

An overview of Pocket-based drug design

Pablo Varas Pardo,^{†,‡} Eugenia Ulzurrun,^{¶,‡} Nuria Campillo,[¶] and David Quesada^{*,†}

[†]*Aitenea Biotech, SL*

[‡]*Instituto de Ciencias Matemáticas (ICMAT-CSIC)*

[¶]*Centro de Investigaciones Biológicas Margarita Salas (CIB Margarita Salas-CSIC)*

E-mail: david.quesada@aitenea.es

Abstract

hacer abstract

INTRODUCTION

In recent years, advances in artificial intelligence (AI) techniques have greatly impacted the field of drug discovery.¹ The rise of deep learning has enabled computational drug design to increasingly become a reality. One of the main advantages of using these algorithms is the possibility of exploring relevant areas of a chemical space that is just immense.² Researchers have tried to estimate its size, which refers to the number of possible compounds that could be synthesized. However, there is no scientific consensus on the number, with some estimating it to be between 10^{30} to 10^{60} .³ The discrepancy in the estimations is due to different criteria applied in the studies, such as the maximum size of the molecules, the types of atoms that compose them, or the presence of physicochemical restrictions like the Lipinski rules.^{4–6}

Such a vast search space poses significant challenges when optimizing or searching for molecules with specific properties. As a result, the development of methods to navigate this

space has gained considerable interest in recent years. Several techniques are now available to explore the huge chemical space and identify new potential drug candidates. The current methods can be classified into two families: ligand and pocket-based (Figure 1). Ligand-based methods use a set of high-affinity molecules to a target protein as input to create new ligands.^{7,8} This is useful when the research focuses on an interesting part of the chemical space where the corresponding algorithms generate new potential drug candidates around a specific point that represents a molecule.

On the other hand, pocket-based methods, are used to generate appropriate ligands for protein pockets by utilizing detailed structural information of the target binding site.⁹ For this reason, pocket-based algorithms are highly effective when the three-dimensional structure of a protein is available, allowing for precise molecular docking and virtual screening to design molecules that fit within the structural constraints of the binding site. Such approaches can accurately model interactions between the ligand and the protein, simplifying the design of molecules optimally configured to the target site and making them especially useful to create new drugs that inhibit certain proteins involved in the appearance of diseases. One of the main problems with pocket-based algorithms is the lack of experimental validation.¹⁰ Therefore, developing a protocol to select the best candidates for subsequent synthesis in the laboratory is essential.

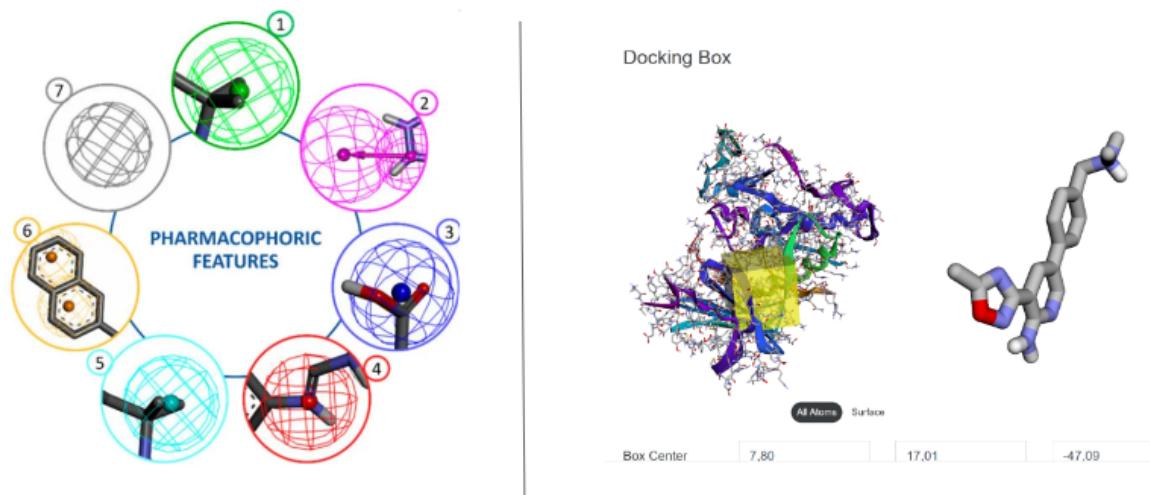


Figure 1. Ligand *vs* Pocket-based drug design. Adapted from.¹¹

A typical pipeline of pocket-based drug discovery is shown in Figure 2. Steps 1 and 2 correspond to getting the PDB file of the interest protein and its binding site respectively. The remaining steps complete the process of generating drug candidates with optimal properties. This article will actually focus on current generative models based on the pocket structure of proteins and the evaluation metrics of the molecules, corresponding to steps 3 and 4 of Figure 2. Its main contributions are¹:

- To create a common framework of current pocket-based molecular generation methods.
- The benchmarking of these methods to assess which are the most promising in pocket-based drug design, creating a reference framework to evaluate additional algorithms.
- To develop a protocol to filter optimal compounds given a target protein.
- To validate experimentally pocket-based algorithms to generate potential drug candidates for DYRK1A protein, an enzyme directly implied in Alzheimer's Disease (AD).

The manuscript is organized as follows. POCKET-BASED GENERATIVE MODELS section presents the current pocket-based generative models models. In MAIN EVALUATION METRICS section the characteristics for choosing a drug candidate are explained. CASE STUDY section presents a specific case study, where the generative models and the evaluation metrics are used to generate optimized molecules for the DYRK1A protein. Finally, in CONCLUSIONS we finished with the analysis extracted from the article and the future lines of pocket-based generative models.

¹For reproducibility reasons, all the code used in this article can be found in <https://github.com/pvaras8/pocketdrugdesign>.

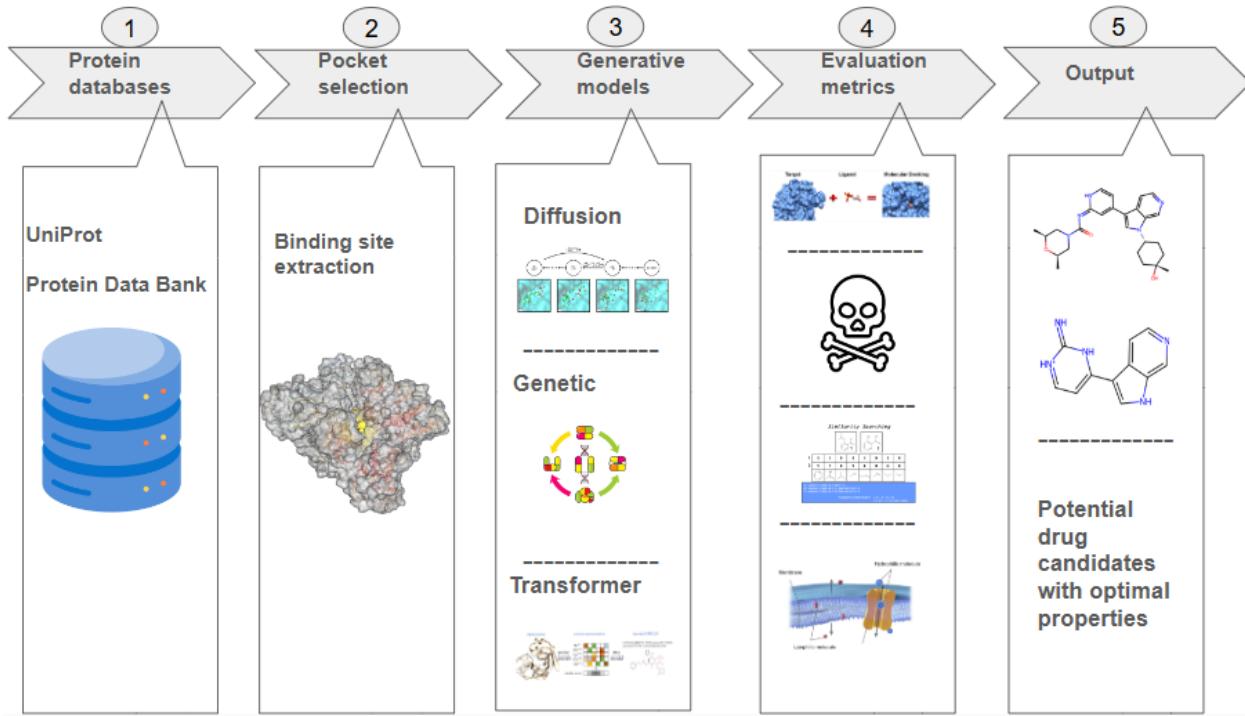


Figure 2. Pipeline of pocket-based drug discovery.

