

AI-driven *de-novo* design and development of non-toxic DYRK1A inhibitors for Alzheimer's disease

Eduardo González,^{†,⊥} Pablo Varas,^{‡,†,⊥} Pedro González-Naranjo,[¶] Eugenia Ulzurrun,^{§,†} Concepción Pérez,[¶] Juan Antonio Páez,[¶] David Ríos Insua,[†] Simón Rodríguez Santana,^{*,||} and Nuria E. Campillo^{*,§}

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

[‡]*AItenea Biotech S.L*

[¶]*Instituto de Química Médica (IQM-CSIC)*

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

^{||}*Universidad Pontificia Comillas (ICAI-UPC)*

[⊥]*These authors contributed equally.*

E-mail: srsantana@icai.comillas.edu; nuria.campillo@csic.es

Abstract

In humans, dual-specificity Tyrosine-Phosphorylation-Regulated Kinase 1A is an enzyme encoded by the DYRK1A gene involved in various diseases, including DYRK1A syndrome, cancer, diabetes, and neurodegenerative pathologies such as Alzheimer's disease (AD). AD is the most prevalent form of dementia, accounting for 60–80% of cases and remains an unmet medical challenge with no cure and just palliative treatments. Recent studies have identified DYRK1A as a promising therapeutic target in AD, given its involvement in multiple biological functions and its alterations correlated with AD progression.

In this work, we leverage multiple Artificial Intelligence (AI) tools, including predictive models and generative algorithms, to design non-toxic DYRK1A inhibitors. We construct a dual-target drug discovery framework integrating AI-driven methods with classical techniques to identify novel compounds. An ensemble Quantitative Structure-Activity Relationship (QSAR) model is employed for predicting compound affinities, while Directed Message Passing Neural Networks (DMPNN) are used to assess toxicity. In a generative phase, a Hierarchical Graph Generation model (HGG) facilitated the design of potential DYRK1A inhibitors. Promising candidate molecules were refined through classical docking studies, leading to their synthesis and experimental validation. As a result, pyrazolyl-1*H*-pyrrolo[2,3-*b*]pyridine was identified as a potent DYRK1A inhibitor, leading to the synthesis of a new derivative series. Enzymatic assays demonstrated nanomolar-level inhibitory activity, while anti-inflammatory and antioxidant properties were confirmed through ORAC assays and LPS-induced pro-inflammatory response evaluations in BV2 microglial cells. Pharmacological testing revealed that the mentioned compound and its derivatives exhibit significant DYRK1A inhibition alongside robust antioxidant and anti-inflammatory effects.

Introduction

Drug discovery, particularly when aiming for addressing complex diseases like Alzheimer's, relies on a deep understanding of biological mechanisms and identifying potential therapeutic targets. Among these, the dual-specificity Tyrosine-Phosphorylation-Regulated Kinase 1A (DYRK1A) family is notable for its role in supporting fundamental biological processes and its association with important diseases, including DYRK1A syndrome, cancer, diabetes,^{1,2} and neurodegenerative disorders such as AD.^{2,3}

AD, commonly characterized through cognitive decline, presents a major health challenge, with many unanswered questions regarding its prevention, treatment, and potential cure. Despite extensive research, its underlying mechanisms remain elusive. However, recent studies connect DYRK1A dysregulation with various aspects of AD pathology,

highlighting its potential as a promising therapeutic target.^{2,4-6}

The design of new molecules against a target has traditionally been addressed using classic design methods, which face significant and complex challenges, *e.g.*, navigating the vast chemical space to identify compounds that meet desired properties. In recent years, artificial intelligence (AI) has emerged as a promising methodology providing tools for *de novo* molecule generation, leveraging extensive databases in conjunction with novel AI generative methods.⁷⁻⁹ By integrating these approaches, researchers aim to accelerate the discovery of novel therapeutics. Numerous successful candidates have emerged from similar strategies, garnering widespread interest in the machine learning, statistics, and chemical design communities.⁹⁻¹³ This interest has been most pronounced in instances with abundant data, while successful cases based on limited datasets remain comparatively rare. Furthermore, most studies focus on proposing promising molecule candidates based on computational models, with far fewer extending to full *in vitro* validation with synthesis and laboratory measurement of molecular properties. This limitation is understandable, as such efforts require significant resources, including access to qualified experts and well-equipped laboratories.

This paper demonstrates a successful integration of AI-based techniques for *de novo* molecule generation. Our generative pipeline is validated through experimental *in vitro* studies of candidate molecules proposed to inhibit the DYRK1A enzyme, a promising therapeutic target for Alzheimer's disease and related conditions. Conducted in a small-data regime, this process employed a range of AI techniques to develop a robust model for suitable candidate generation. The main candidate identified was synthesized alongside its derivatives to evaluate their biological activity. Enzymatic inhibition, anti-inflammatory effects, and antioxidant capacity were experimentally confirmed, showcasing highly promising results.

***De-novo* molecular design**

Drug development is a long and intricate process, widely regarded as one of the most complex industrial tasks. It typically spans 10 to 12 years and incurs costs exceeding a

trillion dollars. Furthermore, around 90% of the proposed new compounds fail to secure approval from the FDA (Food and Drug Administration).¹⁴ An increasing number of animals are also used in preclinical experimentation, despite raising ethical concerns and is still regarded as a necessary step in the drug development process. These factors underscore the immense complexity and challenges of drug discovery and design, which involve navigating a virtual space of over 10^{60} molecules to identify a drug candidate with the desired properties.¹⁵ Over the past decades, various strategies have been implemented to accelerate and reduce the costs of this procedure, focusing on minimizing time, laboratory efforts, and the number of animal lives employed during the experimental phases.^{16,17} AI has emerged as providing promising tools in this domain, aiding in the optimization of the drug development process.^{13,18}

Indeed AI tools are now embedded across multiple steps of drug development, including the identification and validation of molecular targets, optimizing hit and lead compounds, suggesting synthesis pathways for drug-like molecules, predicting critical molecular properties (*e.g.*, ADME-Tox), and even refining the formulation of clinical trials.^{19–21} Notably, one of the most promising applications of AI in this context is the *de novo* molecular design, which involves generating new molecules with optimized property profiles by exploring and exploiting the vast chemical space. Figure 1 illustrates a general workflow of a typical *de novo* drug design process where red boxes represent inputs requiring chemical expertise (*e.g.* data, targets, or expert judgment), green boxes indicate modeling steps, and blue boxes correspond to laboratory work. The outcome of this pipeline is a newly generated set of candidate molecules, proposed based on state-of-the-art models, existing data, and expert insights, and validated through biological testing. Additionally, this process does not require large quantities of data, making it suitable for a wide range of scenarios.

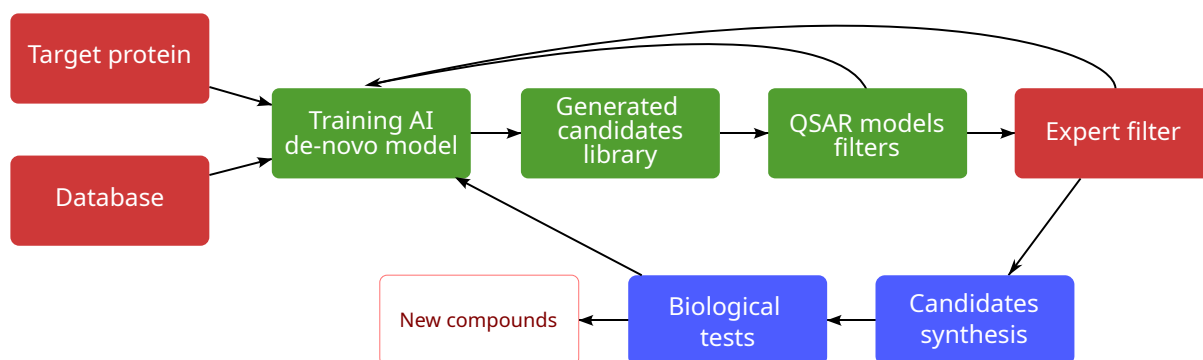


Figure 1: Flowchart for a typical process of *de-novo* design molecules

Recent successes in *de novo* AI molecule generation have led to the development of several AI-based approaches tailored to the diverse needs of different tasks in drug discovery. We categorize these approaches into two main types:

1. **QSAR models**, they estimate compound properties using information collected in chemical databases. This can be achieved in a *regression* setting to predict continuous values, *e.g.* molecular affinity, logP, or Quantitative Estimate of Druglikeness (QED), or in a *classification* setting to assign discrete labels, *e.g.* toxic *vs.* non-toxic. These models rely on training databases, where the quantity and quality of data significantly influence method selection and performance. Deep learning approaches, such as neural networks,²² typically require larger datasets. However, public databases and frameworks like Chemprop²³ mitigate these limitations in many cases, enabling the application of advanced predictive techniques.
2. **Generative models**, Generative AI models serve to create new data from existing datasets. In AI-assisted *de novo* molecule design, these models extrapolate from data or expert knowledge to propose new compounds. Generative approaches can be categorized as *data-heavy* (*e.g.*, deep learning) or *data-sparse* (*e.g.*, evolutionary methods). We focus on multi-target *de novo* generation using methods capable of extracting insights from small datasets to produce promising molecules. While combining predictive and generative models can be effective, the limited size of our DYRK1A inhibition dataset necessitated treating these tasks separately. To address this, we fine-tuned state-of-the-art models with minimal computational resources

to design molecular candidates efficiently.

Related work

The drug development process poses significant challenges for the pharmaceutical industry due to the extensive time and financial resources required to bring a new drug to market. Virtual screening accelerates the discovery of potential drug candidates by enabling the rapid evaluation of large compound libraries. Among the various strategies aimed at reducing the cost and time of drug development, computer-aided drug design (CADD) has proven to be highly effective. Within CADD, structure-based virtual screening (SBVS) is regarded as one of the most promising *in silico* techniques for drug discovery. SBVS focuses on predicting optimal binding interactions between molecules to form stable complexes, using scoring functions to estimate the strength of non-covalent interactions between a ligand and its molecular target. Numerous studies have highlighted the profound impact of CADD on the development of new therapeutics,²⁴ with a marked increase in the application of virtual screening (VS) techniques in drug discovery.²⁵

While the concepts behind computer-assisted *de novo* molecular design have long been established,^{26,27} recent advances have driven unprecedented improvements in the field. In particular, AI-based methods are increasingly being applied to accelerate and optimize drug design processes. Seminal works, such as Gómez-Bombarelli *et al.* on generative models, have paved the way for future developments, significantly influencing advancements in QSAR model formulations.²¹

Given the rapid expansion of research in this area, an exhaustive list of references is beyond the scope of this paper and would be more appropriate for a comprehensive review, such as that conducted by Pang *et al.*²⁹ As highlighted in their work, deep generative models have demonstrated remarkable efficiency in generating drug-like molecules with tailored properties. Li *et al.*³⁰ further advanced these ideas by introducing a 3D deep generative model capable of designing molecules that fit specific target binding sites, a method successfully applied to the design of inhibitors for the main protease of SARS-CoV-2.

Some earlier works²⁶ emphasized the critical importance of considering synthetic feasibility during the molecular design process, an issue that AI-driven models increasingly address by incorporating constraints related to synthetic accessibility. More recent studies^{31–33} have highlighted the potential of AI to accelerate drug discovery by refining molecule generation through machine learning techniques. These approaches enable the prediction of biological activity and the optimization of multiple pharmacological properties. Collectively, these advances represent a shift toward generative models that are highly adaptable to real-world drug development challenges.

From the perspective of experimental validation, a 2019 study employed deep learning algorithms to identify potential inhibitors of discoidin domain receptor 1 kinase (DDR1).³¹ By leveraging extensive datasets of biological activity and molecular structures, the study predicted candidate molecules with DDR1 inhibitory activity. These candidates underwent rigorous validation through molecular docking simulations to assess their binding affinities and structural compatibility within DDR1's active site. This integrated approach demonstrated that the AI-predicted molecules not only exhibited high binding affinities but also conformed structurally to DDR1's requirements for effective inhibition. Consequently, combining AI-driven predictions with docking-based experimental validation significantly accelerated the discovery of novel DDR1-targeting therapeutics.³¹

Building on this methodology, our objective in this work is to harness AI tools, including predictive models and generative algorithms, to design non-toxic inhibitors of DYRK1A. To achieve this, we utilize both our own developed models and state-of-the-art approaches from the literature, leveraging our DYRK1A affinity dataset for this task. The structurally novel molecules generated by the compound of AI models underwent comprehensive experimental validation, including protein-ligand docking simulations. Following synthesis, these molecules were further evaluated through enzymatic and cellular assays, culminating in the design of a family of DYRK1A inhibitors with a good drug-like profile. Such extensive validation is uncommon in similar studies, underscoring the robustness and effectiveness of our approach.

Results and Discussion

The main focus of this work is the design of DYRK1A inhibitors as potential therapeutic agents for Alzheimer's disease (AD) using AI tools. Following an initial design phase, we synthesize and biologically evaluate the performance of each compound to determine its suitability for this task. This evaluation includes both *in vitro* and cellular assays for the most promising molecules. To approach this systematically, we propose the protocol detailed in Figure 2, which outlines the complete process and involves two complementary strategies: AI *de novo* design and classical drug development.

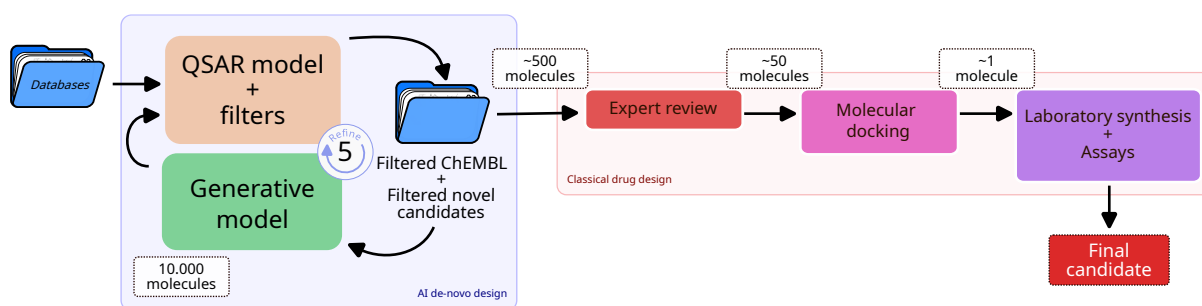


Figure 2: Protocol for DYRK1A inhibitors development.

Following Figure 2, utilizing our DYRK1A affinity dataset, we constructed and fitted QSAR predictive models to estimate the properties of different compounds. By combining this with other public datasets and available models, we also predicted additional chemical properties for proposed compounds with unknown affinity levels, such as their potential toxicity. Similarity and exploitability filters were applied to ensure that the proposed molecules were synthesizable in the laboratory and distinct from those in pre-existing databases. This entire process was coupled with a generative model, which was iteratively refined using the curated selection of candidates to thoroughly explore the most relevant part of the chemical space. Please refer to the *Experimental section* for additional experimental details. Next, thousands (10^4) of candidates were generated using generative models and filtered based on predictions of binding affinity, toxicity, and similarity to known inhibitors. Approximately the top 1% of the most promising candidate molecules were retained through this part of the pipeline. Following a review by expert chemists, the top candidates underwent molecular docking studies to rank their

potential. Finally, the highest-ranked molecules were handpicked, synthesized, and tested in the laboratory.

AI-assisted *de-novo* design for DYRK1A inhibitors

QSAR models for DYRK1A inhibitors

As mentioned earlier, the first step in the *de-novo* molecule design process involves training a model capable of assessing the quality of molecules based on our primary objectives: DYRK1A inhibition, as measured by binding affinity, and non-toxicity. The resulting QSAR model from this process is a key component for the later generative effort, as it enables more efficient navigation of the chemical space, guiding the search towards the most promising candidate regions. This approach estimates these key properties without requiring each proposed compound to be explicitly present in the database. The process was conducted under a low-data regime, as we did not have access to a large database of affinity values. While this limitation poses challenges for certain techniques, we view it as a significant point of interest, as successfully conducting molecule generation in such a context could pave the way for similar efforts in other low-data scenarios.

To predict the affinity of each compound, we selected an ensemble model comprising XGBoost,³⁴ support vector regressors (SVR), k-nearest Neighbors (KNN), and a directed message-passing neural network (DMPNN),²³ following extensive experimental evaluation. While each model performed well individually, integrating these methods into an ensemble predictor provided more robust and reliable forecasts. This ensemble model was compared with other predictive models, including MolCLR,³⁵ a graph neural network (GNN) pre-trained using contrastive learning; SPGNN,³⁶ another GNN pre-trained with a combination of self-supervised and multitask supervised tasks; and traditional models such as random forests (RF), gaussian processes (GP), and a multilayer perceptron (MLP). Each of these models was applied with various molecular representations and optimized individually.

Figure 3 displays the mean ranking of each predictive method, along with their respective standard error bars. Rankings were constructed by sorting the models based

on their predictive performance for each evaluation metric: Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Explained Variance Score (EVS), and Coefficient of Determination (R^2), across the complete dataset, assigning positions from first to last, and calculating the mean positions. The ensemble consistently achieved the lowest mean absolute error, as evidenced by its superior mean ranking and minimal variance around this leading position. Furthermore, we employed the most effective molecular representations for each method in the ensemble, utilizing a rigorous cross-validation procedure to optimize the combination of hyper-parameters and molecular representations. In this implementation, we chose to maintain homogeneous weights for all the ensemble models. However, a further generalization of this approach could employ a weighted average for the distinct model's predictions that constitute the ensemble, potentially achieving even higher performance. For further details and explicit results across the different metrics, please refer to the section on experimental details.

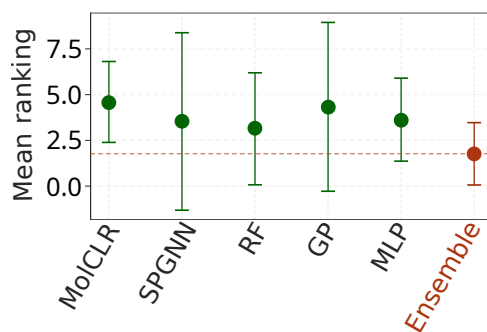


Figure 3: Performance ranking of models. The selected ensemble is highlighted in red. Lower is better.

To evaluate the toxicity of the identified compounds, we employed the Directed MPNN from the Chemprop package.²³ This state-of-the-art model outputs 12 values, each representing the probability of a compound belonging to a specific toxicity class. The model was utilized both as a predictive tool and a filtering mechanism, requiring each candidate molecule to be classified as non-toxic across all 12 metrics to qualify as non-toxic.

Finally, several filters were employed during the generative process to classify potential candidate molecules as promising alongside the predictive affinity and toxicity models. In particular, we applied filters to ensure the internal consistency of the QSAR ensemble model and to assess the similarity of generated molecules to those in pre-existing

datasets. For further details, please refer to the description of the screening filters in the Experimental Section.

