; Molina, R. S.; Thomas, N.; Khan, Y. A.; Mishra, C.; Kim, C.; Bartie, L. J.; Nemeth, M.; Hsu, P. D.; Sercu, T.; Candido, S.; Rives, A. Simulating 500 Million Years of Evolution with a Language Model. *Science* **2025**, *387*, 850–858. doi: 10.1126/science.ads0018.

83. Ingraham, J. B.; Baranov, M.; Costello, Z.; Barber, K. W.; Wang, W.; Ismail, A.; Frappier, V.; Lord, D. M.; Ng-Thow-Hing, C.; Van Vlack, E. R.; Tie, S.; Xue, V.; Cowles, S. C.; Leung, A.; Rodrigues, J. V.; Morales-Perez, C. L.; Ayoub, A. M.; Green, R.; Puentes, K.; Oplinger, F.; Panwar, N. V.; Obermeyer, F.; Root, A. R.; Beam, A. L.; Poelwijk, F. J.; Grigoryan, G. Illuminating Protein Space with a Programmable Generative Model. *Nature* **2023**, *623*, 1070–1078. doi: 10.1038/s41586-023-06728-8.

84. Fey, M.; Lenssen, J. E. Fast Graph Representation Learning with PyTorch Geometric. *arXiv* **2019.** https://doi.org/10.48550/ARXIV.1903.02428.

85. Wang, M.; Zheng, D.; Ye, Z.; Gan, Q.; Li, M.; Song, X.; Zhou, J.; Ma, C.; Yu, L.; Gai, Y.; Xiao, T.; He, T.; Karypis, G.; Li, J.; Zhang, Z. Deep Graph Library: A Graph-Centric, Highly-Performant Package for Graph Neural Networks. *arXiv* **2019**. https://doi.org/10.48550/ARXIV.1909.01315.

86. Huang, K.; Fu, T.; Gao, W.; Zhao, Y.; Roohani, Y.; Leskovec, J.; Coley, C. W.; Xiao, C.; Sun, J.; Zitnik, M. Therapeutics Data Commons: Machine Learning Datasets and Tasks for Drug Discovery and Development. *arXiv* **2021**. https://doi.org/10.48550/ARXIV.2102.09548.

87. Krishnan, K.; Kassab, R.; Agajanian, S.; Verkhivker, G. Interpretable Machine Learning Models for Molecular Design of Tyrosine Kinase Inhibitors Using Variational Autoencoders and Perturbation-Based Approach of Chemical Space Exploration. *Int. J. Mol. Sci.* **2022**, *23*, 11262. doi: 10.3390/ijms231911262.

88. Davies, M.; Nowotka, M.; Papadatos, G.; Dedman, N.; Gaulton, A.; Atkinson, F.; Bellis, L.; Overington, J. P. ChEMBL web services: streamlining access to drug discovery data and utilities. *Nucleic Acids Res* **2015,** *43* (W1), W612-W620. doi: 10.1093/nar/gkv352.

89. Wishart, D. S.; Feunang, Y. D.; Guo, A. C.; Lo, E. J.; Marcu, A.; Grant, J. R.; Sajed, T.; Johnson, D.; Li, C.; Sayeeda, Z.; Assempour, N.; Iynkkaran, I.; Liu, Y.; Maciejewski, A.; Gale, N.; Wilson, A.; Chin, L.;

Cummings, R.; Le, D.; Pon, A.; Knox, C.; Wilson, M. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res* **2018,** *46* (D1), D1074-D1082. doi: 10.1093/nar/gkx1037.