POCKET-BASED GENERATIVE MODELS

Protein pocket-based models are currently being used to generate new ligands and explore vast chemical space. This section examines key technical features of major models, with the Supporting Information displaying key aspects of algorithms implementing them.

Transformer based models

Transformer-based algorithms are gaining popularity in drug discovery because they can capture complex relations within training data through the celebrated *attention mechanism*, and generate high-affinity ligands for a given protein using a sequential-based approach,¹⁰ creating new molecules by joining atoms or molecular substructures.

These models typically require a tokeniser to convert the Simplified Molecular Input Line Entry System (SMILES) code of the molecule and the protein's amino acid chain into tokens.

Its ligand and protein embeddings are then obtained and used to train the weights of the so-called transformer multi-attention layers. Once trained, the model can generate molecules related to a given amino acid sequence.

Following this approach, DrugGPT¹² employs the GPT-2 model and the Byte Pair Encoding (BPE) tokenizer, which can represent a huge number of compounds using a limited vocabulary, to explore the chemical space and discover new ligands for specific proteins. Lingo3DMol⁹ proposes a new representation for SMILES called FSMILES, which fragments SMILES into molecular groups according to certain rules. The protein pocket is then encoded through an encoder, and, starting from the initially calculated growth point, the decoder builds the molecule by iteratively joining molecular groups produced.

TamGent¹³ incorporates a variant of the Transformer encoder designed to process 3D geometric information of targets. Moreover, Ang et al.¹⁴ employ an Encoder-Decoder Transformer combined with Reinforcement Learning through an Adaptive Monte Carlo Tree Search, emphasizing the generation of valid small molecules with desirable drug-like characteristics and binding affinities.

Difussion based models

Diffusion models have attracted special interest in recent years due to their capacity to generate novel compounds. They create a Markov chain of progressive noising steps to add random Gaussian noise to real data until the original sample becomes unrecognizable. Consequently, a model is trained to reverse this process. Once trained, this model can generate new molecules by sampling from a normal distribution and denoising this data until a new compound is created. In the field of pocket-based drug design, this denoising procedure is conditioned on the protein pocket for which the model will create new potential drug candidates.¹⁵

One example is DiffSBDD,¹⁶ which has proved its capability to generate novel ligands with high predicted binding affinities to given protein pockets, highlighting its potential as a

tool for molecules design in structure-based drug design. Moreover, TargetDiff¹⁷ introduces an advanced diffusion model to generate molecules in the 3D space. This ensures that molecular generation is sensitive to the spatial conformation of protein targets.

Genetic algorithms

Genetic Algorithms (GA) implement heuristics inspired by natural evolutionary processes. They use mutation and/or crossover operations to explore the chemical space and maximize a target property, generally the docking value. Following this approach, after several generations, these algorithms can create novel ligands with high affinity given a target protein.¹⁸

One example of GA for pocket-based drug design is AutoGrow4,¹⁹ which generates novel drug-like molecules by applying a series of mutations and crossovers to an initial population of seed molecules. This process is further refined through a fitness function that ranks compounds based on their predicted binding affinities, continuously selecting the top performers for subsequent generations.

Another innovative approach, the Reinforced Genetic Algorithm (RGA),²⁰ proposes a technique that relies on reinforcement learning to choose mutation and crossover operations. The goal is to maximize the docking value, which serves as the fitness function to optimize. The process starts with the selection of approximately 100 drug candidates from an initial database. These compounds are then subjected to mutation and crossover conditioned to the protein pocket to create new ones that improve the docking score.

Graph Neural Networks based models

Graph neural networks have received special attention in recent years within the field of drug discovery. These models are very important for data processing based on structured graphs.²¹ Molecules and proteins are represented as graphs, denoted $G = (V, E)$, whereby:

- V represents the nodes of the graph, each corresponding to an atom in the molecule

or an amino acid protein.

- E represents the edges of the graph, signifying the bonds between atoms or the sequential proximity between amino acids in proteins.

Following this approach, Pocket2Mol²² introduces an equivariant generative network aiming to efficiently sample molecular structures based on the 3D structure of protein pockets. Its innovation lies in its dual-module design: a novel graph neural network capturing spatial and bonding relationships, and an efficient algorithm for conditional molecular sampling.

MAIN EVALUATION METRICS

To create a benchmark for evaluating the molecules generated by the different models, the metrics considered are:

- *Virtual Docking.* It predicts the interaction between a drug molecule and a target protein by simulating their binding affinity.²³ This helps to identify potential drug candidates by assessing how well they fit into the target site.
- *Pharmacological activity.* It measures the biological effects of a drug molecule on the body or specific cells or tissues, therefore assessing the action mechanism of a compound in producing a therapeutic effect, such as inhibiting a disease-related enzyme or activating a receptor. The continuous variable pChEMBL, defined as the $-\log_{10}$ of molar concentration (IC_{50} , XC_{50} , EC_{50} , AC_{50} , Ki , Kd , or Potency), is employed to assess the pharmacological activity of a compound.²⁴
- *Quantitative estimation of drug likeliness (QED).* It measures the likelihood of a chemical compound to be a successful drug candidate based on its physicochemical properties. It quantifies drug-like properties through a single score, considering factors such as solubility, permeability, and molecular weight. These factors are indicative of the ability of a compound to become an effective oral drug in humans.²⁵

- *Lipophylicity (LogP)*. It assesses the tendency of a compound to dissolve in fats, oils, and lipids over aqueous (water-based) solutions, therefore indicating the compound's ability to penetrate cell membranes, affecting its absorption, distribution, metabolism, and excretion properties. Often expressed as LogP, which is the logarithm of the partition coefficient between N-octanol and water.²⁶
- *Molecular Diversity*. It assesses the structural variety among A and B through the formula

$$\text{Tanimoto Similarity} = \frac{c}{a + b - c},$$

where a and b are the counts of features present in molecules A and B respectively, and c is the count of features common to both molecules. Specifically, the features usually employed are represented by the Morgan Fingerprints, converting the molecular structure into a binary vector representing the presence or absence of certain chemical substructures in the molecule. A higher Tanimoto score indicates greater similarity.²⁷

- *Molecular Weight*. Its analysis is crucial for evaluating the molecule's atomic composition and determining the drug pharmacokinetics.²⁶
- *Synthetic Accessibility Score (SAS)*. It is calculated as the sum of the molecular fragment scores plus a complexity penalty based on the presence of certain molecular groups. This metric assesses the ease with which a chemical compound can be synthesized, providing crucial insights about the feasibility of its production on a larger scale.²⁸
- *Toxicity*. In the evaluation of potential drug candidates, assessing the toxicity profile is crucial to determine their safety and viability of molecules to guarantee further development. These toxicity classes include interactions with various nuclear receptors (e.g., androgenic, estrogenic), responses to environmental and endogenous stress signals (e.g., oxidative stress response, heat shock response), and effects on key cellular

processes (e.g., DNA damage response) as shown in Table 1. Each class represents a distinct mechanism through which a compound might exhibit toxicological effects, providing a comprehensive view of its safety profile.²⁹

Table 1. Toxicity classes.