Generative models and *de-novo* design

For the generative process, we explored several models to generate potential DYRK1A inhibitors, selecting each method based on data availability and the chemical viability of the proposed compounds. In all cases, candidate molecules were rigorously filtered according to our criteria for binding affinity, toxicity, and novelty.

The chosen model was the HGG model,³⁷ which was trained on a dataset of DYRK1A inhibitors. Through an iterative process, it produced five batches of 10^4 new molecules. After each generation, the newly generated molecules were evaluated for binding affinity, toxicity, and structural similarity to known inhibitors. Molecules passing these filters were added back to the training set, and the model was retrained, enabling iterative refinement of the candidate list over five cycles. We restrict the usage of this recursive approach to a few iterations (≤ 5) to avoid convergence, which may induce a decrease in the diversity of the outputs after excessive iterations. Empirically, this process ensured a final set of molecules with high affinity, low toxicity, and high chemical novelty, resulting in a robust selection of viable inhibitors. The model consistently generated compounds that met these criteria and demonstrated structural characteristics aligned with known effective compounds. Expert chemists reviewing these structures deemed them promising in terms of chemical properties and drug-likeness. Subsequent experimental results confirmed these initial assessments, further validating the effectiveness of the HGG approach. For additional experimental details, as well as information on other models evaluated during this step, please refer to the *Experimental section*.

Finally, the top candidates generated by the Hierarchical Graph Generation model underwent conventional molecular docking calculations to check their experimental performance.

Molecular Docking

A structure-based virtual screening (SBVS) was performed for ~ 50 novel inhibitors targeting the ATP-binding zone of the DYRK1A protein. As control it was redocked the 4E3 (5t) compound into the ATP binding site of the chain A of the published DYRK1A structure (PDB code 4YLL³⁸), reproducing the original pose as the highest ranked posed (RMSD value of 0.715 Å, see experimental section). Table 1 shows the top compounds (rank I-IX) obtained from virtual screening, all of which score higher than the control (4E3). As additional information, Table S1 (see Supporting Information, SI) shows the top 50 molecules resulting from the docking studies.

The inhibitor 4E3 forms bonds with Lys167, Lys188, Glu239, and Leu241 (Table 1). Remarkably, interactions with catalytic residue Glu239 and Leu241 are present in 67% of the compounds I-IX, suggesting their crucial role in ligand binding typical for the other inhibitors.³⁹ On the other hand, only compounds II and III interact with the catalytic lysine Lys188. However, the remaining seven compounds have an arene oriented towards this catalytic residue.

Significantly, the key hinge interactions are conserved across several newly synthesized compounds. The conserved hinge motif in protein kinases, comprising three amino acids, is defined by their positions relative to the sequence downstream of the “gatekeeper” residue, designated as gk+1, gk+2, and gk+3. This motif is well-known for its role in forming traditional hydrogen bonds with inhibitors.⁴⁰ In our case, the reference compound does not exhibit these hydrogen bonds. In contrast, the novel inhibitors I, III, and VI interact with the hinge backbone residues Glu239 (gatekeeper + 1) and Leu241 (gatekeeper + 3) (Table 1). Particularly, the ligand’s chemical moieties interact with the hinge region through three key hydrogen bonds, the ligand donates an H-bond to the main-chain carbonyl of gk+1, while a nitrogen atom accepts an H-bond from the main-chain amide of gk+3. The third interaction in which the ligand donates a proton to the main-chain carbonyl of gk+3 in the third interaction. Compounds I and III exhibit interactions with the hinge backbone similar to those observed in the adenine moiety of ATP within the ATP-binding pocket, involving both canonical and noncanonical hy-

drogen bonds (exemplified in Figure 4). Although compound VI also forms hydrogen bonds with gk+1 (Glu239) and gk+3 (Leu241), these interactions are exclusively canonical (Table 1). In contrast, none of the remaining compounds exhibit these three specific interactions with the hinge region.

The identification of both canonical and noncanonical hydrogen bonds aligns with the established mechanism through which inhibitor scaffolds mimic adenine's interaction with the hinge. These scaffolds incorporate hydrogen bond donors and acceptors that engage with the carbonyl groups of gk+1 and gk+3.^{41,42}

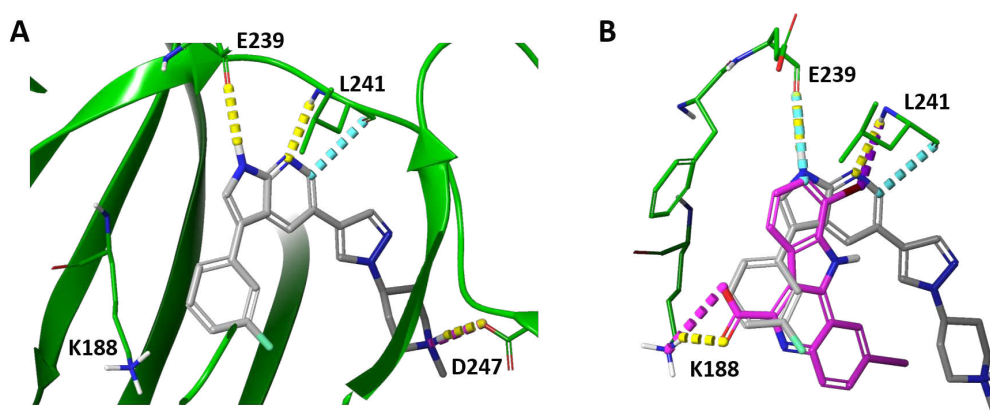
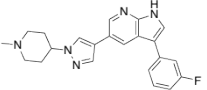
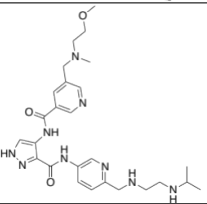
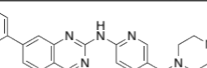
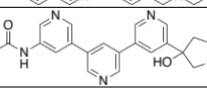
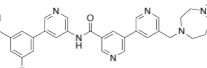
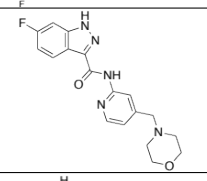
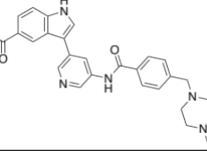
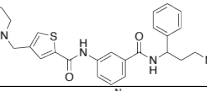
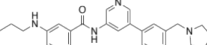


Figure 4: Interactions of residues gk+1 (E239) and gk+3 (L241) in the ATP binding site of DYRK1A structure (PDB 4YLL) with inhibitors. **A:** Hit compound I; **B:** Superposition of the reference compound 4E3 (magenta) and the top hit compound I (grey).

Considering both the docking score and the hinge hydrogen bond interactions, 3-(3-fluorophenyl)-5-(1-(1-methylpiperidin-4-yl)-1*H*-pyrazol-4-yl)-1*H*-pyrrolo[2,3-*b*]pyridine (compound **1**) was selected as the top candidate for synthesis and subsequent biological assays.

Table 1: Interactions of the top-10 ranked compounds with residues in the binding site of DYRK1A. **AHB**: Aromatic Hydrogen Bond; **HB**: Hydrogen Bond; **SBr**: Salt Bridge; **HalB**: Halogen Bond.

Rank	Compound	Docking score (kcal/mol)	Lys167Lys188Glu239		Leu241	Ser242	Asp247	Asn292	Asp307	
			AHB	HB/SB	AHB	HalB	-	-	-	-
Control			AHB	HB/SB	AHB	HalB	-	-	-	-
I (1)		-13.99	-	-	HB	HB/AHB	-	HB/SBr	-	-
II		-13.21	-	HB	HB	HB	AHB	2x(HB/SBr)	-	SBr
III		-12.74	-	HB	HB	2x(HB)	-	HB/SBr	-	-
IV		-12.54	-	-	-	HB/AHB	-	2x(HB)/SBr	-	-
V		-12.38	-	-	-	HB/(AHB)x2	-	HB/SBr	-	-
VI		-12.00	-	-	HB	2x(HB)	-	-	-	-
VII		-11.75	AHB	-	HB	AHB	-	-	AHB	-
VIII		-11.69	-	-	AHB	HB/AHB	-	HB/SBr	-	-
IX		-11.46	HB	-	AHB	-	HB	-	AHB	2x(AHB)

Synthesis of the new compounds

The 3-(3-fluorophenyl)-5-(1-(1-methylpiperidin-4-yl)-1*H*pyrazol-4-yl)-1*H*-pyrrolo[2,3-*b*]pyridine (**1**) was selected as the candidate for synthesis (Figure 5). A literature search for 5-(4-piperidinyl-1*H*--pyrazolyl)-1*H*--pyrrolo[2,3-*b*]pyridine derivatives yielded only three papers^{43–45} and several patents^{46–51} related to this family of candidates. Based on these references, a general synthetic procedure was proposed (Scheme 1).

The synthetic methodology for obtaining the target compound, 3-fluorophenyl-5--(pyrazol-4-yl)-1*H*-pyrrolo[2,3-*b*]pyridine (**1**), involves two stages. The first stage (A) com-

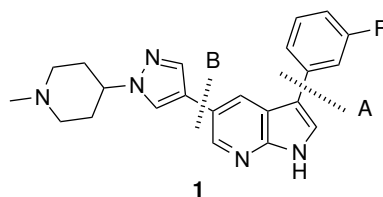


Figure 5: Retrosynthesis to access compound **1**

prises the formation of 3-fluorophenyl-1*H*-pyrrolo[2,3-*b*]pyridine (Scheme 1), followed by the second stage (B), which introduces the 1-(1-methylpiperidin-4-yl)-1*H*-pyrazol-4-yl group (also included in Scheme 1).

The first stage (A) represents a structurally versatile and general synthetic route for the formation of 3-fluorophenyl-1*H*-pyrrolo[2,3-*b*]pyridine (**2**). Initial attempts to synthesize compound **2** directly from compound **3** without protection were unsuccessful. Consequently, the preparation of **2** was accomplished via a two-step route: first, protecting the N-1 position of the pyrrole ring in compound **3** to yield intermediate **15**, followed by the introduction of the aryl substituent at position 3 using 2-(3-fluorophenyl)-4,4,5,5-tetramethyl-1,3,2-dioxaborolane **4** (Scheme 1).

The second stage (B) involves the preparation of 3-fluorophenyl-1*H*-pyrazol-4-yl-1*H*-pyrrolo[2,3-*b*]pyridine **1** through a four-step sequence. This sequence includes the initial introduction of the piperidinepyrazolyl group at position 5 of the pyrrolo[2,3-*b*]pyridine ring, followed by methylation of the nitrogen in the piperidinyl group, as outlined in Scheme 1.

After successfully synthesizing compound **1**, a virtual library of 247 potential derivatives was proposed (see Table S2 in the SI) to predict their affinity for DYRK1A using the AI-based QSAR predictive model previously discussed. This virtual library was designed based on two key criteria: (i) structural variability of substituents on the phenyl ring and (ii) synthetic accessibility during step B (Scheme 1) using commercially available 4,4,5,5-tetramethyl-1,3,2-dioxaborolanes with 2-substituents. Consequently, the aryl-substituted groups at position 3 of the pyrrolo[2,3-*b*]pyridine ring consisted of phenyl groups bearing one to three substituents selected from fluoro, chloro, methyl, trifluoromethyl, methoxy, nitro, and dimethylamino.

The activity of this chemical library was predicted using the AI-based QSAR models described above. Most of the proposed derivatives exhibited affinity at the micro- or sub- μM level (see Table S2 in the SI). Three compounds were chosen to represent a diverse set of 3-aryl-5-pyrazolyl-1*H*-pyrrolo[2,3-*b*]pyridine derivatives (**5-7**). Additionally, mono-, di-, and tri-substituted phenyl derivatives with favorable scores were selected. Thus, along with the initial compound **1**, a representative set of 3-aryl-5-pyrazolyl-1*H*-pyrrolo[2,3-*b*]pyridine derivatives (**5-7**) and the corresponding NH-piperidino derivatives (**8-11**) were proposed as potential candidates (Figure 6).

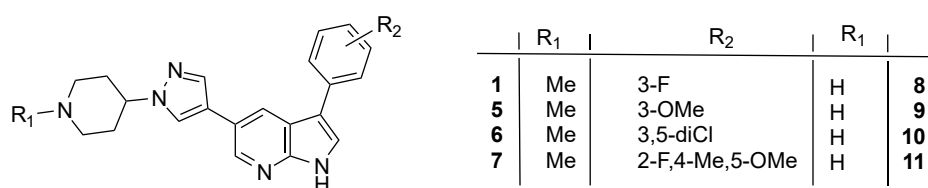
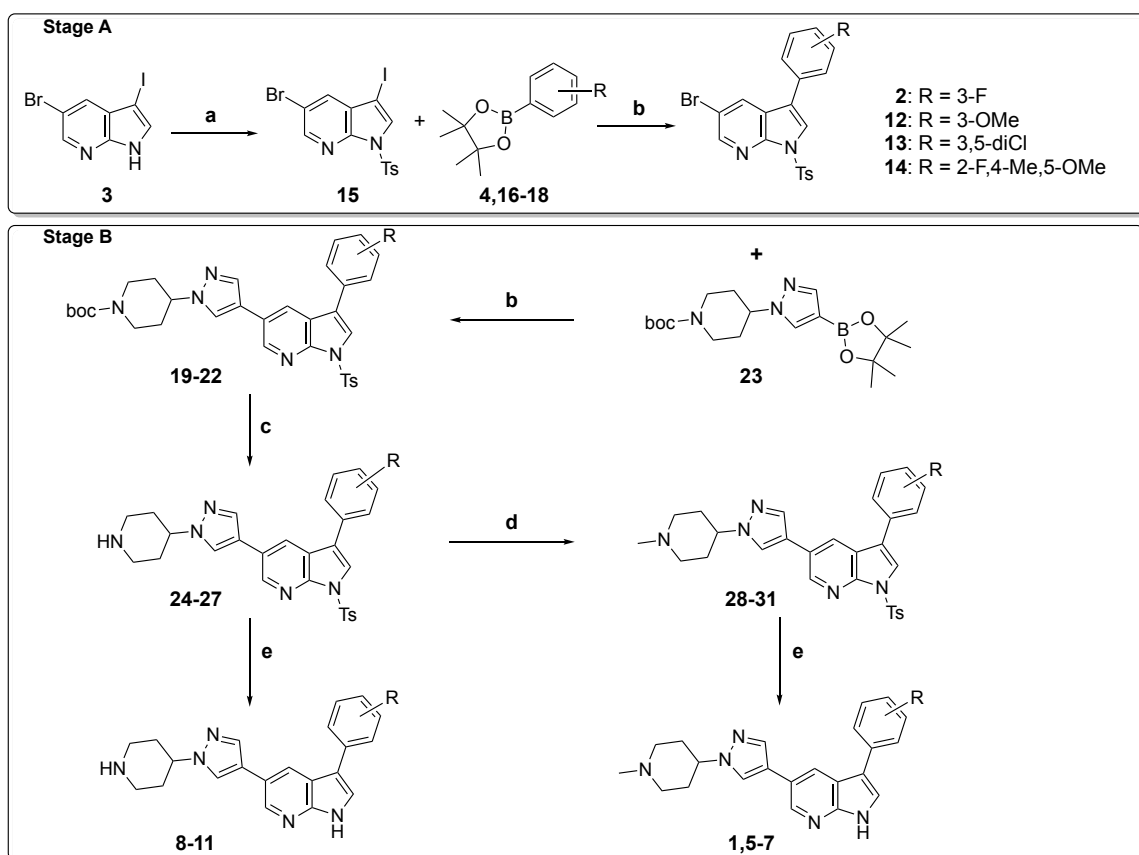


Figure 6: Representative set of 3-aryl-5-pyrazolyl-1*H*-pyrrolo[2,3-*b*]pyridine derivatives proposed

According to stage A (Scheme 1), the first step involves the protection of position N-1 of 5-bromo-3-iodo-1*H*-pyrrolo[2,3-*b*]pyridine (**3**) with tosyl chloride. Subsequently, the synthesis of 3-aryl-1*H*-pyrrolo[2,3-*b*]pyridine derivatives (**2** and **12-14**) was carried out by reacting 1-tosyl-5-bromo-3-iodo-1*H*-pyrrolo[2,3-*b*]pyridine (**15**) with the corresponding 4,4,5,5-tetramethyl-1,3,2-dioxaborolane derivatives (**4** and **16-18**), including 3-fluoro, 3-methoxy, 3,5-dichloro, or 2-fluoro-4-methyl-5-methoxy substituents, using Pd(dppf)Cl₂ as a catalyst.

The general synthetic route for the introduction of the 1-(1-methylpiperidin-4-yl)-1*H*-pyrazol-4-yl group comprises four steps (stage B, steps b-e)(Scheme 1). The synthesis of 3-aryl-5-pyrazolyl-1*H*-pyrrolo[2,3-*b*]pyridine derivatives (**19-22**) was carried out by reacting 5-bromo-1-tosyl-1*H*-pyrrolo[2,3-*b*]pyridine derivatives (**2,12-14**) with tert-butyl 4-(4-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)-1*H*-pyrazol-1-yl)piperidine-1-carboxylate (**23**), using Pd(dppf)Cl₂ as a catalyst. The removal of the Boc group then afforded the corresponding NH-piperidino derivatives (**24-27**). The methylpiperazinylpyrrolo[2,3-*b*]pyridines (**28-31**) were prepared from **24-27** by reacting with formaldehyde in formic acid. Finally, deprotection of the N-1-tosyl group yielded the selected candidates (**1,5-**

7).



Reagents and conditions: (a) Tosyl chloride, NaH, DMF, rt; (b) Pd(dppf)Cl₂, K₂CO₃, 1,4-dioxane:water (6:1), argon, microwave, 100 °C, 2h; (c) TFA, CH₂Cl₂, rt, 90 min; (d) formaldehyde, formic acid, 70 °C, overnight; (e) 0.4 M NaOH in MeOH, rt, 4 h.

Scheme 1: General synthetic route for the preparation of 3-arylpyrrolo[2,3-b]pyridine derivatives (A) and selected 1-methylpiperidin-4-yl-1H-pyrazol-4-yl-pyrrolo[2,3-b]pyridine derivatives (B)

Last, the preparation of the corresponding NH-piperidinyl derivatives (**8-11**) was achieved by deprotection of the N-1-tosyl group from the piperidin-4-yl-1H-pyrrolo[2,3-b]pyridines (**24-27**).

The structures of all newly synthesized compounds were confirmed based on their analytical and spectroscopic data. Detailed spectroscopic characterization is provided in the SI. Specifically, ¹H NMR and ¹³C NMR chemical shifts are summarized in Tables S3 and S4, with Figures S1-S16 presenting the corresponding spectra in the SI.