90. Knox, C.; Wilson, M.; Klinger, C. M.; Franklin, M.; Oler, E.; Wilson, A.; Pon, A.; Cox, J.; Chin, N. E. (Lucy); Strawbridge, S. A.; Garcia-Patino, M.; Kruger, R.; Sivakumaran, A.; Sanford, S.; Doshi, R.; Khetarpal, N.; Fatokun, O.; Doucet, D.; Zubkowski, A.; Rayat, D. Y.; Jackson, H.; Harford, K.; Anjum, A.; Zakir, M.; Wang, F.; Tian, S.; Lee, B.; Liigand, J.; Peters, H.; Wang, R. Q. (Rachel); Nguyen, T.; So, D.; Sharp, M.; da Silva, R.; Gabriel, C.; Scantlebury, J.; Jasinski, M.; Ackerman, D.; Jewison, T.; Sajed, T.; Gautam, V.; Wishart, D. S. DrugBank 6.0: The DrugBank Knowledgebase for 2024. *Nucleic Acids Res.* **2024** , *52(D1)*, D1265-D1275. doi: 10.1093/nar/gkad976.

91. Liu, T.; Hwang, L.; Burley, S. K.; Nitsche, C. I.; Southan, C.; Walters, W. P.; Gilson, M. K. BindingDB in 2024: A FAIR Knowledgebase of Protein-Small Molecule Binding Data. *Nucleic Acids Res.* **2025** , *53(D1)*, D1633-D1644. doi: 10.1093/nar/gkae1075.

92. Ahmed, A.; Smith, R. D.; Clark, J. J.; Dunbar, J. B., Jr.; Carlson, H. A. Recent improvements to Binding MOAD: a resource for protein-ligand binding affinities and structures. *Nucleic Acids Res* **2015,** *43*, D465-D469. doi: 10.1093/nar/gku1088.

93. Hastings, J.; de Matos, P.; Dekker, A.; Ennis, M.; Harsha, B.; Kale, N.; Muthukrishnan, V.; Owen, G.; Turner, S.; Williams, M.; Steinbeck, C. The ChEBI reference database and ontology for biologically relevant chemistry: enhancements for 2013. *Nucleic Acids Res* **2013,** *41*, D456-D463. doi: 10.1093/nar/gks1146.

94. Irwin, J. J.; Sterling, T.; Mysinger, M. M.; Bolstad, E. S.; Coleman, R. G. ZINC: a free tool to discover chemistry for biology. *J. Chem. Inf. Model.* **2012,** *52*, 1757-1768. doi: 10.1021/ci3001277.

95. Sterling, T.; Irwin, J. J. ZINC 15--Ligand Discovery for Everyone. *J. Chem. Inf. Model.* **2015,** *55*, 2324-2337. doi: 10.1021/acs.jcim.5b00559.

96. Irwin, J. J.; Tang, K. G.; Young, J.; Dandarchuluun, C.; Wong, B. R.; Khurelbaatar, M.; Moroz, Y. S.; Mayfield, J.; Sayle, R. A. ZINC20—A Free Ultralarge-Scale Chemical Database for Ligand Discovery. *J. Chem. Inf. Model.* **2020,** *60*, 6065–6073. doi: 10.1021/acs.jcim.0c00675.

97. Ruddigkeit, L.; van Deursen, R.; Blum, L. C.; Reymond, J. L. Enumeration of 166 billion organic small molecules in the chemical universe database GDB-17. *J. Chem. Inf. Model.* **2012,** *52*, 2864-2875. doi: 10.1021/ci300415d.

98. Ruddigkeit, L.; Blum, L. C.; Reymond, J. L. Visualization, and virtual screening of the chemical universe database GDB-17. *J. Chem. Inf. Model.* **2013,** *53*, 56-65. doi: 10.1021/ci300535x.

99. Visini, R.; Awale, M.; Reymond, J. L. Fragment Database FDB-17. *J. Chem. Inf. Model.* **2017,** *57*, 700-709.

100. Xerxa, E.; Bajorath, J. Data Sets of Human and Mouse Protein Kinase Inhibitors With Curated Activity Data Including Covalent Inhibitors. *Future Sci OA*. **2023** , 9, FSO892. doi: 10.2144/fsoa-2023-0088.