Nuclear Receptor Panel (biomolecular targets)	Stress Response Panel
ER-LBD: estrogen receptor, luciferase	ARE: nuclear factor (erythroid-derived 2)-like 2 antioxidant responsive element
ER: estrogen receptor alpha	HSE: heat shock factor response element
Aromatase	ATAD5: genotoxicity indicated by ATAD5
AhR: aryl hydrocarbon receptor	MMP: mitochondrial membrane potential
AR: androgen receptor	p53: DNA damage p53 pathway
AR-LBD: androgen receptor, luciferase	
PPAR: peroxisome proliferator-activated receptor gamma	

CASE STUDY

In this section we present a case study for evaluating the performance of different generative algorithms.

In recent years, the DYRK1A enzyme has been identified as a promising target for therapeutic intervention in AD, as it is involved in multiple biological functions. Several studies have shown that DYRK1A undergoes alterations linked to the progression of AD, such as the phosphorylation of proteins like TAU³⁰ and APP.³¹ Therefore, DYRK1A is a highly promising enzyme as a therapeutic target for designing new drugs that could be used potentially to treat AD.

The protocol that we have followed to obtain potential lead compounds is shown in Figure

Experimental section

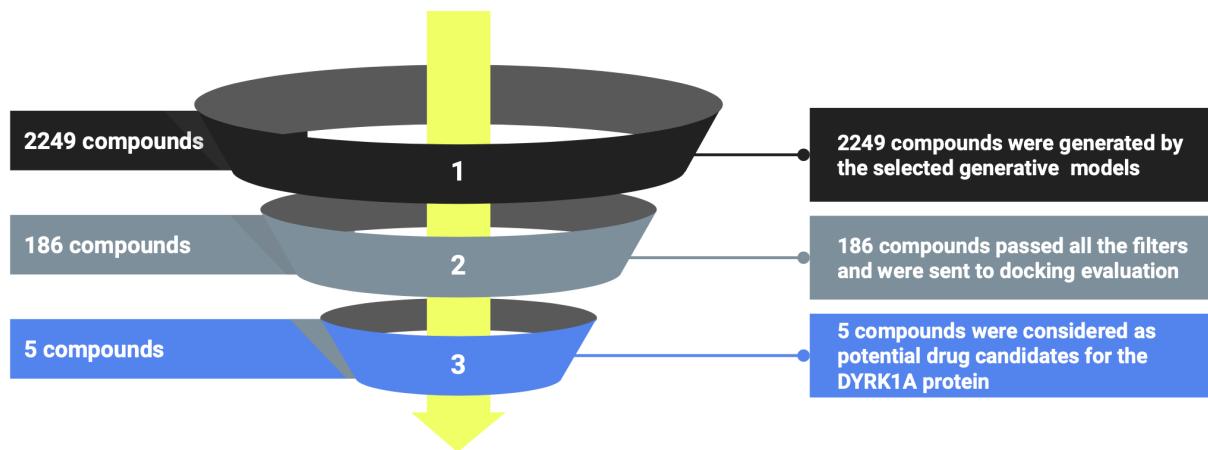


Figure 3. Candidates selection protocol.

Binding site definition

referencia (PS 12 - ref 44)

To begin with, we obtained the DYRK1A 3D structure from the Protein Data Bank (PDB code: 6EIF). We extracted the mass centre of the crystallographic ligand B5T of chain A by taking the average of the x , y and z coordinates, which were 7.80, 17.01, and -47.09 respectively. Molecules were then generated around a box of 15 Å from this point.

Selected generative models

The models used to develop new ligands against the DYRK1A protein were DiffSBDD, Drug-GPT, Lingo3DMol, Pocket2Mol and RGA. These models were chosen for their open-source nature, effectiveness and accessibility. As illustrated in Table 2, these models were trained using different datasets, each with its unique size and composition of training compounds. The DrugGPT model utilizes a substantial dataset from ‘jglaser/binding affinity’³² and ZINC 20³³ totalling 1.9 million and 2 billion compounds, respectively. In contrast, RGA, DiffSBDD and Pocket2Mol are trained on the CrossDocked dataset,³⁴ with DiffSBDD further supplemented by Binding MOAD³⁵ with 41 thousand protein-ligand pairs for training and RGA with 250 thousand compounds of the ZINC 15³⁶ database. In turn, Lingo3DMol is

trained with 20 million commercially available compounds, and in its fine-tuning phase, uses a smaller subset of CrossDocked and DUD-E,³⁷ focusing on precise adjustments with 11.8 thousand and 6.5 thousand compounds, respectively.

Table 2. Summary of models and databases used for training.

Model	Databases	Training Compounds
DiffSBDD	CrossDocked, Binding MOAD	100 K, 41 K
DrugGPT	jglaser/binding affinity, ZINC 20	1.9 M, 2 B
Lingo3DMol (fine-tuning phase)	CrossDocked, DUD-E	11.8 K, 6.5 K
Pocket2Mol	CrossDocked	22.5 M
RGA	CrossDocked, ZINC 15	22.5 M, 250 K

generate new

These models were run to obtain 2249 compounds for the given pocket. For DrugGPT and Lingo3DMol, the number of molecules generated can be chosen as a hyperparameter. For the other three algorithms, the models were run 5 times to obtain a significant number of molecules for comparison, as shown in Table 3.

Filters

The parameters used to filter the molecules are explained in the Supporting Information). For toxicity prediction, the Chemprop algorithm³⁸ was used, which employs a directed message-passing neural network to predict molecular properties. The remaining properties were calculated using the RDKIT package.³⁹

The simulated docking value for all the molecules was obtained using Smina, provided by PyScreener,⁴⁰ allowing efficient and flexible computation of docking scores. Pharmacological activity was calculated using Morgan fingerprints to process the initial SMILES and a Gaussian process to obtain the pChEMBL value. These two metrics were used to compare the models performance in Table 3 In the end, 186 molecules passed all the filters as shown in Figure 3

↳ no están incluidos. Incluirlos aquí o en la figura

Docking studies

The following steps were taken to validate the 186 resulting molecules experimentally:

- *Ligand Preparation.* The conversion from SMILES to SD format was carried out using the structconvert tool available in the Schrödinger suite.⁴¹ Ligand preparation was performed utilizing the LigPrep tool included in the Maestro package.^{42,43} Progressive levels were generated, encompassing possible ionization states at physiological pH and potential tautomers. Final energy minimization was implemented using the OPLS4 force field, with default settings applied for stereoisomers.
- *Protein Preparation.* Human DYRK1A (PDB code 6EIF,⁴⁴) was prepared for subsequent computational analyses using the Protein Preparation Wizard^{45,46} integrated within Maestro.⁴³ The preparation protocol included preprocessing steps such as bond order assignment and structural adjustments carried out using Prime.^{47–49} Protonation and metal charge states for cofactors and metals at pH 7 ± 2 were generated using Epik.^{50,51} The hydrogen-bonding network was optimized, and residue protonation states at pH 7 were calculated using PROPKA.⁵² Water molecules beyond a 5 Å radius from protein residues were excluded, and a final restrained minimization was performed using the OPLS4 force field.
- *Ligand Docking.* The centroid of the crystallized ligand in the catalytic pocket was used as the grid center. During grid generation, a van der Waals radius scaling factor of 1.0 and a partial charge cutoff of 0.25 were applied. Docking was performed using the Glide extra precision (XP) mode available in the Schrödinger software suite,^{53–57} without applying any constraints. Default parameters were employed for ligand setting, including flexible ligand sampling and the incorporation of Epik state penalties into the docking score. The final step involved post-docking minimization using default settings.