Biological studies

Once derivatives **1**, **5-11** were synthesized, their biological evaluation was conducted. The results obtained, see Table 2, indicate that all compounds exhibit inhibitory activity in the nanomolar range against DYRK1A. Notably, compounds **1** and **5**, along with their demethylated analogues **8** and **9**, demonstrated IC₅₀ values comparable to the reference compound, harmine. Overall, there was no significant difference in IC₅₀ values between the methylated compounds and their demethylated analogues.

Regarding the antioxidant capacity of compounds **1**, **5-11**, the ORAC assay revealed values around 1 trolox equivalent for most compounds, except for the dichlorinated derivatives **6** and **10**, which showed values around 0.5 trolox equivalents. As with the inhibitory activity, no notable differences were observed between the methylated compounds and their demethylated analogues (Table 2).

Table 2: Inhibition of DYRK1A (IC₅₀, nM) and Oxygen Radical Absorbance Capacity (ORAC, Trolox equivalents) of compounds **1**, **5 - 11**. ^a Compounds were evaluated using ATP (10 μmol/well) and DYRKtide (4 μmol/well) as substrate. Experiments were performed in triplicate. ^b Data are expressed as μmol of Trolox equivalents/μmol of tested compound. ^c See ref.⁵²

Compound	IC ₅₀ DYRK1A(nM) ^a	ORAC ^b
1	41 ± 3	1.02
5	79 ± 5	1.25
6	459 ± 24	0.44
7	231 ± 20	1.3
8	48 ± 3	1.0
9	81 ± 4	1.3
10	450 ± 18	0.55
11	165 ± 17	1.2
Harmine	80 ^c	-

In addition, an interesting property closely related to AD is the one that refers to the anti-inflammatory capacity of a drug. Thus, the compounds were studied using LPS-induced proinflammatory responses in BV2 microglial cells. First, the toxicity of the compounds was assessed via an MTT assay. Compounds **6**, **7**, **10**, and **11** were found to be non-toxic at concentrations up to 10 μM, whereas compounds **1**, **5**, **8**, and **9** exhibited

toxicity at 10 μM but were non-toxic at 5 μM .

Subsequently, the effect of the compounds on LPS-induced proinflammatory responses in BV2 microglial cells was investigated. As illustrated in Figure 7, all compounds reduced LPS-induced NO production, with compounds **1** and **8** showing the most pronounced effects. The inhibition was dose-dependent and was particularly strong for compounds **1** and **5**, along with their demethylated analogues **8** and **9** (Figures S17-S18 in the SI). These findings align with the high DYRK1A inhibition observed previously for these compounds.

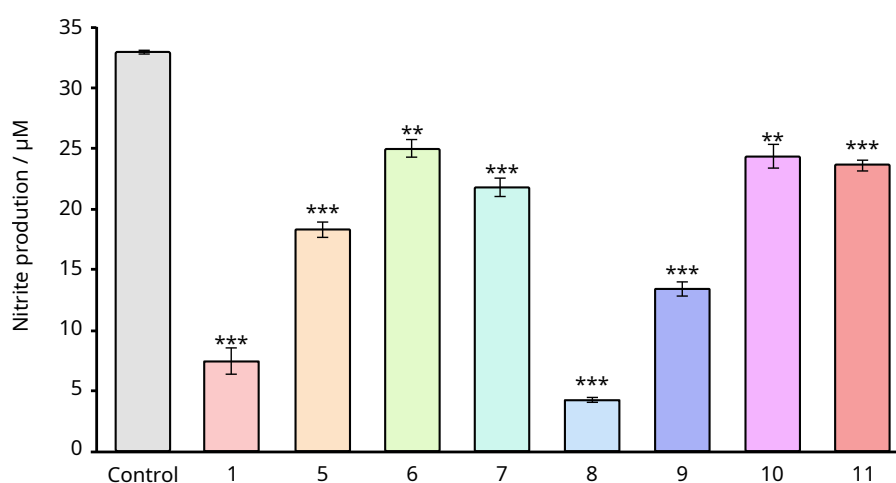


Figure 7: Antiinflammatory effect of compounds **1** and **5-11** at their maximum dose assayed (5 μM (**1**, **5**, **8** and **9**) or 10 μM (**6,7,10** and **11**)). BV2 cells were incubated for 24h with lipopolysaccharide (LPS; 200 ng mL⁻¹) in the absence or presence of inhibitors, and the production of nitrite was evaluated by the Griess reaction. Cells were pretreated with inhibitors for 1h before lipopolysaccharide (LPS) stimulation. Values represent the mean and their respective standard deviations from three independent experiments. **: $p < 0.01$; ***: $p < 0.001$ versus LPS-treated cells.

Conclusions

The objective of this research was to apply a comprehensive range of AI-based techniques to develop an effective model to generate candidate compounds with good drug-like properties as DYRK1A inhibitors. Conducted under a small-data regime, this study utilized a robust pipeline encompassing *de novo* molecular generation, AI-QSAR modeling, expert knowledge integration, and docking studies.

The strategic application of AI tools, including predictive and generative models,

proved highly effective in designing non-toxic DYRK1A inhibitors within a dual-target drug discovery framework. An ensemble model comprising XGBoost, S support vector regressors, K-nearest Neighbors, and a directed message passing neural network (DMPNN) was developed to predict the binding affinity of each compound, while the DMPNN further assessed toxicity profiles. For the generative phase, a hierarchical graph generation model enabled the design of promising DYRK1A inhibitors, facilitating the identification of molecular structures with favorable binding affinity, toxicity, and drug-like properties.

Classical docking studies were employed to prioritize candidates for synthesis and experimental validation. Among these, fluorophenyl-5-methylpiperidinopyrazolyl-1*H*-pyrrolo[2,3-*b*]pyridine **1** emerged as the top candidate based on its superior docking score and hinge hydrogen bond interactions. Compound **1** was synthesized and pharmacologically evaluated, demonstrating potent DYRK1A inhibitory activity at the nanomolar level.

Further exploration of this novel compound family resulted in synthesising and evaluating derivatives (**5-11**), all of which exhibited comparable efficacy. These derivatives also possess additional antioxidant and anti-inflammatory properties, broadening their therapeutic potential.

In conclusion, this study successfully identified a novel DYRK1A inhibitor with nanomolar potency using AI-guided methodologies and established a new family of pyrazolylpyrrolo[2,3-*b*]pyridine derivatives with promising pharmacological profiles, paving the way for further exploration in therapeutic applications.

Experimental section

AI-assisted *de-novo* design

This section provides a comprehensive and detailed account of the complete process undertaken with the *de novo* generative models that led to the identification of the proposed candidate molecules. It includes an outline of the key characteristics of the dataset used, its application in the generative workflow, and additional details necessary

to facilitate the reproduction of similar procedures in analogous contexts.

Datasets

Two datasets containing molecules in SMILES format were utilized in this work:

- The primary dataset consists of 1,782 active inhibitors targeting DYRK1A, sourced from the ChEMBL database.⁵³ It includes pharmacological activity data for each molecule measured using various methods, such as K_i , K_d , IC_{50} , or EC_{50} . These values are converted to pChEMBL scores, defined as the negative \log_{10} of the molar concentration.
- The Tox21 dataset,⁵⁴ comprising 12,060 training samples with 12 binary labels representing the outcomes of 12 distinct toxicological experiments.

AI-based QSAR models

We aim to develop new compounds with high affinity for DYRK1A while maintaining non-toxic profiles, framing this as a dual-target problem. To achieve this, we first construct QSAR models for each property of interest (*affinity* and *toxicity*), enabling us to evaluate the quality of the proposed molecules.

1. **Affinity target:** To construct the *affinity model*, we employed several approaches trained on the primary DYRK1A database. Compared to typical datasets used for QSAR model development, this dataset is relatively small, presenting challenges in achieving high-quality predictions. The pChEMBL values were derived from various measurement methods (K_i , K_d , IC_{50} , or EC_{50}), leading to observable differences among recorded values solely due to the measurement method, as illustrated in Figure 8. To account for this confounding factor, each measurement type was standardized and normalized separately using a Box-Cox transformation,⁵⁵ and the standardized target values were used for performance assessments. This correction was also considered when evaluating each model's performance.

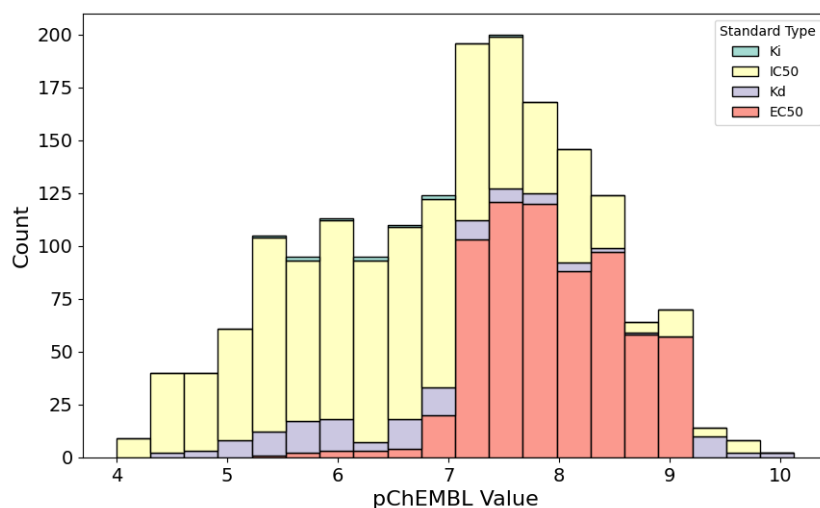


Figure 8: Stacked-bars plot of the frequencies of DYRK1A inhibitors affinity featured in the database by measurement type.

We employed multiple approaches to develop the predictive affinity model, integrating various molecular representations for each compound to improve accuracy. This process involved selecting the most suitable descriptors for each model, tailored to its specific requirements. The available descriptors are:

- **Graph:** This representation models a molecule as a graph,⁵⁶ where atoms are nodes and bonds are edges, effectively capturing molecular connectivity and structural relationships.
- **Morgan:** (*Morgan fingerprints*⁵⁷) Circular fingerprints that encode the local environment around each atom, capturing atom neighborhoods. These are widely used for similarity searches and structure-activity relationship (SAR) analysis.
- **Rdkitfpbits:**⁵⁸ A representation using bit vectors that denote specific substructures and functional groups within a molecule, enabling rapid identification of molecular features.
- **M3C:** A frequency-based encoding that quantifies how often each substructure appears within a molecule. This descriptor provides a detailed measure of molecular features and is obtained using the `DescriptaStorus` package.

- **SPGNN-e:**³⁶ Learned molecular representations derived from graph neural networks (GNNs). These representations capture complex atomic and bonding relationships in a data-driven manner, enhancing predictive accuracy.

Given the available representations of the dataset, we employed various predictive algorithms to construct our QSAR model. For each candidate model, multiple combinations of molecular descriptors were tested, selecting the configuration that yielded the most favorable results. Hyper-parameter selection was performed using a grid search with 10-fold cross-validation for each algorithm. The algorithms tested for constructing the QSAR model were:

- **MolCLR:**³⁵ A self-supervised learning framework applied to **Graph** representations. It leverages large unlabeled datasets to pre-train graph neural networks through contrastive learning by maximizing agreement between augmented views of the same molecule, thereby learning meaningful molecular representations.
- **SPGNN:**³⁶ A graph neural network model that operates on **Graph** representations, pre-trained with tasks at both the node and the graph level. This approach enhances the model's ability to learn detailed structural relationships.
- **GPs** (Gaussian Processes): Utilizes **M3C** representations to predict molecular properties with uncertainty estimation. This probabilistic model provides confidence intervals for predictions, making it particularly valuable for small datasets.
- **RFs** (Random Forest): Relies on **RDKit** fingerprints to predict molecular properties. It constructs an ensemble of decision trees, averaging predictions across multiple trees to produce robust results.
- **MLP** (Multi-layer Perceptron): Combines **M3C** and **RDKit** fingerprints to enhance predictive power. MLP captures complex relationships by passing the input through multiple layers of interconnected nodes.

- **KNN** (K-nearest Neighbors): Uses a combination of **M3C** and **SPGNN-e** representations to predict molecular properties. It determines a molecule's properties by analyzing its closest neighbors in the dataset.
- **SVR** (Support Vector Regressor): Utilizes both **M3C** and **RDKit** fingerprints for regression-based predictions. SVR identifies a hyperplane in high-dimensional space that best fits the data points for accurate property estimation.
- **Chemprop**:²³ A directed message-passing neural network that uses **Graph** representations to predict molecular properties. It computes edge embeddings through message passing and aggregates them into a molecular embedding for prediction.
- **XGBoost** (Extreme Gradient Boosting):³⁴ Employs both **Morgan** and **RDKit** representations to deliver highly accurate and efficient predictions. XGBoost constructs an ensemble of weak learners using gradient boosting, iteratively refining the model to improve performance.

Based on the performance metrics, we constructed our primary predictive model as an ensemble consisting of XGBoost,³⁴ Support Vector Regressors (SVR), K-nearest Neighbors (KNN), and the **Chemprop** directed message-passing neural network.²³ SVR contributed strong regression capabilities by employing kernel methods to model non-linear relationships using M3C and RDKit fingerprints. KNN, combining M3C and SPGNN-e representations, captured local molecular similarities by predicting properties based on the nearest neighbors in the dataset. XGBoost, utilizing Morgan and RDKit fingerprints, provided robust and efficient predictions through its gradient boosting algorithm, which iteratively and effectively improves weak learners. Finally, **Chemprop** enhanced the ensemble with deep learning-based structural insights by leveraging graph representations and directed message-passing mechanisms. This diverse combination allowed the ensemble to produce robust and reliable predictions.

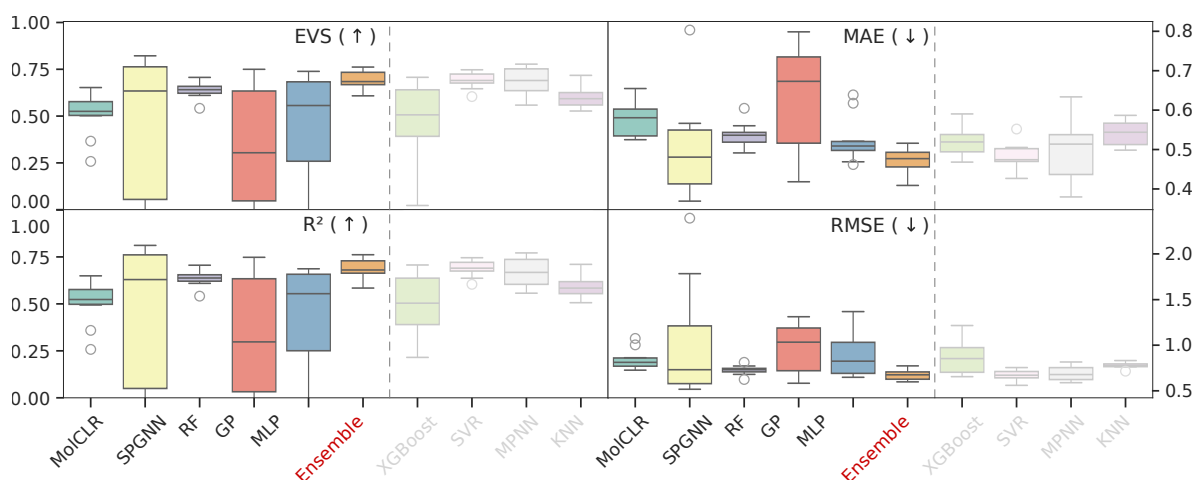


Figure 9: Metric-wise results for each method. Each metric is detailed at the top of each sub-figure, indicating whether it should be maximized (\uparrow) or minimized (\downarrow). The individual models that constitute the ensemble are included on the right of each plot after the dashed line.

We employed the ensemble to predict the performance of each molecule and filter potential candidates. If the predictions from the models within the ensemble differed by more than a specified threshold, the proposed compound was discarded. This served as a consistency check, ensuring alignment among the models regarding the predicted affinity value of promising compounds. This heuristic rule provided greater stability during the subsequent generative phases. Initially, the threshold was set to approximately 1 unit for the predicted affinity; however, later experiments indicated that a more flexible (higher) threshold could also be effective.

Figure 9 presents the performance of all models evaluated using four key metrics: EVS, MAE, R^2 , and RMSE. The EVS quantifies the proportion of variance explained by the model, with higher values indicating better performance; MAE measures the average absolute error between predicted and actual values, while R^2 assesses the goodness of fit; RMSE, similar to MAE, evaluates prediction errors while placing greater emphasis on larger deviations. The objective is to maximize EVS and R^2 while minimizing MAE and RMSE. As illustrated, the ensemble consistently delivered superior median performance across all metrics, with minimal variance across the dataset. Based on these results, this ensemble was selected as the QSAR model for this work. Figure 3 provides additional insights by sum-

marizing the methods through their mean rankings and corresponding standard deviations, offering a clearer comparison.

2. **Toxicity target:** To complete the binary target QSAR model, we developed a model to predict the toxicity of each compound. For this task, we utilized the larger Tox21 dataset.⁵⁴ In this dataset, each compound is assessed for toxicity across 12 biomarkers, such as the aryl hydrocarbon receptor (AhR) and the estrogen receptor (ER). The selected model is the Directed Message-Passing Neural Network, implemented within the Chemprop package,²³ a state-of-the-art algorithm for molecular property prediction. This model operates on molecular graphs, passing messages between atoms and bonds to capture intricate structural relationships and predict chemical properties.

The output of the model is an array of 12 probability values, each representing the likelihood (ranging from 0 to 1) of the compound belonging to a specific toxicity class. This model is employed both as a predictive tool and as a filter in the generative process, requiring candidate molecules to achieve a probability of toxicity below 0.5 across all 12 metrics. This serves as an initial screening step to ensure that generated compounds do not exhibit toxic traits, thereby aligning with the desired drug-like properties. While the cutoff at 0.5 has yielded promising experimental results, stricter thresholds can be applied for individual toxicity labels if a more conservative approach is desired.

In addition to the QSAR models, we implemented a **similarity** function to enhance the diversity of the proposed molecules. Specifically, the Tanimoto similarity metric was employed to prevent the generation of molecules that were overly similar to one another. For each candidate molecule, its similarity was calculated against all compounds in the existing database, ensuring that the maximum similarity value remained below 0.5. This threshold can also be adjusted to promote an even greater diversity in the exploration of the chemical space, depending on the specific goals of the study.

Together, these QSAR models collectively predict the properties of the proposed compounds and form the foundation for the filters used to screen molecules, which are ex-

plicitly detailed in the section on screening filters.