101. Koch, S.; Kullmann, E.; Bajorath, J. Opportunities for Protein Kinase Drug Discovery – 2025 Update on the Chemically Underexplored Human Kinome. *Eur J Med Chem Rep* **2025**, *15*, 100305. https://doi.org/10.1016/j.ejmcr.2025.100305.

102. Hu, H.; Laufkötter, O.; Miljković, F.; Bajorath, J. Data set of competitive and allosteric protein kinase inhibitors confirmed by X-ray crystallography. *Data Brief.* **2021**, *35*, 106816. doi: 10.1016/j.dib.2021.106816.

103. Laufkötter, O.; Hu, H.; Miljković, F.; Bajorath, J. Structure- and Similarity-Based Survey of Allosteric Kinase Inhibitors, Activators, and Closely Related Compounds. *J. Med. Chem.* **2022**, *65*, 922-934. doi: 10.1021/acs.jmedchem.0c02076.

104. Hu, H.; Laufkötter, O.; Miljković, F.; Bajorath, J. Systematic comparison of competitive and allosteric kinase inhibitors reveals common structural characteristics. *Eur. J. Med. Chem.* **2021**, *214*, 113206. doi: 10.1016/j.ejmech.2021.113206.

105. Kanev, G. K.; de Graaf, C.; Westerman, B. A.; de Esch, I. J. P.; Kooistra, A. J. KLIFS: An Overhaul after the First 5 Years of Supporting Kinase Research. *Nucleic Acids Res.* **2021** , 49(D1), D562-D569. doi: 10.1093/nar/gkaa895.

106. Xerxa, E.; Laufkötter, O.; Bajorath, J. Systematic Analysis of Covalent and Allosteric Protein Kinase Inhibitors. *Molecules* **2023**, *28*, 5805. doi: 10.3390/molecules28155805.

107. Bento, A.P.; Hersey, A.; Félix, E.; Landrum, G.; Gaulton, A.; Atkinson, F.; Bellis, L.J.; De Veij, M.; Leach, A.R. An open-source chemical structure curation pipeline using RDKit. *J. Cheminform.* **2020,** *12*, 51. doi: 10.1186/s13321-020-00456-1.

108. Kruger, F.; Stiefl, N.; Landrum, G.A. rdScaffoldNetwork: The Scaffold Network Implementation in RDKit. *J. Chem. Inf. Model.* **2020**, *60*, 3331-3335. doi: 10.1021/acs.jcim.0c00296.

109. Abadi, M.; Agarwal, A.; Barham, P.; Brevdo, E.; Chen, Z.; Citro, C.; Corrado, G. S.; Davis, A.; Dean, J.; Devin, M.; Ghemawat, S.; Goodfellow, I.; Harp, A.; Irving, G.; Isard, M.; Jia, Y.; Jozefowicz, R.; Kaiser, L.; Kudlur, M.; Levenberg, J.; Mane, D.; Monga, R.; Moore, S.; Murray, D.; Olah, C.; Schuster, M.; Shlens, J.; Steiner, B.; Sutskever, I.; Talwar, K.; Tucker, P.; Vanhoucke, V.; Vasudevan, V.; Viegas, F.; Vinyals, O.; Warden, P.; Wattenberg, M.; Wicke, M.; Yu, Y.; Zheng, X. TensorFlow: A System for Large-Scale Machine Learning. In *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)* ***2016***, *16, 265-283.* https://doi.org/10.48550/arXiv.1603.04467

110. Boulesteix, A.; Janitza, S.; Kruppa, J.; König, I. Overview of random forest methodology and practical guidance with emphasis on computational biology and bioinformatics. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery,* **2012***, 2*, 493-507. https://doi.org/10.1002/widm.1072.

111. Godden, J.W.; Xue, L.; Bajorath, J. Combinatorial preferences affect molecular similarity/diversity calculations using binary fingerprints and Tanimoto coefficients. *J. Chem. Inf. Comput. Sci.* **2000,** *40*, 163-166. doi: 10.1021/ci990316u.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.