- *Docking Validation Protocol.* To validate the docking protocol for DYRK1A using the Glide program, we redocked the ligand B5T (XMD7-117) into the binding site of the crystal structure 6EIF.

RESULTS

This section provides an in-depth analysis of the performance of the molecular generation models. The evaluation includes the metrics described in Section 3, designed to provide a comprehensive view of each model’s capabilities. Taken together, these metrics provide insights about the ability of the models to generate structurally innovative and diverse molecules that may be potential therapeutic agents. The results are critical to understanding the current state of the art in molecule generation algorithms and highlight the strengths and weaknesses of each model.

First, it is important to understand whether these models are exploring different parts of the chemical space, or whether they are focusing on a narrow part. Figure 4 displays a t-distributed Stochastic Neighbour Embedding (t-SNE) plot to illustrate the molecular diversity generated by different computational models to help answer this question. Each point in the plot represents a molecule, and the proximity of the points indicates their structural similarity.

The distinct clustering of points suggests that each model has a unique signature in terms of molecular generation, with some models, as indicated by the concentrated clusters, producing molecules with higher structural homogeneity. In contrast, the more dispersed clusters suggest models that generate a more structurally diverse set of molecules. The presence of distinct and well-defined clusters also implies that certain models may specialize in particular regions of the chemical space, potentially aligning with specialized drug discovery objectives. The original molecules are marked separately, serving as a baseline reference for the diversity introduced by each model. Overall, this figure highlights the importance of diversity in molecular design and the capacity of different models to explore the vast chemical

Se necesita referencia!

space.

power of number coupled to entire percentas RGA

Thus, the RGA model produces molecules grouped in a molecular space quite similar to the original database. This is expected since the genetic algorithm forms molecules by joining common molecular substructures given an initial population of molecules.

Similarly, the DIFSBDD molecules are grouped in a very specific chemical space, similar to those of DrugGPT. On the other hand, both Pocket2Mol and Lingo3DMol molecules achieve greater molecular diversity by exploring different areas of the chemical space that are distant from the molecules in the original database.

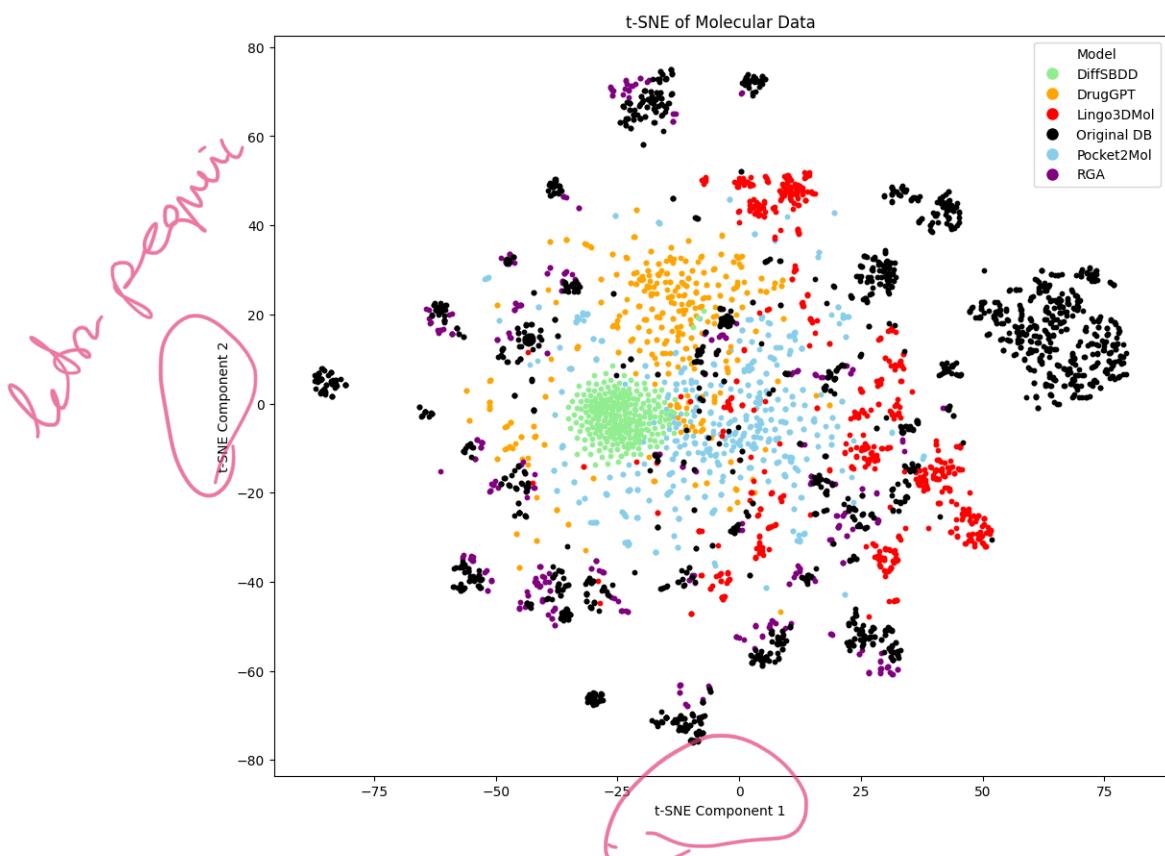


Figure 4. Diversity comparison of proposed generative models.

Table 3 showcases the algorithmic performance through different evaluation metrics. The results underscore the efficacy of Pocket2Mol in docking score averages and top-ranking molecules, suggesting a superior fit for potential drug candidates within the target protein's binding site. Meanwhile, RGA distinguishes itself in predicting binding affinities, hinting

at its utility in identifying potential drug candidates. In addition, all the models except DiffSBDD and DrugGPT create molecules with a median logP between 1.5-3.5, which is essential in our case study, where the compounds have to penetrate the blood-brain barrier (BBB).⁵⁸ By its way, DrugGPT, Lingo3DMol, Pocket2Mol, and RGA design molecules with a QED higher than 0.5, so the ~~de-novo~~ *corsic* design molecules are highly likely to become a drug candidate.⁵⁹ Looking at the SAS, DrugGPT excels among the others, so the molecules created by this algorithm are easier to synthesise.⁶⁰ Finally, all the models create molecules with a reasonable molecular weight.

Table 3. Model performance evaluation

	DiffSBDD	DrugGPT	Lingo3DMol	Pocket2Mol	RGA
Docking (Mean ↓) ^a	-6.77	-8.72	-8.71	-9.21	-8.53
Docking Top1 (↓)	-9.2	-10.80	-11.30	-12.51	-11.2
Docking Top10 (↓)	-8.77	-10.47	-10.64	-12.08	-10.61
pCHEMBL (Mean ↑) ^b	5.87	5.99	6.08	5.96	6.18
pCHEMBL Top1 (↑)	6.83	7.08	8.03	7.13	9.01
pCHEMBL Top10 (↑)	6.59	6.86	7.69	6.89	8.22
logP (Mean 1.5-3.5) ^c	0.43	3.98	2.46	2.68	2.80
QED (Mean ↑) ^b	0.48	0.53	0.65	0.67	0.62
SAS (Mean ↓) ^a	4.76	2.57	2.83	3.25	3.52
Validity (↑) ^b	1	1	0.99	1	0.85
Molecular Weight (180-480) ^c	268.67	418.80	335.16	309.73	314.41
Molecules evaluated	271	400	600	562	416

^a Indicates better performance with lower values.

^b Indicates better performance with higher values.