Generative models

To generate new candidate molecules, we primarily relied on pre-trained generative models, as these typically require large datasets for effective training. This approach enabled us to generate high-quality candidates despite the limited data availability. Given our specific dataset constraints, we repurposed pre-existing, pre-trained approaches to suit our needs. Among the models considered, the HGG³⁷ proved particularly effective due to its ability to process complex molecular data and produce viable molecular structures that met our stringent criteria for affinity, toxicity, and novelty.

The HGG model is a hierarchical graph encoder-decoder model that constructs molecules using structural motifs as building blocks. Initially trained on a dataset containing SMILES representations of DYRK1A inhibitors, the model generated five batches of 10,000 molecules, iteratively filtering them for binding affinity, toxicity, and structural similarity to known inhibitors. Molecules passing all filtering criteria were reintegrated into the training dataset, enabling the model to retrain and progressively refine the candidate list over five iterations. This iterative approach improved the quality of the final molecule selection, ensuring that each candidate satisfied stringent standards for chemical properties and drug-likeness. While applying such a recursive process blindly could raise concerns about overfitting, in our case, the limited number of iterations ensured the results remained focused on the relevant regions of the chemical space for this specific task. Furthermore, expert chemists reviewed the resulting structures and deemed them both synthetically feasible and chemically interesting. Subsequent experimental validation confirmed these initial evaluations, reinforcing the effectiveness of the HGG model.

For comparison, we briefly explored other models, including Pocket2Mol,⁵⁹ an E(3)-equivariant generative network leveraging protein pocket data, as well as several additional algorithms. These included genetic algorithms like the Reinforced Genetic Algorithm (RGA)⁶⁰ and diffusion models such as DiffSBDD.⁶¹ Although some methods

showed potential (*e.g.* Pocket2Mol) and should be further explored in extensions of this current work, others frequently produced molecules with less desirable chemical properties, making them unsuitable for further development.

Finally, the candidates generated by the HGG model underwent conventional molecular docking calculations, yielding scores that surpassed those of the reference ligand. The top-scoring molecules from this process were filtered and ranked using the QSAR ensemble model, with the top 9 compounds displayed in Table 1.

Screening filters

To screen the generated molecules, we apply FOUR different types of filters. When needed, we will refer to the predictive QSAR models for pChEMBL affinity and toxicity as f_{pch} , f_{tox} , respectively.

- **Affinity filter:** Given a molecule G , its predicted affinity (pChEMBL) must be higher than the third quartile Q_3 of our primary dataset:

$$f_{pch}(G) > Q_3.$$

This ensures that the molecules generated are somewhat promising candidates for our affinity target.

- **Toxicity filter:** G must be classified as nontoxic in all 12 toxicity classes:

$$\sum_{i=1}^{12} f_{tox}^{(i)}(G) = 0, \quad \text{where} \quad f_{tox}^{(i)}(G) = \begin{cases} 0 & \text{if } p(tox_{(i)}|G) < 0.5, \\ 1 & \text{if } p(tox_{(i)}|G) \geq 0.5. \end{cases}$$

where $p(tox_{(i)}|G)$ is the estimated probability that the compound G is deemed toxic in the i -th category across all 12 available toxicity classes $i \in \{1, \dots, 12\}$. As mentioned earlier in the article, this 0.5 threshold can be changed if a more conservative estimate of the toxicity of the compounds is needed.

- **Similarity filter:** We define the similarity between a molecule G and our primary

dataset D through:

$$s_D(G) = \max_{G' \in D} \{T(G, G')\},$$

where T is the Tanimoto coefficient between two molecules. To ensure that a molecule G is sufficiently different from the known inhibitors of the dataset, its similarity will have to be less than a predefined value:

$$s_D(G) < 0.5$$

Increasing this threshold encourages the model to explore more diverse regions of the chemical space but comes with an increased risk of generating non-chemically viable compounds. Conversely, selecting lower values keeps the model closer to the existing dataset, prioritizing chemically sound candidates. Empirically, a threshold of 0.5 provided a good balance between these two behaviors. However, depending on the nature of the task and the available data, alternative threshold values may provide a more suitable exploration of the chemical space.

- **Exploitability filter:** Given a molecule G , the variance of the predicted affinities in the ensemble model, $\sigma_{ens}(G)$, must not exceed a pre-set threshold, σ_{thr} . We consider $\sigma_{thr} = 1$ for our experiments, although further tests suggest that larger and more permissive thresholds may also work well. This helps ensure certain stability regarding the proposed candidate molecules so no single candidate presents structures that exploit particular parts of the ensemble model.

$$\hat{\sigma}_{ens}(G) < \sigma_{thr}$$

This highlights one of the key strengths of the ensemble model in this context: by incorporating diverse methods within the ensemble, enforcing this condition makes it challenging for the generative process to propose a candidate compound that exploits the specific formulation of any single algorithm. Instead, the compound must perform well across the other components of the ensemble. Therefore, we

consider the diversity and collective performance of the ensemble model to be a crucial aspect of our generative process.

Molecular Docking

For implementing molecular docking, the following pipeline was applied.

- **Ligand Preparation:** The conversion from SMILES to SD format was performed using the `structconvert` tool available in the Schrödinger module.⁶² Ligand preparation was conducted using the LigPrep tool included in the Maestro package.^{63,64} Progressive levels were generated, encompassing possible ionization states at physiological pH and potential tautomers. Final energy minimization was carried out using the OPLS4 force field, with default parameters set for stereoisomers.
- **Protein Preparation:** Human DYRK1A (PDB code 4YLL³⁸) underwent preparation for subsequent computational analyses utilizing `Protein Preparation Wizard`,^{65,66} a tool integrated into `Maestro`.⁶⁴ As part of the protocol, the protein structure underwent preprocessing, including bond order assignment and structural adjustments using Prime.^{67–69} Additionally, protonation at pH 7 ± 2 was generated using Epik.^{70,71} Subsequently, optimization of the hydrogen-bonding network and calculation of residue protonation states at pH 7 were performed using PROPKA,⁷² followed by a final restrained minimization employing the OPLS4 force field.
- **Ligand Docking:** The centroid of the crystallized ligand in the catalytic pocket served 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 carried out using the Glide extra precision mode (XP) within the Schrodinger software suite,^{73–77} with no constraints applied. Default parameters were utilized for ligand settings, including flexible ligand sampling and the addition of epik state penalties to 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 4E3 (5t) into the binding site of the crystal structure 4YLL (Table 3).

Table 3: Ligand interactions for redocked 4E3 ligand into 4YLL (HB: Hydrogen bond)

Compound _{Pose}	RMSD (Å)	Docking Score (kcal/mol)	Residue; Interaction; Distance (Å)
4E3 _{Pose 1}	0.715	-8.417	Lys167; Aromatic HB; 2.77 Lys188; HB; 1.81 Lys188; Salt bridge; 3.55 Glu239; Aromatic HB; 2.39 Leu241; Halogen Bond (Br); 2.78
4E3 _{Pose 2}	0.739	-8.247	Lys167; Aromatic HB; 2.76 Lys188; HB; 1.87 Lys188; Salt bridge; 3.61 Glu239; Aromatic HB; 2.35 Leu241; Halogen Bond (Br); 2.70

Chemistry

Chemistry Melting points were determined using a MP70 (Mettler Toledo) apparatus and are uncorrected. ¹H NMR spectra (400 or 500 MHz) and ¹³C NMR spectra (100 or 125 MHz) were recorded on BRUKER AVANCE III HD-400 (400 MHz) and VARIAN SYSTEM-500 (500 MHz) spectrometers and are reported in ppm on the δ scale. The signal of the solvent was used as a reference. High-performance liquid chromatography (HPLC) was performed using a Waters 2695 apparatus with a diode array UV/Vis detector Waters 2996 and coupled to a Waters micromass ZQ using a Sunfire C18 column (4.6 \times 50 mm, 3.5 μ m) at 30°C, with a flow rate of 0.35 mL/min. The mobile phases used were: CH₃CN and 0.1% formic acid in H₂O. Electrospray in positive mode was used for ionisation. The sample injection volume was set to 3 μ L of a solution of 1 mg/mL CH₃CN. Gradient conditions, time of gradient (gt) and time of retention (rt) are specified for each case and a different gradient elution was specified for each case. Flash chromatography

was performed in an Isolera Prime (Biotage) equipment with a variable detector, using silica gel 60 (230 – 400 mesh) cartridges or KP C18-HS cartridges, both from Biotage. Elemental analyses were performed on a Heraeus CHN-O Rapid analyzer. Reactions heated by microwaves were realized in a Biotage Initiator microwave oven reactor (frequency of 2045 GHz). The purity of all compounds was greater than 95% before biological testing (SI). Reagents and solvents were purchased from common commercial suppliers, mostly Scharlau, BLD and FluoroChem, and were used without further purification. The compound 5-bromo-3-iodo-1-tosyl-1*H*-pyrrolo[2,3-*b*]pyridine (**15**) was prepared from the procedure reported in Goodfellow *et al.*⁷⁸

General procedure for the synthesis of the 5-bromo-3-(aryl)-1-tosyl-1*H*-pyrrolo[2,3-*b*]pyridine compounds **2**, **12-14**.

A microwave vial was charged with 5-bromo-3-iodo-1-tosyl-1*H*-pyrrolo[2,3-*b*]pyridine (**15**), the corresponding aryl-4,4,5,5-tetramethyl-1,3,2-dioxaborolane (**4**, **16-18**), potassium carbonate (K₂CO₃) and [1,1'-Bis(diphenylphosphino)ferrocene] dichloropalladium (II)(Pd(dppf)Cl₂). The vial was sealed with a septum cap and purged with argon. 1,4-dioxane and water were added. The mixture was stirred at rt and bubbled with argon during 5 min. The reaction mixture was irradiated in a microwave for 2h at 100°C. The crude reaction mixture was diluted with dichloromethane (CH₂Cl₂) and filtered. The solvents were evaporated under a vacuum, and the product was purified by flash chromatography (0~15% EtOAc in hexane).

5-bromo-3-(3-fluorophenyl)-1-tosyl-1*H*-pyrrolo[2,3-*b*]pyridine (2**)**. From 2-(3-fluorophenyl)-4,4,5,5-tetramethyl-1,3,2-dioxaborolane (**4**) (53 mg, 0.25 mmol), 5-bromo-3-iodo-1-tosyl-1*H*-pyrrolo[2,3-*b*]pyridine (**15**) (102 mg, 0.21 mmol), K₂CO₃ (116mg, 0.84 mmol, 4% eq) and Pd(dppf)Cl₂ (8.1mg, 0.011 mmol, 5% eq), 1,4-dioxane (3 mL) and H₂O (0.5 mL). Yield: (51 mg, 55%). Mp: 163.8 – 164.4 °C. ¹H NMR: CDCl₃ (400 MHz) δ: 8.50 (d, 1H, 6-H); 8.20 (d, 1H, 4-H); 8.09 (d, 2H, Ts); 7.90 (s, 1H, 2-H); 7.47 – 7.42 (m, 1H, Ar); 7.34 - 7.30 (m, 3H, Ar, Ts); 7.26 – 7.22 (m, 1H, Ar); 7.11 – 7.06 (m, 1H, Ar); 2.39 (s, 3H, CH₃). ¹³C NMR: CDCl₃ (100 MHz) δ: 163.3 (d, J = 246 Hz; Ph); 146.1

(C-6); 145.9 (Ts); 145.8 (C-7a); 135.0 (Ts); 134.3 (d, $J = 8$ Hz; Ph); 131.1 (C-4); 131.0 (d, $J = 8$ Hz; Ph); 130.0 (2C, Ts); 128.4 (2C, Ts); 124.5 (C-2); 123.2 (d, $J = 3$ Hz; Ph); 122.9 (C-3a); 118.6 (d, $J = 2$ Hz; C-3); 115.8 (C-5); 115.0 (d, $J = 21$ Hz; Ph); 114.4 (d, $J = 22$ Hz; Ph); 21.8 (CH₃). HPLC-MS (ES⁺): CH₃CN/H₂O 60:40 – 95:5, gt: 5 min; rt: 5.97; [M+H]⁺ = 445 / 447.

5-bromo-3-(3-methoxyphenyl)-1-tosyl-1*H*-pyrrolo[2,3-*b*]pyridine (12). From 2-(3-methoxyphenyl)-4,4,5,5-tetramethyl-1,3,2-dioxaborolane (**16**) (96 mg, 0.42 mmol), 5-bromo-3-iodo-1-tosyl-1*H*-pyrrolo[2,3-*b*]pyridine (**15**) (152 mg, 0.32 mmol), K₂CO₃ (227 mg, 1.64 mmol, 4%eq) and Pd(dppf)Cl₂ (12 mg, 0.016 mmol, 5% eq), 1,4-dioxane (3 mL) and H₂O (0.5 mL). Yield: (133 mg, 58%). Mp: 165.8 – 166.3 °C. ¹H NMR: CDCl₃ (400 MHz) δ : 8.48 (d, 1H, 6-H); 8.21 (d, 1H, 4-H); 8.09 (d, 2H, Ts); 7.89 (s, 1H, 2-H); 7.38 (t, 1H, Ar); 7.29 (d, 2H, Ts); 7.13 – 7.11 (m, 1H, Ar); 7.07 – 7.06 (m, 1H, Ar); 6.94 – 6.91 (m, 1H, Ar); 3.87 (s, 3H, OCH₃); 2.38 (s, 3H, CH₃). ¹³C NMR: CDCl₃ (100 MHz) δ : 163.2 (d, $J = 246$ Hz; Ph); 154.6 (CO); 146.3 (C-7a); 145.5 (Ts); 143.3 (C-6); 136.6 (Ind); 135.2 (Ts); 134.9 (d, $J = 8$ Hz; Ph); 130.8 (d, $J = 8$ Hz; Ph); 129.8 (2C, Ts); 128.2 (2C, Ts); 125.1 (C-4); 124.9 (C-2); 124.0 (C-5); 123.8 (Ind); 123.2 (d, $J = 3$ Hz; Ph); 121.4 (C-3a); 119.7 (Ind); 119.1 (d, $J = 3$ Hz; C-3); 114.6 (d, $J = 21$ Hz; Ph); 114.4 (d, $J = 22$ Hz; Ph); 80.0 (OC); 59.7 (CH); 32.5 (Pip); 28.5 (5C, 3*CH₃, Pip); 24.9 (Pip); 21.7 (CH₃). HPLC-MS (ES⁺): CH₃CN/H₂O 60:40 – 95:5, gt: 5 min; rt = 5.84; [M+H]⁺ = 457/459.

5-bromo-3-(3,5-dichlorophenyl)-1-tosyl-1*H*-pyrrolo[2,3-*b*]pyridine (13). From 2-(3,5-dichlorophenyl)-4,4,5,5-tetramethyl-1,3,2-dioxaborolane (**17**) (249 mg, 0.91 mmol), 5-bromo-3-iodo-1-tosyl-1*H*-pyrrolo[2,3-*b*]pyridine (**15**) (400 mg, 0.83 mmol), K₂CO₃ (573 mg, 4.15 mmol, 5 eq) and Pd(dppf)Cl₂ (32 mg, 0.04 mmol, 5% eq), 1,4-dioxane (3 mL) and H₂O (0.5 mL). Yield: (297 mg, 72%). Mp: 187.6 – 188.2 °C. ¹H NMR: CDCl₃ (400 MHz) δ : 8.52 (d, 1H, 6-H); 8.15 (d, 1H, 4-H); 8.10 (d, 2H, Ts); 7.91 (s, 1H, 2-H); 7.42 (d, 2H, Ar); 7.37 (t, 1H, Ph); 7.31 (d, 2H, Ts); 2.40 (s, 3H, CH₃). ¹³C NMR: CDCl₃ (100 MHz) δ : 146.4 (C-6); 146.0 (Ts); 145.7 (C-7a); 135.9 (2C, Ph); 135.2 (Ts); 134.8 (Ph); 130.7 (C-4); 130.0 (2C, Ts); 128.4 (2C, Ts); 128.0 (Ph); 125.8 (2C, Ph); 125.1 (C-2); 122.4

(C-3a); 117.1 (C-3); 115.9 (C-5); 21.8 (CH₃). HPLC-MS (ES⁺): CH₃CN/H₂O 80:20 – 95:5, gt: 5 min; rt = 4.27; [M+H]⁺ = 497.

5-bromo-3-(2-fluoro-5-methoxy-4-methylphenyl)-1-tosyl-1*H*-pyrrolo[2,3-*b*]pyridine (14). From 2-(2-fluoro-5-methoxy-4-methylphenyl)-4,4,5,5-tetra-methyl-1,3,2-dioxaborolane (**18**) (149 mg, 0.56 mmol), 5-bromo-3-iodo-1-tosyl-1*H*-pyrrolo[2,3-*b*]pyridine (**15**) (200 mg, 0.41 mmol), K₂CO₃ (283 mg, 2.05 mmol, 4 eq) and Pd(dppf)Cl₂ (15 mg, 0.02 mmol, 5% eq), 1,4-dioxane (3 mL) and H₂O (0.5 mL). Yield: (113 mg, 55%). Mp: 169.3 – 169.6 °C. ¹H NMR: CDCl₃ (400 MHz) δ: 8.48 (d, 1H, 6-H); 8.10 (d, 2H, Ts); 8.06 (t, 1H, 4-H); 7.90 (s, 1H, 2-H); 7.30 (d, 2H, Ts); 7.00 (d, 1H, Ar); 6.85 (d, 1H, Ar); 3.86 (s, 3H, OCH₃); 2.39 (s, 3H, CH₃); 2.26 (s, 3H, CH₃). ¹³C NMR: CDCl₃ (100 MHz) δ: 154.3 (d, J = 2 Hz; Ph); 153.7 (d, J = 239 Hz; Ph); 145.8 (C-6); 145.8 (Ts); 145.5 (C-7a); 135.0 (Ts); 131.9 (d, J = 5 Hz; C-4); 129.9 (2C, Ts); 128.9 (d, J = 8 Hz; Ph); 128.4 (2C, Ts); 125.7 (d, J = 3 Hz; C-2); 123.6 (C-3a); 118.4 (d, J = 24 Hz; Ph); 116.6 (d, J = 16 Hz; Ph); 115.6 (C-5); 114.3 (C-3); 110.9 (d, J = 4 Hz; Ph); 56.2 (OCH₃); 21.8 (CH₃); 16.3 (CH₃). HPLC-MS (ES⁺): CH₃CN/H₂O 60:40 – 95:5, gt: 5 min; rt = 7.02; [M+H]⁺ = 489/491.

General procedure for the synthesis of the tert-butyl 4-(4-(3-(aryl)-1-tosyl-1*H*-pyrrolo[2,3-*b*]pyridin-5-yl)-1*H*-pyrazol-1-yl)piperidine-1-carboxylate compounds 19-22

A microwave vial was charged with the corresponding 5-bromo derivative (**2,12-14**), tert-butyl 4-(4-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)-1*H*-pyrazol-1-yl)piperidine-1-carboxylate, potassium carbonate (K₂CO₃) and [1,1'-Bis(diphenylphosphino)ferrocene]dichloropalladium(II) (Pd(dppf)Cl₂). The vial was sealed with a septum cap and purged with argon. 1, 4-dioxane and water were added. The mixture was stirred at rt and bubbled with argon during 5 min. The reaction mixture was irradiated in microwave for 2 h at 100 °C. The crude reaction mixture was diluted with dichloromethane (CH₂Cl₂) and filtered. The solvents were evaporated under a vacuum, and the product

was purified by flash chromatography (0 – 50 % EtOAc in hexane).

Tert-butyl 4-(4-(3-(3-fluorophenyl)-1-tosyl-1*H*-pyrrolo[2,3-*b*]pyridin-5-yl)-1*H*-pyrazol-1-yl)piperidine-1-carboxylate (19). From 5-bromo-3-(3-fluoro-phenyl)-1-tosyl-1*H*-pyrrolo[2,3-*b*]pyridine (**2**) (90 mg, 0.2 mmol), tert-butyl 4-(4-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)-1*H*-pyrazol-1-yl)piperidine-1-carboxylate (**23**) (87 mg, 0.23 mmol), K₂CO₃ (126 mg, 0.91 mmol, 4.6 eq) and Pd(dppf)Cl₂ (8.0 mg, 0.011 mmol, 5% eq), dioxane (3 mL) and H₂O (0.5 mL). Yield: (116 mg, 94%). Mp: 106.8 – 107.2 °C. ¹H NMR: CDCl₃ (400 MHz) δ: 8.61 (d, 1H, 6-H); 8.13 (d, 2H, Ts); 8.08 (d, 1H, 4-H); 7.89 (s, 1H, 2-H); 7.79 (d, 1H, Ind); 7.73 (d, 1H, Ind); 7.49 – 7.43 (m, 1H, Ar); 7.40 - 7.37 (m, 1H, Ar); 7.32 – 7.29 (m, 3H, Ar, Ts); 7.11 – 7.08 (m, 1H, Ar); 4.37 – 4.28 (m, 3H, Pip); 2.96 - 2.89 (m, 2H, Pip); 2.38 (s, 3H, CH₃); 2.19 (d, 2H, Pip); 2.03 – 1.93 (m, 2H, Pip); 1.49 (s, 9H, CH₃). ¹³C NMR: CDCl₃ (100 MHz) δ: 163.2 (d, J = 246 Hz; Ph); 154.6 (CO); 146.3 (C-7a); 145.5 (Ts); 143.3 (C-6); 136.7 (Ind); 135.4 (Ts); 134.9 (d, J = 8 Hz; Ph); 130.8 (d, J = 8 Hz; Ph); 129.8 (2C, Ts); 128.2 (2C, Ts); 125.1 (C-4); 124.9 (C-5); 124.0 (Ind); 123.2 (C-2); 123.2 (d, J = 3 Hz; Ph); 121.4 (C-3a); 119.7 (Ind); 119.1 (d, J = 2 Hz; C-3); 114.6 (d, J = 21 Hz; Ph); 114.4 (d, J = 22 Hz; Ph); 80.0 (OC); 59.7 (CH); 32.5 (Pip); 24.5 (5C, 3*CH₃, Pip); 24.9 (Pip); 21.7 (CH₃). HPLC-MS (ES⁺): CH₃CN/H₂O 60:40 – 95:5, gt: 5 min; rt: 5.77; [M+H]⁺ = 616.

Tert-butyl 4-(4-(3-(3-methoxyphenyl)-1-tosyl-1*H*-pyrrolo[2,3-*b*]pyridin-5-yl)-1*H*-pyrazol-1-yl)piperidine-1-carboxylate (20). From 5-bromo-3-(3-methoxyphenyl)-1-tosyl-1*H*-pyrrolo[2,3-*b*]pyridine (**12**) (150 mg, 0.33 mmol), tert-butyl 4-(4-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)-1*H*-pyrazol-1-yl)piperidine-1-carboxylate (**23**) (137 mg, 0.36 mmol), K₂CO₃ (137 mg, 0.99 mmol, 3 eq) and Pd(dppf)Cl₂ (12.1 mg, 0.017 mmol, 5% eq), dioxane (3 mL) and H₂O (0.5 mL). Yield: (176 mg, 85%). Mp: 88.5 – 89.2 °C. ¹H NMR: CDCl₃ (400 MHz) δ: 8.59 (d, 1H, 6-H); 8.11 (d, 2H, Ts); 8.08 (d, 1H, 4-H); 7.86 (s, 1H, 2-H); 7.77 (d, 1H, Ind); 7.67 (d, 1H, Ind); 7.41 (t, 1H, Ar); 7.29 (d, 2H, Ts); 7.18 (m, 1H, Ar); 7.11 (m, 1H, Ar); 6.93 (m, 1H, Ar); 4.34 – 4.26 (m, 3H, Pip); 3.88 (s, 3H, OCH₃); 2.90 (t, 2H, Pip); 2.37 (s, 3H, CH₃); 2.18 – 2.14 (m, 2H, Pip); 2.01 – 1.90 (m, 2H, Pip); 1.48 (s, 9H, CH₃). ¹³C NMR: CDCl₃ (100 MHz) δ: 160.3

(Ph); 154.7 (CO); 146.5 (C-7a); 145.4 (Ts); 143.2 (C-6); 136.7 (Ind); 135.5 (Ts); 134.1 (Ph); 130.3 (Ph); 129.8 (2C, Ts); 128.2 (2C, Ts); 125.5 (C-4); 124.8 (C-2); 124.0 (C-5); 123.5 (Ind); 121.8 (C-3a); 120.3 (Ph); 120.1 (Ind); 119.9 (C-3); 113.5 (Ph); 113.1 (Ph); 80.1 (OC); 59.8 (CH); 55.6 (OCH₃); 32.6 (2C, Pip); 28.6 (5C, 3*CH₃, Pip); 21.8 (CH₃). HPLC-MS (ES⁺): CH₃CN/H₂O 60:40 – 95:5, gt: 5 min; rt = 5.61; [M+H]⁺ = 628. **Tert-butyl 4-(4-(3-(3,5-dichlorophenyl)-1-tosyl-1*H*-pyrrolo[2,3-*b*]pyridin-5-yl)-1*H*-pyrazol-1-yl)piperidine-1-carboxylate (21)**. From 5-bromo-3-(3,5-di-chloro phenyl)-1-tosyl-1*H*-pyrrolo[2,3-*b*]pyridine (**13**) (182 mg, 0.37 mmol), tert-butyl 4-(4-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)-1*H*-pyrazol-1-yl)-piperidine-1-carboxylate (**23**) (153 mg, 0.41 mmol), K₂CO₃ (153 mg, 1.11 mmol, 3 eq) and Pd(dppf)Cl₂ (14 mg, 0.019 mmol, 5% eq), dioxane (3 mL) and H₂O (0.5 mL). Yield: (173 mg, 71%). Mp: 117.2 – 117.6 °C. ¹H NMR: CDCl₃ (400 MHz) δ: 8.54 (d, 1H, 6-H); 8.10 (d, 2H, Ts); 8.01 (d, 1H, 4-H); 7.86 (s, 1H, 2-H); 7.79 (s, 1H, Ind); 7.76 (s, 1H, Ind); 7.44 (d, 2H, Ar); 7.27 – 7.25 (m, 3H, Ts, Ar); 4.34 – 4.22 (m, 3H, Pip); 2.89 (bs, 2H, Pip); 2.33 (s, 3H, CH₃); 2.13 (d, 2H, Pip); 2.04 – 1.90 (m, 2H, Pip); 1.46 (s, 9H, CH₃). ¹³C NMR: CDCl₃ (100 MHz) δ: 154.2 (CO); 145.7 (C-7a); 145.3 (Ts); 143.2 (C-6); 136.3 (Ind); 135.6 (Ts); 135.3 (2C, Ph); 134.8 (Ph); 129.5 (2C, Ts); 128.0 (2C, Ts); 127.2 (Ph); 125.5 (2C, Ph); 125.5 (C-4); 124.9 (C-2); 124.5 (Ph); 124.2 (C-5); 124.1 (Ind); 121.8 (C-3a); 119.1 (Ind); 117.3 (C-3); 79.5 (OC); 59.3 (CH); 32.1 (2C, Pip); 28.2 (3C, 3*CH₃); 24.6 (2C, Pip); 21.4 (CH₃). HPLC-MS (ES⁺): CH₃CN/H₂O 80:20 – 95:5, gt: 5 min; rt = 3.95; [M+H]⁺ = 666.

Tert-butyl 4-(4-(3-(2-fluoro-5-methoxy-4-methylphenyl)-1-tosyl-1*H*-pyrrolo[2,3-*b*]pyridin-5-yl)-1*H*-pyrazol-1-yl)piperidine-1-carboxylate (22). From 5-bromo-3-(2-fluoro-5-methoxy-4-methylphenyl)-1-tosyl-1*H*-pyrrolo[2,3-*b*]pyridine (**14**) (200 mg, 0.41 mmol), tert-butyl 4-(4-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)-1*H*-pyrazol-1-yl)piperidine-1-carboxylate (**23**) (170 mg, 0.45 mmol), K₂CO₃ (170 mg, 1.23 mmol, 3 eq) and Pd(dppf)Cl₂ (14.6 mg, 0.017 mmol, 5% eq), dioxane (3 mL) and H₂O (0.5 mL). Yield: (195 mg, 73%). Mp: 163.1 – 163.5 °C. ¹H NMR: CDCl₃ (400 MHz) δ: 8.58 (d, 1H, 6-H); 8.12 (d, 2H, Ts); 7.94 (d, 1H, 4-H); 7.87 (s, 1H, 2-H); 7.75 (s, 1H, Ind); 7.66 (s, 1H, Ind); 7.29 (d, 2H, Ts); 7.00 (d, 1H, Ar); 6.89 (d, 1H, Ar); 4.33 – 4.24 (m,

3H, Pip); 3.86 (s, 3H, OCH₃); 2.94 – 2.86 (m, 2H, Pip); 2.37 (s, 3H, CH₃); 2.26 (s, 3H, CH₃); 2.17 – 2.13 (m, 2H, Pip); 2.00 – 1.89 (m, 2H, Pip); 1.47 (s, 9H, CH₃). ¹³C NMR: CDCl₃ (100 MHz) δ: 154.7 (CO); 154.2 (d, J = 2 Hz; Ph); 153.8 (d, J = 239 Hz; Ph); 146.0 (C-7a); 145.5 (Ts); 143.2 (C-6); 136.7 (Ind); 135.4 (Ts); 129.8 (2C, Ts); 128.6 (d, J = 8 Hz; Ph); 128.3 (2C, Ts); 126.1 (d, J = 3 Hz; C-4); 124.9 (d, J = 3 Hz; C-2); 124.7 (C-5); 123.9 (Ind); 122.2 (C-3a); 120.0 (Ind); 118.3 (d, J = 24 Hz, Ph); 117.2 (d, J = 16 Hz; Ph); 114.9 (C-3); 111.1 (d, J = 4 Hz; Ph); 80.1 (OC); 59.7 (CH); 56.2 (OCH₃); 32.5 (2C, Pip); 28.5 (5C, 3*CH₃, Pip); 21.8 (CH₃); 16.2 (d, J = 1 Hz, CH₃). HPLC-MS (ES⁺): CH₃CN/H₂O 60:40 – 95:5, gt: 5 min; rt = 6.58; [M+H]⁺ = 660.

General procedure for the synthesis of the 3-(aryl)-5-(1-(piperidin-4-yl)-1*H*-pyrazol-4-yl)-1-tosyl-1*H*-pyrrolo[2,3-*b*]pyridine compounds 24-27.

Trifluoroacetic acid (TFA) was added to a solution of the corresponding Boc-protected compound (**19-22**) in dichloromethane at rt. The reaction is stirred until the end of the reaction. The solvent was evaporated under a vacuum. The residue was suspended in NaHCO₃ aq 1 M (10 mL) and the resulting suspension was cooled to 4 °C (overnight). The final product was obtained by filtration, washed with water and air dried.

3-(3-fluorophenyl)-5-(1-(piperidin-4-yl)-1*H*-pyrazol-4-yl)-1-tosyl-1*H*-pyrrolo[2,3-*b*]pyridine (24). From tert-butyl 4-(4-(3-(3-fluorophenyl)-1-tosyl-1*H*-pyrrolo[2,3-*b*]pyridin-5-yl)-1*H*-pyrazol-1-yl)piperidine-1-carboxylate (**19**) (106 mg, 0.17 mmol) in CH₂Cl₂ (10 mL) and TFA (2 mL). Yield: (68 mg, 78%). Mp: 141.8 – 142.6 °C. ¹H NMR: CDCl₃ (400 MHz) δ: 8.61 (d, 1H, 6-H); 8.12 (d, 2H, Tos); 8.06 (d, 1H, 4-H); 7.88 (s, 1H, 2-H); 7.77 (s, 1H, Ind); 7.70 (s, 1H, Ind); 7.47 – 7.43 (m, 1H, Ar); 7.39 - 7.37 (m, 1H, Ar); 7.31 – 7.28 (m, 3H, Ar, Tos); 7.11 – 7.06 (m, 1H, Ar); 4.29 – 4.24 (m, 1H, CH); 3.29 – 3.24 (m, 2H, Pip); 2.79 (t, 2H, Pip); 2.38 (s, 3H, CH₃); 2.22 – 2.18 (m, 2H, Pip); 1.98 – 1.88 (m, 2H, Pip). ¹³C NMR: Acetone-d₆ (100 MHz) δ: 164.0 (d, J = 243 Hz; Ph); 146.9 (C-7a); 145.6 (Ts); 143.6 (C-6); 136.9 (Ind); 136.2 (Ts); 136.0 (d, J = 8 Hz; Ph);

131.7 (d, $J = 9$ Hz; Ph); 130.6 (2C, Ts); 129.0 (2C, Ts); 126.5 (C-5); 125.7 (C-4); 125.6 (C-2); 125.0 (Ind); 124.3 (d, $J = 2$ Hz; Ph); 121.9 (C-3a); 119.8 (Ind); 119.7 (d, $J = 2$ Hz; C-3); 115.0 (d, $J = 21$ Hz; Ph); 114.6 (d, $J = 21$ Hz; Ph); 60.2 (CH); 45.6 (2C, Pip); 33.9 (2C, Pip); 21.5 (CH₃). HPLC-MS (ES⁺): CH₃CN/H₂O 15:85 – 95:5, gt: 5 min; rt: 5.09; [M+H]⁺ = 516.

3-(3-methoxyphenyl)-5-(1-(piperidin-4-yl)-1*H*-pyrazol-4-yl)-1-tosyl-1*H*-pyrrolo[2,3-*b*]pyridine (25). From tert-butyl 4-(4-(3-(3-methoxyphenyl)-1-tosyl-1*H*-pyrrolo[2,3-*b*]pyridin-5-yl)-1*H*-pyrazol-1-yl)piperidine-1-carboxylate (**20**) (230 mg, 0.37 mmol) in CH₂Cl₂ (10 mL) and TFA (2 mL). Yield: (192 mg, 98%). Mp: 121.8 – 122.3 °C. ¹H NMR: CDCl₃ (400 MHz) δ : 8.59 (d, 1H, 6-H); 8.12 – 8.09 (m, 3H, Ts, 4-H); 7.86 (s, 1H, 2-H); 7.76 (d, 1H, Ind); 7.69 (d, 1H, Ind); 7.41 (t, 1H, Ar); 7.28 (d, 2H, Ts); 7.18 (d, 1H, Ar); 7.12 – 7.11 (m, 1H, Ar); 6.95 – 6.92 (m, 1H, Ar); 4.28 – 4.22 (m, 1H, Pip); 3.88 (s, 3H, OCH₃); 3.28 – 3.23 (m, 2H, Pip); 2.81 – 2.74 (m, 2H, Pip); 2.37 (s, 3H, CH₃); 2.21 – 2.16 (d, 2H, Pip); 1.98 – 1.87 (m, 2H, Pip). ¹³C NMR: CDCl₃ (100 MHz) δ : 160.3 (Ph); 146.4 (C-7a); 145.4 (Ts); 143.3 (C-6); 136.5 (Ind); 135.5 (Ts); 134.1 (Ph); 130.3 (Ph); 129.8 (2C, Ts); 128.2 (2C, Ts); 125.5 (C-4); 125.0 (C-2); 123.8 (C-5); 123.5 (Ind); 121.8 (C-3a); 120.3 (Ind); 120.1 (C-3); 119.8 (C-3); 113.5 (Ph); 113.1 (Ph); 60.1 (CH); 55.5 (OCH₃); 45.8 (2C, Pip); 34.1 (2C, Pip); 21.8 (CH₃). HPLC-MS (ES⁺): CH₃CN/H₂O 15:85 – 95:5, gt: 5 min; rt = 4.97; [M+H]⁺ = 528.

3-(3,5-dichlorophenyl)-5-(1-(piperidin-4-yl)-1*H*-pyrazol-4-yl)-1-tosyl-1*H*-pyrrolo[2,3-*b*]pyridine (26). From tert-butyl 4-(4-(3-(3,5-dichlorophenyl)-1-tosyl-1*H*-pyrrolo[2,3-*b*]pyridin-5-yl)-1*H*-pyrazol-1-yl)piperidine-1-carboxylate (**21**) (306 mg, 0.46 mmol) in CH₂Cl₂ (10 mL) and TFA (2 mL). Yield: (213 mg, 82%). Mp: 132.9 – 133.5 °C. ¹H NMR: CDCl₃ (400 MHz) δ : 8.62 (d, 1H, 6-H); 8.12 (d, 2H, Ts); 8.01 (d, 1H, 4-H); 7.88 (s, 1H, 2-H); 7.78 (s, 1H, Ind); 7.71 (s, 1H, Ind); 7.46 (s, 2H, Ar); 7.37 (s, 1H, Ar); 7.30 (d, 2H, Ts); 4.31 – 4.23 (m, 1H, Pip); 3.28 – 3.23 (m, 2H, Pip); 2.78 (t, 2H, Pip); 2.38 (s, 3H, CH₃); 2.21 – 2.17 (m, 2H, Pip); 1.99 – 1.90 (m, 2H, Pip). ¹³C NMR: CDCl₃ (100 MHz) δ : 146.2 (C-7a); 145.7 (Ts); 143.8 (C-6); 136.5 (Ind); 135.9 (Ts); 135.8 (2C, Ph); 135.2 (Ph); 129.9 (2C, Ts); 128.4 (2C, Ts); 127.8 (Ph); 125.9 (2C, Ph); 125.3 (C-4);

124.9 (C-2); 124.3 (C-5); 123.9 (Ind); 121.0 (C-3a); 119.5 (Ind); 117.8 (C-3); 60.2 (CH); 45.8 (2C, Pip); 34.1 (2C, Pip); 21.8 (CH₃). HPLC-MS (ES⁺): CH₃CN/H₂O 15:85 – 95:5, gt: 5 min; rt = 5.41; [M+H]⁺ = 566 / 568.