^c Optimal range specified.

Figure 5 presents a visual comparison of the docking score distributions for the five molecular generation models. The median docking score, indicated by the horizontal line within each boxplot, serves as a robust indicator of the central tendency among the scores achieved by each model. To determine whether there were significant differences in docking scores among the five models, we conducted a one-way ANOVA test,⁶¹ which indicated a significant difference ($p < 0.001$). Given this, we performed Dunn's post-hoc test with Bonferroni correction.⁶² The results showed significant differences in docking scores between most pairs of models, except between DrugGPT and Lingo3DMol. Notably, Pocket2Mol achieves the

most favourable median docking score, suggestive of its adeptness in generating stable for the target protein pocket. The spread and range of scores, as denoted by the boxes and whiskers, also offer insights into the consistency and reliability of each model's predictions. For instance, while RGA shows a comparatively tight distribution signalling consistency in scoring, Lingo3DMol demonstrates a broader range, indicating greater variability in its docking score predictions. Outliers, represented by individual dots, highlight exceptional cases where molecules exhibit either particularly high or low docking scores compared to the typical range for the model output. Collectively, this graphical representation underscores the performance landscape of these models, guiding researchers to prudent selections.

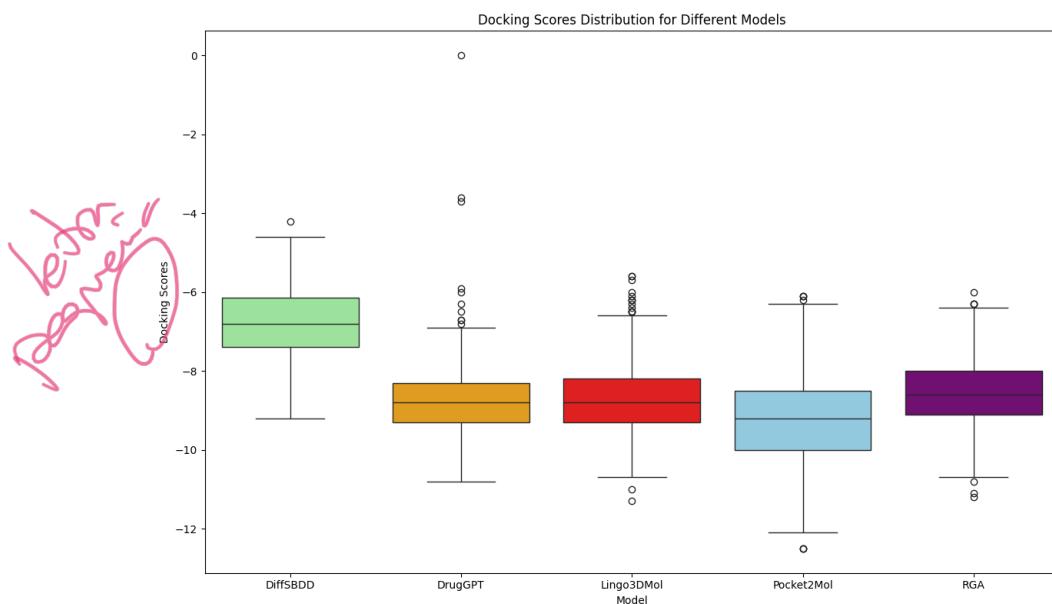


Figure 5. Docking score *Comparison* by Model.

The distribution of ring sizes across the generated molecules through the five models, as presented in Table 4, offers a glimpse about the structural diversity each algorithm brings to the table. The models display a varied propensity towards certain ring sizes, which could be indicative of their bias towards or against specific molecular frameworks. This diversity in ring sizes is essential in the exploration of the vast chemical space. The most common ring sizes are 5 and 6, with DiffSBDD exploring a wider range of ring sizes compared with the other models.

Table 4. Ring size distribution

Ring Size	DiffSBDD	DrugGPT	Lingo3DMol	Pocket2Mol	RGA
3	36.71%	0.27%	0%	0.05%	1.36%
4	5.27%	0.27%	0.28%	0%	0.41%
5	21.94%	25.72%	36.44%	23.96%	31.45%
6	30.38%	72.61%	63.14%	72.21%	66.78%
7	4.22%	0.94%	0.14%	2.37%	0%
8	0.84%	0.07%	0%	0.33%	0%
9	0.42%	0%	0%	0.09%	0%
≥ 10	1.91%	0.20%	0%	2.15%	0%

Figure 6 illustrates the percentage of molecules across non-toxic classes for different models. Models like DiffSBDD and DrugGPT demonstrate a substantial percentage of molecules falling into non-toxic classes. These models are not directly trained to generate non-toxic medicines, but as they create molecules around a narrow part of the chemical space, the results suggest that these models have examined a region of the chemical space where less toxic molecules exist.

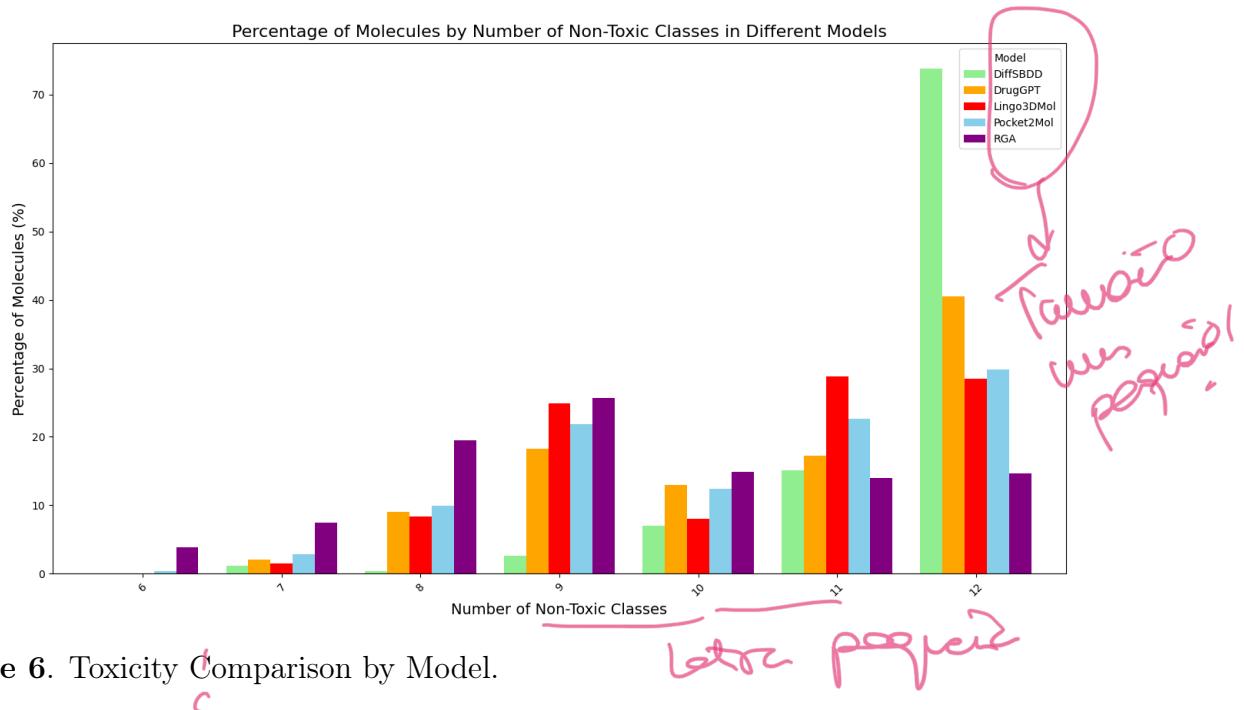


Figure 6. Toxicity Comparison by Model.