3-(2-fluoro-5-methoxy-4-methylphenyl)-5-(1-(piperidin-4-yl)-1*H*-pyrazol-4-yl)-1-tosyl-1*H*-pyrrolo[2,3-*b*]pyridine (27). From tert-butyl 4-(4-(3-(2-fluoro-5-methoxy-4-methylphenyl)-1-tosyl-1*H*-pyrrolo[2,3-*b*]pyridin-5-yl)-1*H*-pyrazol-1-yl)piperidine-1-carboxylate (**22**) (246 mg, 0.37 mmol) in CH₂Cl₂ (10 mL) and TFA (2 mL). Yield: (203 mg, 98%). Mp: 117.9 – 118.6 °C. ¹H NMR: CDCl₃ (400 MHz) δ: 8.59 (d, 1H, 6-H); 8.12 (d, 2H, Ts); 7.95 (t, 1H, 4-H); 7.88 (s, 1H, 2-H); 7.74 (s, 1H, Ind); 7.68 (s, 1H, Ind); 7.29 (d, 2H, Ts); 7.01 (d, 1H, Ar); 6.90 (d, 1H, Ar); 4.29 – 4.21 (m, 1H, CH); 3.86 (s, 3H, OCH₃); 3.28 – 3.24 (m, 2H, Pip); 2.78 (td, 2H, Pip); 2.37 (s, 3H, CH₃); 2.27 (s, 3H, CH₃); 2.20 – 2.16 (m, 2H, Pip); 1.95 – 1.91 (m, 2H, Pip). ¹³C NMR: CDCl₃ (100 MHz) δ: 154.2 (d, J = 2 Hz; Ph); 153.8 (d, J = 239 Hz; Ph); 146.0 (C-7a); 145.4 (Ts); 143.2 (C-6); 136.4 (Ind); 135.4 (Ts); 129.8 (2C, Ts); 128.6 (d, J = 8 Hz; Ph); 128.3 (2C, Ts); 126.1 (d, J = 4 Hz; C-4); 124.9 (d, J = 3 Hz; C-2); 124.8 (C-5); 123.7 (Ind); 122.2 (C-3a); 119.8 (Ind); 118.4 (d, J = 23 Hz, Ph); 117.2 (d, J = 16 Hz; Ph); 115.0 (C-3); 111.2 (d, J = 5 Hz; Ph); 60.0 (CH); 56.2 (OCH₃); 45.8 (2C, Pip); 34.0 (2C, Pip); 21.8 (CH₃); 16.3 (CH₃). HPLC-MS (ES⁺): CH₃CN/H₂O 15:85 – 95:5, gt: 5 min; rt: 5.26; [M+H]⁺ = 560.

General procedure for the synthesis of the 3-(aryl)-5-(1-(1-methylpiperidin-4-yl)-1*H*-pyrazol-4-yl)-1-tosyl-1*H*-pyrrolo[2,3-*b*]pyridine compounds 28-31.

Over a solution of the corresponding piperidine derivative (**24-27**) in formic acid at rt, was added, dropwise, formaldehyde 37% aqueous. The mixture was stirred and heated to 70 °C. until the end of the reaction. The solvent was evaporated under a vacuum. A saturated aqueous solution of sodium carbonate (Na₂CO₃) was added (to pH = 10) and the obtained suspension was cooled to 4 °C. (overnight). The final product was obtained by filtration, washed with water and air dried.

3-(3-fluorophenyl)-5-(1-(1-methylpiperidin-4-yl)-1*H*-pyrazol-4-yl)-1-tosyl-1*H*-pyrrolo[2,3-*b*]pyridine (28). From 3-(3-fluorophenyl)-5-(1-(piperidin-4-yl)-1*H*-pyrazol-4-yl)-1-tosyl-1*H*-pyrrolo[2,3-*b*]pyridine (24) (120 mg, 0.23 mmol), formic acid (2 mL) and formaldehyde 37% aqueous (0.7 mL, 4.4 mmol). Time of reaction: 24h. Yield: (111 mg, 89%). Mp: 191.7 – 192.3 °C. ¹H NMR: CDCl₃ (400 MHz) δ: 8.62 (d, 1H, 6-H); 8.13 (d, 2H, Tos); 8.08 (d, 1H, 4-H); 7.89 (s, 1H, 2-H); 7.79 (s, 1H, Ind); 7.73 (s, 1H, Ind); 7.49 – 7.43 (m, 1H, Ar); 7.40 – 7.36 (m, 1H, Ar); 7.32 – 7.29 (m, 3H, Ar, Tos); 7.11 – 7.06 (m, 1H, Ar); 4.20 – 4.14 (m, 1H, CH); 3.00 (d, 2H, Pip); 2.38 (s, CH₃); 2.34 (s, NCH₃); 2.22 – 2.08 (m, 6H, Pip). ¹³C NMR: CDCl₃ (100 MHz) δ: 163.2 (d, J = 245 Hz; Ph); 146.2 (C-7a); 145.4 (Ts); 143.3 (C-6); 136.3 (Ind); 135.2 (Ts); 134.9 (d, J = 8 Hz; Ph); 130.8 (d, J = 9 Hz; Ph); 129.8 (2C, Ts); 128.2 (2C, Ts); 125.1 (C-4); 125.0 (C-2); 123.7 (C-5); 123.6 (Ind); 123.2 (d, J = 3 Hz; Ph); 121.4 (C-3a); 119.6 (Ind); 119.1 (d, J = 3 Hz; C-3); 114.6 (d, J = 21 Hz; Ph); 114.3 (d, J = 22 Hz; Ph); 59.3 (CH); 54.7 (2C, Pip); 46.0 (NCH₃); 32.6 (2C, Pip); 21.7 (CH₃). HPLC-MS (ES⁺): CH₃CN/H₂O 15:85 – 95:5, gt: 5 min; rt = 5.09; [M+H]⁺ = 530.

3-(3-methoxyphenyl)-5-(1-(1-methylpiperidin-4-yl)-1*H*-pyrazol-4-yl)-1-tosyl-1*H*-pyrrolo[2,3-*b*]pyridine (29). From 3-(3-methoxyphenyl)-5-(1-(piperidin-4-yl)-1*H*-pyrazol-4-yl)-1-tosyl-1*H*-pyrrolo[2,3-*b*]pyridine (25) (120 mg, 0.23 mmol), formic acid (2 mL), and formaldehyde 37% aqueous (0.7 mL, 4.4 mmol). Time of reaction: 18 h. Yield: (111 mg, 89%). Mp: 147.9 – 148.6 °C. ¹H NMR: CDCl₃ (400 MHz) δ: 8.60 (d, 1H, 6-H); 8.13 – 8.08 (m, 3H, Ts, 4-H); 7.87 (s, 1H, 2-H); 7.77 (s, 1H, Ind); 7.71 (s, 1H, Ind); 7.41 (t, 1H, Ar); 7.29 (d, 2H, Ts); 7.19 (d, 1H, Ar); 7.13 (s, 1H, Ar); 6.94 (d, 1H, Ar); 4.19 – 4.13 (m, 1H, Pip); 3.88 (s, 3H, OCH₃); 3.01 – 2.98 (m, 2H, Pip); 2.37 (s, 3H, CH₃); 2.34 (s, NCH₃); 2.22 – 2.07 (m, 6H, Pip). ¹³C NMR: CDCl₃ (100 MHz) δ: 160.2 (Ph); 146.4 (C-7a); 145.3 (Ts); 143.2 (C-6); 136.4 (Ind); 135.4 (Ts); 134.0 (Ph); 130.2 (Ph); 129.8 (2C, Ts); 128.2 (2C, Ts); 125.4 (C-4); 124.9 (C-2); 123.7 (C-5); 123.4 (Ind); 121.8 (C-3a); 120.3 (Ph); 120.0 (Ind); 119.8 (C-3); 113.4 (Ph); 113.1 (Ph); 59.3 (CH); 55.4 (OCH₃); 54.7 (2C, Pip); 46.1 (NCH₃); 32.7 (2C, Pip); 21.7 (CH₃). HPLC-MS (ES⁺): CH₃CN/H₂O 15:85 – 95:5, gt: 5 min; rt = 5.11; [M+H]⁺ = 542.

3-(3,5-dichlorophenyl)-5-(1-(1-methylpiperidin-4-yl)-1*H*-pyrazol-4-yl)-1-tosyl-1*H*-pyrrolo[2,3-*b*]pyridine (30). From 3-(3,5-dichlorophenyl)-5-(1-(piperidin-4-yl)-1*H*-pyrazol-4-yl)-1-tosyl-1*H*-pyrrolo[2,3-*b*]pyridine (**26**) (120 mg, 0.21 mmol), formic acid (5 mL), and formaldehyde 37% aqueous (2 mL, 4.4 mmol). Time of reaction: 28 h. Yield: (103 mg, 84%). Mp: 113.8 – 114.5 °C. ¹H NMR: CDCl₃ (400 MHz) δ: 8.61 (d, 1H, 6-H); 8.12 (d, 2H, Ts); 8.01 (d, 1H, 4-H); 7.88 (s, 1H, 2-H); 7.77 (s, 1H, Ind); 7.70 (s, 1H, Ind); 7.46 (d, 2H, Ar); 7.37 (t, 1H, Ar); 7.30 (d, 2H, Ts); 4.22 – 4.13 (m, 1H, Pip); 3.01 – 2.98 (m, 2H, Pip); 2.38 (s, 3H, CH₃); 2.34 (s, 3H, NCH₃); 2.22 – 2.10 (m, 6H, Pip). ¹³C NMR: CDCl₃ (100 MHz) δ: 146.2 (C-7a); 145.7 (Ts); 143.8 (C-6); 136.5 (Ind); 135.9 (Ts); 135.8 (2C, Ph); 135.2 (Ph); 129.9 (2C, Ts); 128.4 (2C, Ts); 127.8 (Ph); 125.9 (2C, Ph); 125.4 (C-4); 124.9 (C-2); 124.3 (C-5); 124.0 (Ind); 121.0 (C-3a); 119.5 (Ind) 117.8 (C-3); 59.5 (CH); 54.8 (2C, Pip); 46.2 (NCH₃); 32.8 (2C, Pip); 21.8 (CH₃). HPLC-MS (ES⁺): CH₃CN/H₂O 15:85 – 95:5, gt: 5 min; rt = 5.47; [M+H]⁺ = 580 / 582.

3-(2-fluoro-5-methoxy-4-methylphenyl)-5-(1-(1-methylpiperidin-4-yl)-1*H*-pyrazol-4-yl)-1-tosyl-1*H*-pyrrolo[2,3-*b*]pyridine (31). From 3-(2-fluoro-5-methoxy-4-methylphenyl)-5-(1-(piperidin-4-yl)-1*H*-pyrazol-4-yl)-1-tosyl-1*H*-pyrrolo[2,3-*b*]pyridine (**27**) (103 mg, 0.18 mmol), formic acid (5 mL) and formaldehyde 37% aqueous (2 mL, 4.4 mmol). Time of reaction: 18 h. Yield: (90 mg, 87%). Mp: 113.9 – 114.6 °C. ¹H NMR: CDCl₃ (400 MHz) δ: 8.58 (d, 1H, 6-H); 8.12 (d, 2H, Ts); 7.94 (t, 1H, 4-H); 7.88 (s, 1H, 2-H); 7.73 (s, 1H, Ind); 7.66 (s, 1H, Ind); 7.29 (d, 2H, Ts); 7.01 (d, 1H, Ar); 6.90 (d, 1H, Ar); 4.17 – 4.12 (m, 1H, CH); 3.86 (s, 3H, OCH₃); 2.99 – 2.96 (m, 2H, Pip); 2.37 (s, 3H, CH₃); 2.33 (s, NCH₃); 2.27 (s, 3H, CH₃); 2.19 – 2.07 (m, 6H, Pip). ¹³C NMR: CDCl₃ (100 MHz) δ: 154.2 (d, J = 1 Hz; Ph); 153.8 (d, J = 239 Hz; Ph); 146.0 (C-7a); 145.4 (Ts); 143.2 (C-6); 136.4 (Ind); 135.4 (Ts); 129.9 (2C, Ts); 128.6 (d, J = 8 Hz; Ph); 128.3 (2C, Ts); 126.1 (d, J = 5 Hz; C-4); 124.9 (d, J = 3 Hz; C-2); 124.8 (C-5); 123.6 (Ind); 122.2 (C-3a); 119.9 (Ind); 118.4 (d, J = 24 Hz, Ph); 117.3 (d, J = 16 Hz; Ph); 115.0 (C-3); 111.2 (d, J = 5 Hz; Ph); 59.4 (CH); 56.2 (OCH₃); 54.8 (2C, Pip); 46.2 (NCH₃); 32.8 (2C, Pip); 21.8 (CH₃); 16.3 (CH₃). HPLC-MS (ES⁺): CH₃CN/H₂O 15:85 – 95:5, gt: 5 min; rt = 5.28; [M+H]⁺ = 574.

General procedure for the synthesis of the 3-(aryl)-5-(1-(1-methylpiperidin-4-yl)-1*H*-pyrazol-4-yl)-1*H*-pyrrolo[2,3-*b*]pyridine compounds 1,5-11.

A solution of the corresponding tosyl derivative (**24-31**) in 0.4 M NaOH methanolic solution was stirred at room temperature until the end of the reaction. The solvent was evaporated under a vacuum. Water was added and the obtained suspension was cooled to 4 °C (overnight). The final product was obtained by filtration, washed with water and air dried.

3-(3-fluorophenyl)-5-(1-(1-methylpiperidin-4-yl)-1*H*-pyrazol-4-yl)-1*H*-pyrrolo[2,3-*b*]pyridine (1). From 3-(3-fluorophenyl)-5-(1-(1-methylpiperidin-4-yl)-1*H*-pyrazol-4-yl)-1-tosyl-1*H*-pyrrolo[2,3-*b*]pyridine (**28**) (55 mg, 0.1 mmol) and 0.4 M NaOH methanolic solution (20 mL). Time of reaction: 2 h. Yield: (32 mg, 90%). Mp: 234.6 – 235.8 °C. ¹H NMR: DMSO-*d*₆ (400 MHz) δ : 11.99 (bs, 1H, NH); 8.56 (d, 1H, 6-H); 8.41 (d, 1H, 4-H); 8.38 (s, 1H, 2-H); 8.00 (s, 1H, Ind); 7.96 (s, 1H, Ind); 7.59 – 7.47 (m, 3H, Ar); 7.07 (t, 1H, Ar); 4.14 – 4.10 (m, 1H, CH); 2.87 (d, 2H, Pip); 2.21 – 2.31 (s, 3H, CH₃); 2.08 – 1.98 (m, 6H, Pip). ¹³C NMR: DMSO-*d*₆ (100 MHz) δ : 162.8 (d, *J* = 242 Hz; Ph); 147.9 (C-7a); 141.0 (C-6); 137.6 (d, *J* = 8 Hz; Ph); 135.7 (Ind); 130.7 (d, *J* = 9 Hz; Ph); 125.2 (C-2); 124.9 (Ind); 123.3 (C-4); 122.2 (d, *J* = 1 Hz; Ph); 121.7 (C-5); 119.8 (Ind); 117.1 (C-3a); 113.1 (d, *J* = 2 Hz; C-3); 112.6 (d, *J* = 21 Hz; Ph); 112.1 (d, *J* = 21 Hz; Ph); 58.3 (CH); 54.2 (2C, Pip); 45.8 (NCH₃); 32.1 (2C, Pip). HPLC-MS (ES⁺): CH₃CN/H₂O 20:80 – 95:5, gt: 5 min; rt = 4.15; [M+H]⁺ = 376.

3-(3-methoxyphenyl)-5-(1-(1-methylpiperidin-4-yl)-1*H*-pyrazol-4-yl)-1*H*-pyrrolo[2,3-*b*]pyridine (5). From 3-(3-methoxyphenyl)-5-(1-(1-methylpiperidin-4-yl)-1*H*-pyrazol-4-yl)-1-tosyl-1*H*-pyrrolo[2,3-*b*]pyridine (**29**) (75 mg, 0.14 mmol) and 0.4 M NaOH methanolic solution (40 mL). Time of reaction: 4 h. Yield: (42 mg, 85%). Mp: 134.9 - 135.4 °C. ¹H NMR: CDCl₃ (400 MHz) δ : 11.73 (bs, 1H, NH); 8.71 (d, 1H, 6-H); 8.32 (d, 1H, 4-H); 8.28 (d, 1H, Ind); 7.83 (d, 1H, 2-H); 7.51 (d, 1H, Ind); 7.40 (t, 1H, Ar); 7.27 (d, 1H, Ar); 7.21 (s, 1H, Ar); 6.88 (d, 1H, Ar); 4.36 – 4.28 (m, 1H, Pip); 3.89 (s, 3H, OCH₃); 3.15 –

3.11 (m, 2H, Pip); 2.55 – 2.45 (m, 5H, NCH₃, Pip); 2.31 – 2.20 (m, 4H, Pip). ¹³C NMR: CDCl₃ (100 MHz) δ : 160.2 (Ph); 148.6 (C-7a); 141.6 (C-6); 136.6 (Ph); 135.9 (Ind); 130.1 (Ph); 124.7 (C-2); 123.8 (Ind); 123.4 (C-4); 122.2 (C-5); 121.3 (Ph); 119.8 (Ind); 118.6 (C-3a); 116.3 (Ph); 111.6 (C-3); 113.1 (Ph); 111.6 (Ph); 59.7 (CH); 55.5 (OCH₃); 55.1 (2C, Pip); 46.0 (CH₃); 32.6 (2C, Pip). HPLC-MS (ES⁺): CH₃CN/H₂O 20:80 – 95:5, gt: 5 min; rt = 4.12; [M+H]⁺ = 388.

3-(3,5-dichlorophenyl)-5-(1-(1-methylpiperidin-4-yl)-1*H*-pyrazol-4-yl)-1*H*-pyrrolo[2,3-*b*]pyridine (6). From 3-(3,5-dichlorophenyl)-5-(1-(1-methylpiperidin-4-yl)-1*H*-pyrazol-4-yl)-1-tosyl-1*H*-pyrrolo[2,3-*b*]pyridine (**30**) (62 mg, 0.11 mmol) and 0.4 M NaOH methanolic solution (20 mL). Time of reaction: 4 h. Yield: (28 mg, 64%). Mp: 219.8 – 220.5 °C. ¹H NMR: DMSO-*d*₆ (400 MHz) δ : 12.13 (bs, 1H, NH); 8.56 (bs, 1H, 6-H); 8.36 (bs, 2H, 4-H, Ind); 8.07 (s, 1H, 2-H); 7.99 (s, 1H, Ind); 7.78 (bs, 2H, Ar); 7.44 (s, 1H, Ar); 4.13 (bs, 1H, CH); 3.32 (bs, 2H, Pip); 2.87 (bs, 2H, Pip); 2.21 (bs, 2H, Pip); 2.03 (bs, 5H, Pip, NCH₃). ¹³C NMR: DMSO-*d*₆ (100 MHz) δ : 147.9 (C-7a); 141.3 (C-6); 138.9 (Ph); 135.8 (Ind); 134.6 (2C, Ph); 126.4 (C-2); 125.1 (Ph); 124.8 (Ind); 124.4 (2C, Ph); 123.2 (C-4); 122.0 (C-5); 119.6 (Ind); 116.8 (C-3a); 111.6 (C-3); 58.3 (CH); 54.2 (2C, Pip); 45.8 (NCH₃); 32.1 (2C, Pip). HPLC-MS (ES⁺): CH₃CN/H₂O 15:85 – 95:5, gt: 5 min; rt = 4.71; [M+H]⁺ = 426 / 428.

3-(2-fluoro-5-methoxy-4-methylphenyl)-5-(1-(1-methylpiperidin-4-yl)-1*H*-pyrazol-4-yl)-1*H*-pyrrolo[2,3-*b*]pyridine (7). From 3-(2-fluoro-5-methoxy-4-methylphenyl)-5-(1-(1-methylpiperidin-4-yl)-1*H*-pyrazol-4-yl)-1-tosyl-1*H*-pyrrolo[2,3-*b*]pyridine (**31**) (66 mg, 0.12 mmol) and 0.4 M NaOH methanolic solution (20 mL). Time of reaction: 4 h. Yield: (40 mg, 83%). Mp: 226.8 - 227.5 °C. ¹H NMR: DMSO-*d*₆ (400 MHz) δ : 11.93 (bs, 1H, NH); 8.55 (d, 1H, 6-H); 8.31 (s, 1H, Ind); 8.17 (t, 1H, 4-H); 7.93 (s, 1H, Ind); 7.72 (s, 1H, 2-H); 7.15 – 7.11 (m, 2H, Ar); 4.15 – 4.07 (m, 1H, CH); 3.87 (s, 3H, OCH₃); 2.86 (d, 2H, Pip); 2.20 (s, 3H, CH₃); 2.20 (NCH₃); 2.08 – 1.96 (m, 6H, Pip). ¹³C NMR: DMSO-*d*₆ (100 MHz) δ : 153.6 (d, J = 1 Hz; Ph); 152.8 (d, J = 235 Hz; Ph); 147.4 (C-7a); 140.9 (C-6); 135.5 (Ind); 125.9 (d, J = 5 Hz; C-2); 125.1 (d, J = 5 Hz; Ph); 124.7 (Ind); 123.6 (d, J = 4 Hz; C-4); 121.3 (C-5); 119.8 (Ind); 119.6 (d, J = 7 Hz; Ph); 118.0 (C-3a);

117.6 (d, $J = 14$ Hz; Ph); 111.2 (d, $J = 4$ Hz; Ph); 108.5 (C-3); 59.3 (CH); 55.8 (OCH₃); 54.2 (2C, Pip); 45.8 (NCH₃); 32.1(2C, Pip); 15.7 (CH₃). HPLC-MS (ES⁺): CH₃CN/H₂O 15:85 – 95:5, gt: 5 min; rt = 4.52; [M+H]⁺ = 420.

3-(3-fluorophenyl)-5-(1-(piperidin-4-yl)-1*H*-pyrazol-4-yl)-1*H*-pyrrolo[2,3-*b*]pyridine (8). From 3-(3-fluorophenyl)-5-(1-(piperidin-4-yl)-1*H*-pyrazol-4-yl)-1-tosyl-1*H*-pyrrolo [2,3-*b*]pyridine (**24**) (48 mg, 0.09 mmol) and 0.4 M NaOH in MeOH (20 mL). Time of reaction: 70 min. Yield: (28 mg, 89%). Mp: 146.8 – 147.6 °C. ¹H NMR: DMSO-d₆ (400 MHz) δ : 12.02 (bs, 1H, NH); 8.56 (d, 1H, 6-H); 8.41 (d, 1H, 4-H); 8.36 (s, 1H, Ind); 7.99 (s, 1H, Ind); 7.96 (s, 1H, 2-H); 7.65 (d, 1H, Ar); 7.60 – 7.57 (m, 1H, Ar); 7.51 – 7.45 (m, 1H, Ar); 7.09 – 7.04 (m, 1H, Ar); 4.24 – 4.16 (m, 1H, CH); 3.07 – 3.03 (m, 2H, Pip); 2.63 – 2.56 (m, 2H, Pip); 2.02 – 1.98 (m, 2H, Pip); 1.88 – 1.78 (m, 2H, Pip). ¹³C NMR: DMSO-d₆ (100 MHz) δ : 162.8 (d, $J = 244$ Hz; Ph); 147.9 (C-7a); 141.0 (C-6); 137.6 (d, $J = 8$ Hz; Ph); 135.6 (Ind); 130.7 (d, $J = 9$ Hz; Ph); 125.2 (C-2); 124.7 (Ind); 123.3 (C-4); 122.2 (d, $J = 2$ Hz; Ph); 121.8 (C-5); 119.7 (Ind); 117.1 (C-3a); 113.0 (d, $J = 3$ Hz; C-3); 112.5 (d, $J = 22$ Hz; Ph); 112.1 (d, $J = 21$ Hz; C-3); 59.4 (CH); 45.2 (2C, Pip); 33.7 (2C, Pip). HPLC-MS (ES⁺): CH₃CN/H₂O 20:80 – 95:5, gt: 5 min; rt = 1.68; [M+H]⁺ = 362.

3-(3-methoxyphenyl)-5-(1-(piperidin-4-yl)-1*H*-pyrazol-4-yl)-1*H*-pyrrolo[2,3-*b*]pyridine (9). From 3-(3-methoxyphenyl)-5-(1-(piperidin-4-yl)-1*H*-pyrazol-4-yl)-1-tosyl-1*H*-pyrrolo[2,3-*b*]pyridine (**25**) (50 mg, 0.09 mmol) and 0.4 M NaOH in MeOH (20 mL). Time of reaction: 70 min. Yield: (25 mg, 74%). Mp: 246.8 – 247.3 °C. ¹H NMR: DMSO-d₆ (400 MHz) δ : 11.91 (bs, 1H, NH); 8.55 (d, 1H, 6-H); 8.40 - 8.34 (m, 2H, 4-H, Ind); 7.97 (m, 1H, 2-H); 7.87 (s, 1H, Ind); 7.38 – 7.36 (m, 2H, Ar); 7.26 (s, 1H, Ar); 6.86 – 6.83 (m, 1H, Ar); 4.23 – 4.16 (m, 1H, CH); 3.89 (s, 3H, CH₃); 3.15 – 3.03 (m, 2H, Pip); 2.63 – 2.60 (m, 2H, Pip); 2.12 – 1.96 (m, 4H, Pip). ¹³C NMR: DMSO-d₆ (100 MHz) δ : 159.9 (Ph); 147.9 (C-7a); 141.2 (C-6); 136.6 (Ph); 135.7 (Ind); 130.2 (Ph); 124.9 (C-2); 124.6 (Ind); 123.6 (Ph); 121.8 (C-4); 119.9 (C-5); 119.1 (Ind); 117.6 (C-3a); 114.5 (Ph); 111.9 (Ph); 111.6 (C-3); 59.4 (CH); 55.3 (OCH₃); 45.1 (2C, Pip); 33.6 (2C, Pip). HPLC-MS (ES⁺): CH₃CN/H₂O 20:80 – 95:5, gt: 5 min; rt = 4.03; [M+H]⁺ = 374.

3-(3,5-dichlorophenyl)-5-(1-(piperidin-4-yl)-1*H*-pyrazol-4-yl)-1*H*-pyrrolo[2,3-b]pyridine (10). From 3-(3,5-dichlorophenyl)-5-(1-(piperidin-4-yl)-1*H*-pyrazol-4-yl)-1-tosyl-1*H*-pyrrolo[2,3-b]pyridine (**26**) (80 mg, 0.14 mmol) and 0.4 M NaOH in MeOH (20 mL). Time of reaction: 4h. Yield: (31 mg, 71%). Mp: 271.1 - 271.9 °C. ¹H NMR: DMSO-d₆ (400 MHz) δ : 12.08 (bs, 1H, NH); 8.56 (d, 1H, 6-H); 8.36 (d, 1H, 4-H); 8.34 (s, 1H, Ind); 8.07 (s, 1H, 2-H); 7.99 (s, 1H, Ind); 7.78 (d, 2H, Ar); 7.44 (s, 1H, Ar); 4.23 – 4.17 (m, 1H, CH); 3.05 (d, 2H, Pip); 2.59 (t, 2H, Pip); 1.99 (d, 2H, Pip); 1.84 – 1.80 (m, 2H, Pip). ¹³C NMR: DMSO-d₆ (100 MHz) δ : 148.0 (C-7a); 141.3 (C-6); 138.9 (Ph); 135.7 (Ind); 134.6 (2C, Ph); 126.5 (C-2); 124.9 (Ph); 124.7 (Ind); 124.4 (2C, Ph); 123.2 (C-4); 122.0 (C-5); 119.6 (Ind); 116.9 (C-3a); 111.5 (C-3); 59.4 (CH); 45.2 (2C, Pip); 33.7 (2C, Pip). HPLC-MS (ES⁺): CH₃CN/H₂O 15:85 – 95:5, gt: 5 min; rt =4.69; [M+H]⁺ = 412 / 414.

3-(2-fluoro-5-methoxy-4-methylphenyl)-5-(1-(piperidin-4-yl)-1*H*-pyrazol-4-yl)-1*H*-pyrrolo[2,3-b]pyridine (11). From 3-(2-fluoro-5-methoxy-4-methylphenyl)-5-(1-(piperidin-4-yl)-1*H*-pyrazol-4-yl)-1-tosyl-1*H*-pyrrolo[2,3-b]pyridine (**27**) (70 mg, 0.12 mmol) and 0.4 M NaOH in MeOH (20 mL). Time of reaction: 4h. Yield: (37 mg, 75%). Mp: 247.9 - 248.5 °C. ¹H NMR: DMSO-d₆ (400 MHz) δ : 11.96 (bs, 1H, NH); 8.56 (d, 1H, 6-H); 8.29 (s, 1H, Ind); 8.17 (t, 1H, 4-H); 7.92 (s, 1H, Ind); 7.72 (s, 1H, 2-H); 7.15 – 7.13 (m, 2H, Ar); 4.22 – 4.14 (m, 1H, CH); 3.87 (s, 3H, OCH₃); 3.04 (d, 2H, Pip); 2.59 (t, 2H, Pip); 2.20 (s, 3H, CH₃); 1.98 (d, 2H, Pip); 1.86 – 1.76 (m, 2H, Pip). ¹³C NMR: DMSO-d₆ (100 MHz) δ : 153.6 (d, J = 1 Hz; Ph); 152.8 (d, J = 235 Hz; Ph); 147.4 (C-7a); 141.0 (C-6); 135.4 (Ind); 125.8 (d, J = 4 Hz; C-2); 125.1 (d, J = 8 Hz; Ph); 124.5 (Ind); 123.6 (d, J = 4 Hz; C-4); 121.4 (C-5); 119.7 (Ind); 119.6 (d, J = 15 Hz, Ph); 118.0 (C-3a); 117.6 (d, J = 14 Hz; Ph); 111.2 (d, J = 4 Hz; Ph); 108.6 (C-3); 59.3 (CH); 55.8 (OCH₃); 45.2 (2C, Pip); 33.6 (2C, Pip); 15.7 (CH₃). HPLC-MS (ES⁺): CH₃CN/H₂O 15:85 – 95:5, gt: 5 min; rt =4.49; [M+H]⁺ = 406.

Biological studies

Inhibition of Human DYRK1A kinase

The ADP-Glo™+ DYRK1A Kinase Enzyme System (no. catalog VA7425) from Promega was used to screen compounds for activity against DYRK1A. ATP and other reagents were purchased from Sigma-Aldrich (St. Louis, MO). The assays were performed in a buffer solution using 96-well plates. The compound to be tested (5 μ L, 40 μ M dissolved in 4% DMSO) was added to each well followed by ATP (5 μ L, final concentration in the well 10 μ M), DYRKtide (5 μ L, 4 μ g/well) and the enzyme (5 μ L, 25 ng/well). It was then allowed to incubate for 60 min at room temperature and ADP-Glo™ reagent (20 μ L) was added allowing it to incubate again for 40 min at room temperature. After the incubation, the kinase detection agent (40 μ L) was added and allowed to incubate for 30 min at room temperature. Finally, the luminescence was recorded using a FLUOstar Optima (BMG Labtechnologies GmbH, Offenburg, Germany) multimode reader. The inhibition activities were calculated based on the maximum activity measured in the absence of an inhibitor. Experiments were performed in triplicate.

Cell culture

The mouse microglial BV2 cell line was propagated using DMEM, 10% FBS, 1% streptomycin-penicillin, under humidified 5% CO₂ and 95% air. On attaining semiconfluence, cells were treated with 400 ng/ml of LPS for 24 h. Some cultures were pretreated for 1 h with the different compounds at several concentrations ranging from 1 μ M to 10 μ M. After treatment, cultures were processed for cell viability and nitrite production. Experiments were performed in triplicate.

Cell Viability

Cell viability was determined using the MTT assay, which measures the reduction of 3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) to formazan crystals. Briefly, the MTT solution (2 mg/ml) was added to each well and incubated at 37 °C for 2 hours. After removing the culture medium, 100 μ L of dimethyl sulfoxide was added to each well to dissolve the formazan. The optical density was measured at 532 nm using a microplate reader. The absorbance of the control group was considered as 100% of the cell viability.

Nitrite Determination


The assessment of NO production involved quantifying nitrite levels, one of the end products of NO oxidation, by a procedure based on the diazotization of nitrite by sulfanilic acid (Griess reaction). Upon reaching semiconfluence, cells were exposed to 400 ng/ml of LPS for 24 hours. Before this, certain cultures were pretreated with various compounds at concentrations ranging from 0.5 μM to 10 μM . Following a 24-hour incubation period, 50 μL aliquots of the samples were combined with 50 μL of Griess reagent in 96-well plates, and the mixture was allowed to incubate at room temperature for 10 minutes. The absorbance of the resulting mixture was then measured at 520 nm using a microplate reader.

Measurement of the Antioxidant Effect of the Compounds


The antioxidant activity of the newly synthesized compounds was assessed using the oxygen radical absorbance capacity (ORAC) in vitro assay (Ou *et al.*). The FLUOstar Optima plate reader (BMG Labtech GmbH, Offenburg, Germany) was employed, with excitation at 485 nm and emission at 520 nm. 2,2'-Azobis-(amidinopropane) dihydrochloride (AAPH), (\pm)-6-hydroxy-2,5,7,8-tetramethylchromane-2-carboxylic acid (Trolox), and fluorescein (FL) were procured from Sigma-Aldrich. The assay was conducted in 75 mM phosphate buffer (pH = 7.4) with a final reaction volume of 200 μL . Each well of black 96-well plates contained 25 μL of the antioxidant sample and 150 μL of fluorescein (10 nM). After pre-incubation at 37°C for 30 minutes, 25 μL of a 240 mM AAPH solution was rapidly added using a multi-channel pipette. Fluorescence measurements were taken every 90 seconds for 90 minutes, with the plate agitated before each reading. The compounds were tested at four concentrations (10 - 1 μM). A blank containing FL and AAPH in phosphate buffer, as well as four concentrations of Trolox (10 - 1 μM) served as controls. All reactions were performed in duplicate, with at least three independent tests per compound. Data were exported for analysis, plotting absorbance versus time. The area under the fluorescence decay curve (AUC) was calculated for each sample. ORAC values were derived from the AUC values and expressed as Trolox equivalents.⁷⁹

Author information

Corresponding Authors


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

Email: nuria.campillo@csic.es


Simón Rodríguez Santana – Institute for Research in Technology (IIT), ICAI Engineering School, Universidad Pontificia de Comillas;  orcid.org/0000-0003-3760-0520;

Email: srsantana@icai.comillas.edu


Author

Eduardo González – Departamento de Inteligencia Artificial-Universidad Politecnica de Madrid (UPM)), 28660 Madrid, Spain; work initiated during a scholarship in the Institute of Mathematical Sciences (ICMAT-CSIC);  orcid.org/0009-0005-4717-5371;


Email: e.garcia@alumnos.upm.es

Pablo Varas Pardo – Altenea Biotech & Instituto de Ciencias Matemáticas (ICMAT-CSIC), 28049 Madrid, Spain;  orcid.org/0009-0006-1115-4824;


Email: pablo.varas@icmat.es


Pedro González-Naranjo – Instituto de Química Médica (IQM-CSIC);  orcid.org/0000-0001-9726-689X;

Email: pedrojgn@iqm.csic.es

Eugenia Ulzurrún – Centro de Investigaciones Biológicas Margarita Salas (CIB Margarita Salas-CSIC), 28040 Madrid, Spain & Instituto de Ciencias Matemáticas (ICMAT-CSIC), 28049 Madrid, Spain;  orcid.org/0009-0008-7665-2167;

Email: mariaeugenia.ulzurrun@cib.csic.es

Juan Antonio Páez – Instituto de Química Médica (IQM-CSIC);  orcid.org/0000-0003-3874-688X; Email: jpaez@iqm.csic.es

David Ríos Insua – Instituto de Ciencias Matemáticas (ICMAT-CSIC);  orcid.org/0000-0002-5748-9658; Email: david.rios@icmat.es

Concepción Pérez – Instituto de Química Médica (IQM-CSIC);  orcid.org/0000-0001-7183-4035; Email: iqmm305@iqm.csic.es

Author contributions

The manuscript was written with contributions from all authors. All authors have approved the final version of the manuscript

Acknowledgement

This work was financially supported by Projects: "IND2023/BMD-27452" financed by CAM, TED2021-129970B-C21" financed by MCIN/AEI and European Union "NextGenerationEU"/PRTR and Project "SNEO-2022207" (NEOTEC) financed by CDTI (MCIN) and European Union "NextGenerationEU"/PRT. The author thanks Mats Dahlgren for version one of `achemso`, and Donald Arseneau for the code taken from `cite` to move citations after punctuation. Many users have provided feedback on the class, which is reflected in all of the different demonstrations shown in this document.