After validation of docking protocol with the reference compounds B5T (XMD7-117) (RMSD value of 1.21 Å). A virtual screening was made with the 186 molecules that had

passed the filters mentioned above and presented good autodocking values. Of these 186 molecules, 5 of them presented higher docking scores than the crystallized ligand B5T (used as reference). The inhibitor B5T (XMD7-117) forms hydrogen bonds with Ile165, Glu239, Phe238, and Leu241 (Table 5, Panel A Figure 7).

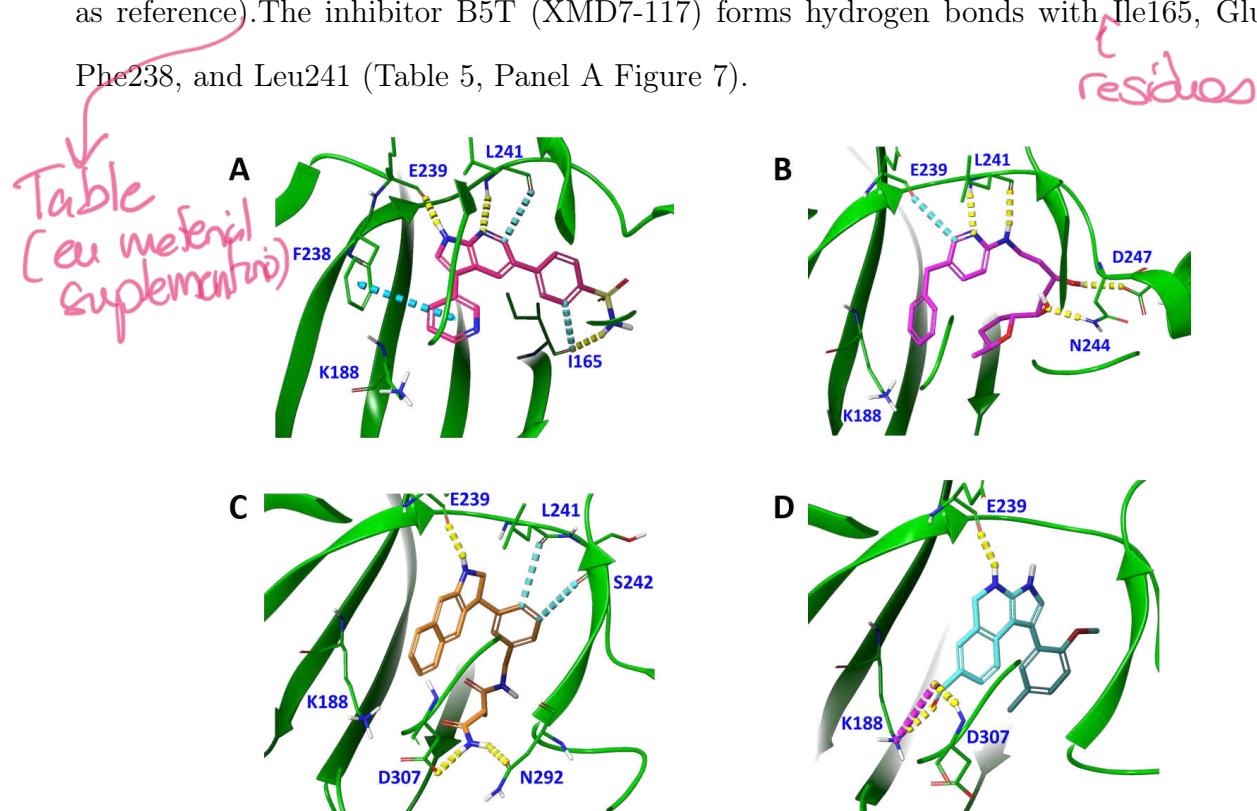


Figure 7. Interactions of key residues in the ATP binding site of DYRK1A structure (PDB 6EIF) with novel inhibitors. Panel A: B5T (reference), Panel B: Compound I, Panel C: Compound II, Panel D: Compound III

Remarkably, interactions with catalytic residue Glu239 and Leu241 are also consistently observed in the top five molecules according to the best docking scores, suggesting their crucial role in ligand binding typical for the other inhibitors.⁴⁴ Notably, compounds IV and V exhibited a pose very similar to that of compound I, with the difference of a Pi-Pi stacking interaction with Phe238 in compound IV and with Ser242 in compound V. Compounds B57, I, and II do not interact with the catalytic lysine Lys188 (Figure 7, Panels A, B and C). However, they possess a terminal arene oriented towards this residue. In contrast, compound III forms a hydrogen bond and a salt bridge with Lys188 (Figure 7, Panel D).

* Compounds I, IV and V exhibit a similar binding pose and their corresponding interactions, with the exception of compound III?

Table 5. Interactions of the top five molecules and control with the protein DYRK1A (PDB: 6EIF)

	Ile165	Lys188	Phe238	Glu239	Leu241	Ser242	Asn244	Asp247	Asn292	Asp307
B5T	AHB ^a HB ^b		PP ^c	HB	AHB HB					
I				AHB	2xHB		HB	HB		
II				HB	AHB	AHB			HB	HB
III		HB SB ^d		HB						HB
IV		PP	AHB	2xHB				2xHB		
V			AHB	2xHB	PP			HB		

^a AHB: Aromatic Hydrogen Bond; ^b HB: Hydrogen Bond; ^c PP: Pi-Pi stacking; ^d SB: Salt Bridge.



The conserved hinge motif found in protein kinases, characterized by two solvent-exposed carbonyl groups and one exposed backbone amide, is widely recognized for its role in forming traditional hydrogen bonds with inhibitors. Similarly, the crystallized inhibitor B5T and the novel inhibitor I exhibit interactions with the hinge backbone, where the hinge-binding motif amide is substituted by an amine (Figure 8). In addition, in both the control and compounds I, IV, and V the hydrogen bonding between residues forming the hinge-like motif involves the same atoms as those in the adenine moiety of ATP (Figure 8)

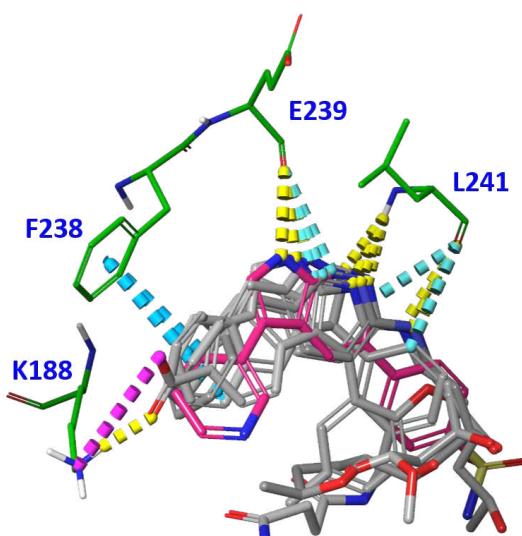


Figure 8. Superposition of the control molecule (magenta) and the top five candidates (grey), showing hydrogen bonding interactions with the backbone scaffold, which closely resembles the hinge backbone crucial for protein function

Faltas un parrafo que resume todo (puesto en overleaf)

- Pablo \Rightarrow se puede convertir algo sencillo en algoritmos que hace generar estas 5 moléculas? Hay más, sobrantes?

CONCLUSIONS

The use of artificial intelligence tools like diffusion models, genetic algorithms, or transformers in designing molecules has gained significant attention in recent years due to their ability to generate new molecules related to a given protein. This article has provided an overview of primary pocket drug design models. It also evaluates the performance of five models in generating molecules against the DYRK1A protein, an enzyme associated with AD. The models evaluated for their ability to generate protein-related molecules included DiffSBDD, DrugGPT, Lingo3DMol, Pocket2Mol and RGA. However, there are many more models under development. These algorithms proved to be highly effective in developing new molecules and exploring new areas of the vast chemical space. They created de novo molecules with high affinity for the DYRK1A protein. The results of virtual docking were later corroborated with traditional docking tools.

A metrics framework has been proposed to assess the performance of protein pocket-based drug design models. This framework is particularly useful as the lack of experimental evaluation is one of the major challenges in the field. It is therefore important to have a filter-based framework to ensure that molecules are screened for suitability before chemists evaluate them in the laboratory. The code used to build this evaluation framework has been made available.

Finally, as running these models can be complicated, a simple code is supplied so that anyone can generate new molecules given the PDB of a protein and the coordinates of its protein pocket.

\rightarrow Yo esto lo pondré al final como conclusión final, convertirlos a los 5 métodos.

- Echo de nuevo, un parágrafo donde se resume los ventajas o desventajas de los mejores métodos

ASSOCIATED CONTENT

Data Availability Statement

The DYRK1A crystal structure is retrievable from the Protein Data Bank

<https://www.rcsb.org/structure/6eif>. The GitHub repository containing the source code and tools used in this article can be accessed at <https://github.com/pvaras8/pocketdrugdesign>.

Supporting Information

AUTHOR INFORMATION

Corresponding Author

David Quesada – Aitenea Biotech;  orcid.org/0000-0002-7280-904X

Author

Pablo Varas Pardo – Aitenea Biotech & Instituto de Ciencias Matemáticas (CSIC), 28049 Madrid, Spain;  orcid.org/0009-0006-1115-4824; Email: pablo.varas@icmat.es

Eugenia Ulzurrun – Centro de Investigaciones Biológicas Margarita Salas (CSIC), 28040 Madrid, Spain;

Nuria E. Campillo – Centro de Investigaciones Biológicas Margarita Salas (CSIC), 28040 Madrid, Spain;  orcid.org/0000-0002-9948-2665

Notes

The authors declare no competing financial interest(s)

Acknowledgement

Hacer

The author thanks

References

- (1) Zhu, H. Big Data and Artificial Intelligence Modeling for Drug Discovery. *Annual review of pharmacology and toxicology* **2020**,
- (2) Vogt, M. Using deep neural networks to explore chemical space. *Expert Opinion on Drug Discovery* **2021**, *17*, 297–304.
- (3) Polishchuk, P. G.; Madzhidov, T. I.; Varnek, A. Estimation of the size of drug-like chemical space based on GDB-17 data. *Journal of computer-aided molecular design* **2013**, *27*, 675–679.
- (4) Henze, H. R.; Blair, C. M. The number of isomeric hydrocarbons of the methane series. *Journal of the American Chemical Society* **1931**, *53*, 3077–3085.
- (5) Weininger, D. Combinatorics of small molecular structures. *Encyclopedia of Computational Chemistry* **2002**, *1*.
- (6) Blair, C. M.; Henze, H. R. The number of stereoisomeric and non-stereoisomeric paraffin hydrocarbons. *Journal of the American Chemical Society* **1932**, *54*, 1538–1545.
- (7) Kaushik, A. C.; Kumar, A.; Bharadwaj, S.; Chaudhary, R.; Sahi, S. Ligand-based approach for in-silico drug designing. *Bioinformatics Techniques for Drug Discovery: Applications for Complex Diseases* **2018**, 11–19.
- (8) Imrie, F.; Hadfield, T. E.; Bradley, A. R.; Deane, C. M. Deep generative design with 3D pharmacophoric constraints. *Chemical science* **2021**, *12*, 14577–14589.

- (9) Feng, W.; Wang, L.; Lin, Z.; Zhu, Y.; Wang, H.; Dong, J.; Bai, R.; Wang, H.; Zhou, J.; Peng, W.; others Generation of 3D molecules in pockets via a language model. *Nature Machine Intelligence* **2024**, 1–12.
- (10) Özçelik, R.; van Tilborg, D.; Jiménez-Luna, J.; Grisoni, F. Structure-Based Drug Discovery with Deep Learning. *ChemBioChem* **2023**, 24, e202200776.
- (11) Giordano, D.; Biancaniello, C.; Argenio, M. A.; Facchiano, A. Drug design by pharmacophore and virtual screening approach. *Pharmaceuticals* **2022**, 15, 646.
- (12) Li, Y.; Gao, C.; Song, X.; Wang, X.; Xu, Y.; Han, S. DrugGPT: A GPT-based strategy for designing potential ligands targeting specific proteins. *bioRxiv* **2023**, 2023–06.
- (13) Wu, K.; Xia, Y.; Fan, Y.; Deng, P.; Liu, H.; Wu, L.; Xie, S.; Wang, T.; Qin, T.; Liu, T.-Y. Tailoring molecules for protein pockets: a transformer-based generative solution for structured-based drug design. *arXiv preprint arXiv:2209.06158* **2022**,
- (14) Ang, D.; Rakovski, C.; Atamian, H. S. De Novo Drug Design Using Transformer-Based Machine Translation and Reinforcement Learning of an Adaptive Monte Carlo Tree Search. *Pharmaceuticals* **2024**, 17, 161.
- (15) Guo, Z.; Liu, J.; Wang, Y.; Chen, M.; Wang, D.; Xu, D.; Cheng, J. Diffusion models in bioinformatics and computational biology. *Nature reviews bioengineering* **2024**, 2, 136–154.
- (16) Schneuring, A.; Du, Y.; Harris, C.; Jamasb, A.; Igashov, I.; Du, W.; Blundell, T.; Lió, P.; Gomes, C.; Welling, M.; others Structure-based drug design with equivariant diffusion models. *arXiv preprint arXiv:2210.13695* **2022**,
- (17) Guan, J.; Qian, W. W.; Peng, X.; Su, Y.; Peng, J.; Ma, J. 3d equivariant diffusion for target-aware molecule generation and affinity prediction. *arXiv preprint arXiv:2303.03543* **2023**,

- (18) Terfloth, L.; Gasteiger, J. Neural networks and genetic algorithms in drug design. *Drug Discovery Today* **2001**, *6*, 102–108.
- (19) Spiegel, J. O.; Durrant, J. D. AutoGrow4: an open-source genetic algorithm for de novo drug design and lead optimization. *Journal of cheminformatics* **2020**, *12*, 1–16.
- (20) Fu, T.; Gao, W.; Coley, C.; Sun, J. Reinforced genetic algorithm for structure-based drug design. *Advances in Neural Information Processing Systems* **2022**, *35*, 12325–12338.
- (21) Sanchez-Lengeling, B.; Reif, E.; Pearce, A.; Wiltschko, A. B. A gentle introduction to graph neural networks. *Distill* **2021**, *6*, e33.
- (22) Peng, X.; Luo, S.; Guan, J.; Xie, Q.; Peng, J.; Ma, J. Pocket2mol: Efficient molecular sampling based on 3d protein pockets. International Conference on Machine Learning. 2022; pp 17644–17655.
- (23) Tang, X.; Dai, H.; Knight, E.; Wu, F.; Li, Y.; Li, T.; Gerstein, M. A Survey of Generative AI for De Novo Drug Design: New Frontiers in Molecule and Protein Generation. *arXiv preprint arXiv:2402.08703* **2024**,
- (24) Chichester, C.; Digles, D.; Siebes, R.; Loizou, A.; Groth, P.; Harland, L. Drug discovery FAQs: workflows for answering multidomain drug discovery questions. *Drug discovery today* **2015**, *20*, 399–405.
- (25) Guan, L.; Yang, H.; Cai, Y.; Sun, L.; Di, P.; Li, W.; Liu, G.; Tang, Y. ADMET-score—a comprehensive scoring function for evaluation of chemical drug-likeness. *Medchemcomm* **2019**, *10*, 148–157.
- (26) Raevsky, O. A. Physicochemical descriptors in property-based drug design. *Mini reviews in medicinal chemistry* **2004**, *4*, 1041–1052.