Abbreviations used

AI, Artificial Intelligence; AD, Alzheimer disease; CADD, Computer-aided drug design; DDR1, discoidin domain receptor 1 kinase; DYRK1A, Dual-Specificity tyrosine Phosphorylation-Regulated Kinase 1 A; EC₅₀, half maximal effective concentration; EVS, Explained Variance Score; FDA, Food and Drug Administration; GNNs, Graph Neural Networks; GPs, Gaussian Processes; HGG, hierarchical graph generation; HPLC, High performance liquid chromatography; IC₅₀, half-maximal inhibitory concentration; Ki, inhibitory constant; Kd, dissociation constant; KNN, K-nearest Neighbors; MAE, Mean absolute error; M3C; MLP, Multi-layer Perceptron; QSAR, Quantitative Structure-Activity Relationship; PDB, Protein Database; QED, Quantitative Estimate of Druglikeness; R², coefficient of determination or R-squared; RFs, Random Forest; RMSE, root

mean square error; RGA, Reinforced Genetic Algorithm, SBVS, Structure-Based Virtual Screening; SVR, Support Vector Regressor; VS, Virtual Screening; SARS CoV 2, Severe acute respiratory syndrome coronavirus 2; XGBoost, Extreme Gradient Boosting.

Supporting Information Available

Table S1. The top 50 de novo molecules. Sorted by docking score value.

Table S2. Virtual chemical library of derivatives of compound **1**.

Table S3: Table of ¹H-RMN of compounds **1, 2, 5 – 14, 19 – 22, 24 – 31**.

Table S4: Table of ¹³C-RMN of compounds **1, 2, 5 – 14, 19 – 22, 24 – 31**.

Figures S1 - S16: ¹H and ¹³C NMR spectra of compounds **1, 5 – 11**.

Figures S17 – S18: Effect of compounds **1, 5 – 11** in the nitrite production of BV2 Cells.

Elemental Analysis Data of compounds **1, 2, 5 – 14, 19 – 22, 24 – 31**.

References

- (1) Abbassi, R.; Johns, T. G.; Kassiou, M.; Munoz, L. DYRK1A in neurodegeneration and cancer: Molecular basis and clinical implications. *Pharmacol. Therapeut.* **2015**, *151*, 87–98.
- (2) Liu, T.; Wang, Y.; Wang, J.; Ren, C.; Chen, H.; Zhang, J. DYRK1A inhibitors for disease therapy: Current status and perspectives. *Eur. J. Med. Chem.* **2022**, *229*, 114062.
- (3) 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. *Curr. Med. Chem.* **2023**, *30*, 669–688.
- (4) Yoon, H. R.; Balupuri, A.; Choi, K.-E.; Kang, N. S. Small molecule inhibitors of

- DYRK1A identified by computational and experimental approaches. *Int. J. Mol. Sci.* **2020**, *21*, 6826.
- (5) Jarhad, D. B.; Mashelkar, K. K.; Kim, H.-R.; Noh, M.; Jeong, L. S. Dual-specificity tyrosine phosphorylation-regulated kinase 1A (DYRK1A) inhibitors as potential therapeutics. *J. Med. Chem.* **2018**, *61*, 9791–9810.
- (6) García-Cerro, S.; Rueda, N.; Vidal, V.; Lantigua, S.; Martínez-Cué, C. Normalizing the gene dosage of Dyrk1A in a mouse model of Down syndrome rescues several Alzheimer's disease phenotypes. *Neurobiol. Dis.* **2017**, *106*, 76–88.
- (7) Vanhaelen, Q.; Lin, Y.-C.; Zhavoronkov, A. The advent of generative chemistry. *ACS Med. Chem. Lett.* **2020**, *11*, 1496–1505.
- (8) Gangwal, A.; Ansari, A.; Ahmad, I.; Azad, A. K.; Kumarasamy, V.; Subramanian, V.; Wong, L. S. Generative artificial intelligence in drug discovery: basic framework, recent advances, challenges, and opportunities. *Front. pharmacol.* **2024**, *15*, 1331062.
- (9) Yan, C.; Grabowska, M. E.; Dickson, A. L.; Li, B.; Wen, Z.; Roden, D. M.; Michael Stein, C.; Embí, P. J.; Peterson, J. F.; Feng, Q.; others Leveraging generative AI to prioritize drug repurposing candidates for Alzheimer's disease with real-world clinical validation. *npj Digit. Med.* **2024**, *7*, 46.
- (10) Ballarotto, M.; Willems, S.; Stiller, T.; Nawa, F.; Marschner, J. A.; Grisoni, F.; Merk, D. De novo design of nurr1 agonists via fragment-augmented generative deep learning in low-data regime. *J. Med. Chem.* **2023**, *66*, 8170–8177.
- (11) Prado-Romero, D. L.; Gómez-García, A.; Cedillo-González, R.; Villegas-Quintero, H.; Avellaneda-Tamayo, J. F.; López-López, E.; Saldívar-González, F. I.; Chávez-Hernández, A. L.; Medina-Franco, J. L. Consensus docking aid to model the activity of an inhibitor of DNA methyltransferase 1 inspired by de novo design. *Front. drug discov.* **2023**, *3*, 1261094.

- (12) Wang, X.; Gao, C.; Han, P.; Li, X.; Chen, W.; Rodríguez Patón, A.; Wang, S.; Zheng, P. PETrans: De novo drug design with protein-specific encoding based on transfer learning. *Int. J. Mol. Sci.* **2023**, *24*, 1146.
- (13) Hasselgren, C.; Oprea, T. I. Artificial intelligence for drug discovery: Are we there yet? *Annu. Rev. Pharmacol. Toxicol.* **2024**, *64*, 527–550.
- (14) Zhou, S.; Johnson, R. Pharmaceutical probability of success. *Alacrita: London, UK* **2019**,
- (15) Polishchuk, P. G.; Madzhidov, T. I.; Varnek, A. Estimation of the size of drug-like chemical space based on GDB-17 data. *J. Comput. -Aided Mol.* **2013**, *27*, 675–679.
- (16) Sharma, T.; Padhy, I.; Sahoo, C. R. *Drug Repurposing and Computational Drug Discovery*; Apple Academic Press, 2023; pp 27–58.
- (17) Khan, J.; Singla, R. K. Bioinformatics Tools for Pharmaceutical Drug Product Development. *Indo Global J. Pharm. Sci.* **2022**, *12*, 281–294.
- (18) Paul, D.; Sanap, G.; Shenoy, S.; Kalyane, D.; Kalia, K.; Tekade, R. K. Artificial intelligence in drug discovery and development. *Drug Discov. Today* **2021**, *26*, 80.
- (19) Gangwal, A.; Lavecchia, A. Unlocking the potential of generative AI in drug discovery. *Drug Dis. Today* **2024**, 103992.
- (20) Tiwari, P. C.; Pal, R.; Chaudhary, M. J.; Nath, R. Artificial intelligence revolutionizing drug development: Exploring opportunities and challenges. *Drug Dev. Res.* **2023**, *84*, 1652–1663.
- (21) Gallego, V.; Naveiro, R.; Roca, C.; Rios Insua, D.; Campillo, N. E. AI in drug development: a multidisciplinary perspective. *Mol. Div.* **2021**, *25*, 1461–1479.
- (22) Schwaller, P.; Laino, T.; Gaudin, T.; Bolgar, P.; Hunter, C. A.; Bekas, C.; Lee, A. A. Molecular transformer: a model for uncertainty-calibrated chemical reaction prediction. *ACS Central Sci.* **2019**, *5*, 1572–1583.

- (23) 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. *J. Chem. Inf. Model.* **2024**, *64*, 9–17.
- (24) Leelananda, S. P.; Lindert, S. Computational methods in drug discovery. *Beilstein J. Org. Chem.* **2016**, *12*, 2694–2718.
- (25) Zhao, H. The science and art of structure-based virtual screening. *ACS Med. Chem. Lett.* **2024**, *15*, 436–440.
- (26) Hartenfeller, M.; Schneider, G. De novo drug design. *Methods Mol. Biol.* **2010**, *672*, 299–323.
- (27) Hartenfeller, M.; Schneider, G. Enabling future drug discovery by de novo design. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2011**, *1*.
- (28) Gómez-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.
- (29) Pang, C.; Qiao, J.; Zeng, X.; Zou, Q.; Wei, L. Deep Generative Models in De Novo Drug Molecule Generation. *J. Chem. Inf. Model.* **2023**,
- (30) Li, Y.; Pei, J.; Lai, L. Structure-based de novo drug design using 3D deep generative models. *Chem. Sci.* **2021**, *12*, 13664 – 13675.
- (31) 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.; others Deep learning enables rapid identification of potent DDR1 kinase inhibitors. *Nat. Biotech.* **2019**, *37*, 1038–1040.
- (32) Méndez-Lucio, O.; Baillif, B.; Clevert, D.-A.; Rouquié, D.; Wichard, J. De novo generation of hit-like molecules from gene expression signatures using artificial intelligence. *Nat. Commun.* **2020**, *11*, 10.

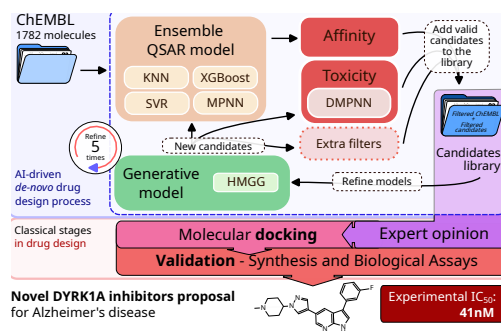
- (33) Maziarka, Ł.; Danel, T.; Mucha, S.; Rataj, K.; Tabor, J.; Jastrzębski, S. Molecule attention transformer. *arXiv preprint arXiv:2002.08264* **2020**,
- (34) Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, NY, USA, 2016; pp 785–794.
- (35) Wang, Y.; Wang, J.; Cao, Z.; Barati Farimani, A. Molecular contrastive learning of representations via graph neural networks. *Nat. Mach. Intell.* **2022**, *4*, 279–287.
- (36) Hu, W.; Liu, B.; Gomes, J.; Zitnik, M.; Liang, P.; Pande, V.; Leskovec, J. Strategies for Pre-training Graph Neural Networks. International Conference on Learning Representations. 2020.
- (37) Jin, W.; Barzilay, R.; Jaakkola, T. Hierarchical generation of molecular graphs using structural motifs. International conference on machine learning. 2020; pp 4839–4848.
- (38) Falke, H.; Chaikuad, A.; Becker, A.; Loaëc, N.; Lozach, O.; Abu Jhaisha, S.; Becker, W.; Jones, P. G.; Preu, L.; Baumann, K.; others 10-Iodo-11 H-indolo [3, 2-c] quinoline-6-carboxylic acids are selective inhibitors of DYRK1A. *J. Med. Chem.* **2015**, *58*, 3131–3143.
- (39) 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. *J. Med. Chem.* **2018**, *61*, 7560–7572.
- (40) Derewenda, Z. S.; Hawro, I.; Derewenda, U. C—H···O hydrogen bonds in kinase-inhibitor interfaces. *IUBMB life* **2020**, *72*, 1233–1242.
- (41) Xing, L.; Klug-Mcleod, J.; Rai, B.; Lunney, E. A. Kinase hinge binding scaffolds and their hydrogen bond patterns. *Bioorg. Med. Chem* **2015**, *23*, 6520–6527.
- (42) Pierce, A. C.; Sandretto, K. L.; Bemis, G. W. Kinase inhibitors and the case for

- CH... O hydrogen bonds in protein–ligand binding. *Proteins: Struct. Funct. Genet.* **2002**, *49*, 567–576.
- (43) Klug, D. M.; Mavrogiannaki, E. M.; Forbes, K. C.; Silva, L.; Diaz-Gonzalez, R.; Pérez-Moreno, G.; Ceballos-Pérez, G.; Garcia-Hernández, R.; Bosch-Navarrete, C.; Cordon-Obras, C.; others Lead optimization of 3, 5-Disubstituted-7-Azaindoles for the treatment of human African trypanosomiasis. *J. Med. Chem.* **2021**, *64*, 9404–9430.
- (44) Park, E.; Lee, S. J.; Moon, H.; Park, J.; Jeon, H.; Hwang, J. S.; Hwang, H.; Hong, K. B.; Han, S.-H.; Choi, S.; others Discovery and biological evaluation of N-methyl-pyrrolo [2, 3-b] pyridine-5-carboxamide derivatives as JAK1-selective inhibitors. *J. Med. Chem.* **2021**, *64*, 958–979.
- (45) Yang, B.; Wu, Q.; Huan, X.; Wang, Y.; Sun, Y.; Yang, Y.; Liu, T.; Wang, X.; Chen, L.; Xiong, B.; others Discovery of a series of 1H-pyrrolo [2, 3-b] pyridine compounds as potent TNIK inhibitors. *Bioorg. Med. Chem Letters* **2021**, *33*, 127749.
- (46) Bearss, D.; Kauwe, J.; Mollard, A. Pyrrole[2,3-b]pyridine derivatives as tyro3 inhibitors. *WO2023049270* **2023**,
- (47) Dorsch, D.; Hoelzemann, G.; Eggenweiler, H.-M.; Czodrowski, P. Preparation of 3-cyanoaryl-1H-pyrrolo[2,3-b]pyridine derivatives useful as TBK1 and IKK $\hat{\mu}$ inhibitors. *WO2013075785* **2023**,
- (48) Li, J.; Wang, Z.; Xu, S. Preparation of pyrrolo[2,3-b]pyridines or pyrrolo[2,3-b]pyrazines as HPK1 inhibitor and the use thereof. *WO2019238067* **2019**,
- (49) Mou, J.; Zhou, T.; Xu, W.; Guo, J.; Qiang, G.; Ling, X.; Pei, D. Preparation of pyrrolopyridine compounds as antitumor drug. *CN115124528* **2022**,
- (50) Vankayalapati, H.; Yerramreddy, V.; Ganipisetty, V. B.; Talluri, S.; Appalaneni, R. P. Preparation of substituted 1H-pyrrolo[2,3-b]pyridine and 1H-pyrazolo[3,4-b]pyridine derivatives as salt inducible kinase 2 (SIK2) inhibitors. *CN115124528* **2014**,

- (51) Zeng, S.; Huang, W.; Wang, Z.; Pan, Y.; others Preparation of pyrrolopyrazine derivative and application as HPK1 or LCK kinase modulator. *CN116162087* **2023**,
- (52) Bain, J.; Plater, L.; Elliott, M.; Shpiro, N.; Hastie, C. J.; Mclauchlan, H.; Klevernic, I.; Arthur, J. S. C.; Alessi, D. R.; Cohen, P. The selectivity of protein kinase inhibitors: a further update. *Biochem. J.* **2007**, *408*, 297–315.
- (53) Mendez, D. *et al.* ChEMBL: towards direct deposition of bioassay data. *Nucleic Acids Res.* **2018**, *47*, D930–D940.
- (54) Huang, R.; Xia, M.; Nguyen, D.-T.; Zhao, T.; Sakamuru, S.; Zhao, J.; Shahane, S. A.; Rossoshek, A.; Simeonov, A. Tox21Challenge to build predictive models of nuclear receptor and stress response pathways as mediated by exposure to environmental chemicals and drugs. *Front. environ. sci.* **2016**, *3*, 85.
- (55) Box, G. E.; Cox, D. R. An analysis of transformations. *J. R. Stat. Soc. Ser. B Methodol.* **1964**, *26*, 211–243.
- (56) Sanchez-Lengeling, B.; Reif, E.; Pearce, A.; Wiltschko, A. B. A gentle introduction to graph neural networks. *Distill* **2021**, *6*, e33.
- (57) Morgan, H. L. The generation of a unique machine description for chemical structures—a technique developed at chemical abstracts service. *J. Chem. Doc* **1965**, *5*, 107–113.
- (58) Landrum, G. Rdkit documentation. *Release* **2013**, *1*, 4.
- (59) 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.
- (60) Fu, T.; Gao, W.; Coley, C.; Sun, J. Reinforced genetic algorithm for structure-based drug design. *Adv. Neural. Inf. Process. Syst.* **2022**, *35*, 12325–12338.

- (61) Schneuing, A.; Du, Y.; Harris, C.; Jamasb, A.; Igashov, I.; Du, W.; Blundell, T.; Lió, P.; Gomes, C.; Welling, M. Structure-based drug design with equivariant diffusion models. *arXiv preprint arXiv:2210.13695* **2022**,
- (62) Schrödinger, LLC Software Release 2021-1 distribution. 2021; New York, NY.
- (63) Schrödinger, LLC Schrödinger Release 2021-1: LigPrep. 2021; New York, NY.
- (64) Schrödinger, LLC Schrödinger Release 2021-1: Maestro. 2021; New York, NY.
- (65) Madhavi Sastry, G.; Adzhigirey, M.; Day, T.; Annabhimoju, R.; Sherman, W. Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *J. Comput. -Aided Mol.* **2013**, *27*, 221–234.
- (66) 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. Schrödinger, LLC: New York, NY, 2021.
- (67) 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: Struct. Funct. Genet.* **2004**, *55*, 351–367.
- (68) Jacobson, M. P.; Friesner, R. A.; Xiang, Z.; Honig, B. On the role of the crystal environment in determining protein side-chain conformations. *J. Mol. Biol.* **2002**, *320*, 597–608.
- (69) Schrödinger, LLC Schrödinger Release 2021-1: Prime. 2021; New York, NY.
- (70) 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: p K a and Protonation State Prediction through Machine Learning. *J. Chem. Theory Comput.* **2023**, *19*, 2380–2388.
- (71) Schrödinger, LLC Schrödinger Release 2021-1: Epik. 2021; New York, NY.

- (72) Olsson Mats, H.; Søndergaard Chresten, R.; Rostkowski Michal, J. J. H. PROPKA3: consistent treatment of internal and surface residues in empirical p K a predictions. *J. Chem. Theory Comput.* **2011**, *7*, 525–537.
- (73) 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. *J. Chem. Theory Comput.* **2021**, *17*, 7106–7119.
- (74) 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. *J. Med. Chem.* **2006**, *49*, 6177–6196.
- (75) 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. *J. Med. Chem.* **2004**, *47*, 1750–1759.
- (76) 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. *J. Med. Chem.* **2004**, *47*, 1739–1749.
- (77) Schrödinger, LLC Schrödinger Release 2021-1: Glide. 2021; New York, NY.
- (78) Goodfellow, V. S.; Loweth, C. J.; Ravula, S. B.; Wiemann, T.; Nguyen, T.; Xu, Y.; Todd, D. E.; Sheppard, D.; Pollack, S.; Polesskaya, O.; others Discovery, synthesis, and characterization of an orally bioavailable, brain penetrant inhibitor of mixed lineage kinase 3. *J. Med. Chem.* **2013**, *56*, 8032–8048.
- (79) Ou, B.; Hampsch-Woodill, M.; Prior, R. L. Development and validation of an improved oxygen radical absorbance capacity assay using fluorescein as the fluorescent probe. *Journal of agricultural and food chemistry* **2001**, *49*, 4619–4626.



Graphical abstract: Integrated workflow for developing DYRK1A inhibitors combining AI-driven de-novo design with experimental validation.