- (27) Bajusz, D.; Rácz, A.; Héberger, K. Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations? *Journal of cheminformatics* **2015**, *7*, 1–13.
- (28) Bonnet, P. Is chemical synthetic accessibility computationally predictable for drug and lead-like molecules? A comparative assessment between medicinal and computational chemists. *European journal of medicinal chemistry* **2012**, *54*, 679–689.
- (29) Garralaga, M. P.; Lomba, L.; Zuriaga, E.; Santander, S.; Giner, B. Key Properties for the Toxicity Classification of Chemicals: A Comparison of the REACH Regulation and Scientific Studies Trends. *Applied Sciences* **2022**, *12*, 11710.
- (30) de Souza, M. M.; Cenci, A. R.; Teixeira, K. F.; Machado, V.; Mendes Schuler, M. C. G.; Gon, A. E.; Paula Dalmagro, A.; André Cazarin, C.; Gomes Ferreira, L. L.; de Oliveira, A. S.; others DYRK1A Inhibitors and Perspectives for the Treatment of Alzheimer’s Disease. *Current Medicinal Chemistry* **2023**, *30*, 669–688.
- (31) Stotani, S.; Giordanetto, F.; Medda, F. DYRK1A inhibition as potential treatment for Alzheimer’s disease. *Future medicinal chemistry* **2016**, *8*, 681–696.
- (32) Glaser, J. Binding Affinity Dataset. https://huggingface.co/datasets/jglaser/binding_affinity, 2024.
- (33) 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. *Journal of Chemical Information and Modeling* **2020**, *60*, 6065–6073.
- (34) Francoeur, P. G.; Masuda, T.; Sunseri, J.; Jia, A.; Iovanisci, R. B.; Snyder, I.; Koes, D. R. Three-dimensional convolutional neural networks and a cross-docked data set for structure-based drug design. *Journal of Chemical Information and Modeling* **2020**, *60*, 4200–4215.

- (35) Hu, L.; Benson, M. L.; Smith, R. D.; Lerner, M. G.; Carlson, H. A. Binding MOAD (mother of all databases). *Proteins: Structure, Function, and Bioinformatics* **2005**, *60*, 333–340.
- (36) Sterling, T.; Irwin, J. J. ZINC 15–ligand discovery for everyone. *Journal of Chemical Information and Modeling* **2015**, *55*, 2324–2337.
- (37) Stein, R. M.; Yang, Y.; Balius, T. E.; O’Meara, M. J.; Lyu, J.; Young, J.; Tang, K.; Shoichet, B. K.; Irwin, J. J. Property-unmatched decoys in docking benchmarks. *Journal of Chemical Information and Modeling* **2021**, *61*, 699–714.
- (38) Heid, E.; Greenman, K. P.; Chung, Y.; Li, S.-C.; Graff, D. E.; Vermeire, F. H.; Wu, H.; Green, W. H.; McGill, C. J. Chemprop: A machine learning package for chemical property prediction. *Journal of Chemical Information and Modeling* **2023**, *64*, 9–17.
- (39) RDKit: Open-source cheminformatics. <https://www.rdkit.org>, Accessed: 12-Jun-2024.
- (40) Graff, D. E.; Coley, C. W. pyscreener: A python wrapper for computational docking software. *arXiv preprint arXiv:2112.10575* **2021**,
- (41) Schrödinger Software Release 2023-2 distribution. 2023.
- (42) Schrödinger Release 2023-2: LigPrep, Schrödinger, LLC, New York, NY, 2023. 2023.
- (43) Schrödinger Release 2023-2: Maestro, Schrödinger, LLC, New York, NY, 2023. 2023.
- (44) Czarna, A.; Wang, J.; Zelencova, D.; Liu, Y.; Deng, X.; Choi, H. G.; Zhang, T.; Zhou, W.; Chang, J. W.; Kildalsen, H.; others Novel scaffolds for Dual specificity tyrosine-phosphorylation-regulated kinase (DYRK1A) inhibitors. *Journal of Medicinal Chemistry* **2018**, *61*, 7560–7572.

- (45) Madhavi Sastry, G.; Adzhigirey, M.; Day, T.; Annabhimoju, R.; Sherman, W. Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *Journal of Computer-Aided Molecular Design* **2013**, *27*, 221–234.
- (46) Schrödinger Release 2021-1: Protein Preparation Wizard; Epik, Schrödinger, LLC, New York, NY, 2021; Impact, Schrödinger, LLC, New York, NY; Prime, Schrödinger, LLC, New York, NY, 2021. 2023.
- (47) Jacobson, M. P.; Pincus, D. L.; Rapp, C. S.; Day, T. J.; Honig, B.; Shaw, D. E.; Friesner, R. A. A hierarchical approach to all-atom protein loop prediction. *Proteins: Structure, Function, and Bioinformatics* **2004**, *55*, 351–367.
- (48) Jacobson, M. P.; Friesner, R. A.; Xiang, Z.; Honig, B. On the role of the crystal environment in determining protein side-chain conformations. *Journal of Molecular Biology* **2002**, *320*, 597–608.
- (49) Schrödinger Release 2023-2: Prime, Schrödinger, LLC, New York, NY, 2023. 2023.
- (50) Johnston, R. C.; Yao, K.; Kaplan, Z.; Chelliah, M.; Leswing, K.; Seekins, S.; Watts, S.; Calkins, D.; Chief Elk, J.; Jerome, S. V.; others Epik: pKa and Protonation State Prediction through Machine Learning. *Journal of Chemical Theory and Computation* **2023**, *19*, 2380–2388.
- (51) Schrödinger Release 2023-2: Epik, Schrödinger, LLC, New York, NY, 2023. 2023.
- (52) Olsson, M. H.; Søndergaard, C. R.; Rostkowski, M.; Jensen, J. H. PROPKA3: consistent treatment of internal and surface residues in empirical pKa predictions. *Journal of Chemical Theory and Computation* **2011**, *7*, 525–537.
- (53) Yang, Y.; Yao, K.; Repasky, M. P.; Leswing, K.; Abel, R.; Shoichet, B. K.; Jerome, S. V. Efficient exploration of chemical space with docking and deep learning. *Journal of Chemical Theory and Computation* **2021**, *17*, 7106–7119.

- (54) Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrin, P. C.; Mainz, D. T. Extra precision glide: Docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. *Journal of Medicinal Chemistry* **2006**, *49*, 6177–6196.
- (55) Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L. Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *Journal of Medicinal Chemistry* **2004**, *47*, 1750–1759.
- (56) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; others Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *Journal of Medicinal Chemistry* **2004**, *47*, 1739–1749.
- (57) Schrödinger Release 2023-2: Glide, Schrödinger, LLC, New York, NY, 2023. 2023.
- (58) ACD/Labs Making Sense of the LogP Value. https://www.acdlabs.com/wp-content/uploads/download/app/physchem/making_sense.pdf, 2024.
- (59) Li, Y.; Pei, J.; Lai, L. Structure-based de novo drug design using 3D deep generative models. *Chemical Science* **2021**, *12*, 7079–7085.
- (60) Ertl, P.; Schuffenhauer, A. Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions. *Journal of Cheminformatics* **2009**, *1*, 8.
- (61) Girden, E. R. *ANOVA: Repeated measures*; Sage, 1992.
- (62) Bobbitt, Z. Dunn's Test for Multiple Comparisons. <https://www.statology.org/dunns-test/>, 2023; Accessed: 12-Jun-2024.