chemistry proceedings

MDPI

Proceeding Paper

Integrated Computational Approach to Rational Drug Design Targeting SIK2/3: From Theory to Practice

Eid Youssef Eid Rashed [1,*] and Alisa Gorislav [2,*]

[1] Department of Pharmacy, Perm State University, 614068 Perm Krai, Russia
[2] Alferov Federal State Budgetary Institution of Higher Education and Science Saint Petersburg, National Research Academic University of the Russian Academy of Sciences, 119991 Moscow, Russia
* Correspondence: edjosef96@gmail.com (E.Y.E.R.); alisagorislav@gmail.com (A.G.)
† Presented at The 28th International Electronic Conference on Synthetic Organic Chemistry (ECSOC 2024), 15–30 November 2024; Available online: https://sciforum.net/event/ecsoc-28.

Abstract: This study presents a comprehensive approach to designing and optimizing small molecule inhibitors targeting Salt-Inducible Kinases 2 and 3 (SIK2 and SIK3), crucial regulators of cellular signaling pathways implicated in various diseases, including cancer, inflammation, and metabolic disorders. By integrating advanced computational methods and expert-driven chemical synthesis, we generated a diverse library of potential inhibitors and meticulously evaluated their pharmacological properties and binding affinities to SIK2 (Appendix C). Through a rigorous analysis of generated data and molecular docking simulations, we successfully identified lead compounds with promising therapeutic potential. Subsequently, employing iterative chemical modifications guided by human expertise, we further optimized these leads, enhancing their efficacy and specificity. Additionally, employing molecular dynamics simulations provided valuable mechanistic insights into the dynamic behavior of optimized compounds within the complex biological environment, elucidating their potential as effective inhibitors of SIK2 activity. Our findings underscore the efficacy and significance of an integrated computational and experimental approach in the development of novel therapeutics targeting SIK2 and SIK3 (Appendix D). By bridging computational predictions with experimental validation, this approach not only accelerates the drug discovery process but also increases the likelihood of identifying clinically relevant compounds. Furthermore, the insights gained from this study lay a solid foundation for future preclinical and clinical investigations, paving the way for the development of effective treatments for diseases associated with dysregulated SIK2 and SIK3 signaling pathways.

Keywords: de novo drug design; fragment-based drug design; molecular docking; covalent docking; molecular dynamics; chemical language models; computational chemistry; SIK2; SIK3; AI-driven drug design

Citation: Rashed, E.Y.E.; Gorislav, A. Proceeding Paper: Integrated Computational Approach to Rational Drug Design Targeting SIK2/3: From Theory to Practice. *Chem. Proc.* **2024**, *6*, x. https://doi.org/10.3390/xxxxx

Academic Editor(s): Name

Published: 15 November 2024

Copyright: © 2024 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Salt-Inducible Kinases 2 and 3 (SIK2 and SIK3) are key regulators of cellular metabolism and transcriptional regulation, implicated in various diseases such as cancer, inflammation, and metabolic disorders. Their pivotal roles make them attractive therapeutic targets, yet the development of selective and potent inhibitors for these kinases presents significant challenges due to their structural intricacies and biological functions.

In this study, we tackle these challenges by adopting a comprehensive computational approach, integrating multiple advanced methodologies, including de novo drug design, fragment-based drug design (FBDD), molecular docking, covalent docking, and molecular dynamics simulations [1]. By leveraging these cutting-edge techniques, we aimed to streamline the identification and optimization of novel small molecule inhibitors with strong binding affinities, high specificity, and favorable pharmacokinetic properties. This

approach not only accelerates the drug discovery process but also enhances the potential for developing effective therapeutics targeting SIK2 and SIK3, addressing diseases associated with dysregulated signalling pathways.

2. Methodology

As the reference molecule for AI-driven molecule generation, we selected "CNc1ncc2cc(-c3ccc(-c4nccc4F)cc3Cl)c(=O)n(C[C@H]3OCC@HCO3)c2n1" (CHEMBL5090394) because of its effective binding within the SIK2/3 active site. We curated a large dataset of potential lead compounds by retrieving over 71 million SMILES strings from the PubChem database (Appendix A). These compounds were first filtered using Lipinski's Rule of Five and the Ghose filter to ensure favorable drug-like properties. The filtering process reduced the dataset to 557,882 compounds. Further refinement was achieved by applying Tanimoto similarity filtering using various fingerprints, which narrowed the dataset to the top 250 candidates. These candidates exhibited high structural similarity to the reference molecule while maintaining diversity(Appendices B and C).

Similarly, we extracted 235,444 compounds from the Zinc database and subjected them to identical filtering criteria. This process reduced the dataset to 187,979 compounds, from which the top 250 candidates were selected based on their Tanimoto similarity scores relative to the reference molecule. This dual-dataset approach provided a robust and diverse starting point for subsequent molecular docking and fragment-based design experiments.

We compiled eight smaller files—four each for the PubChem and Zinc datasets—corresponding to the fingerprint types used. Each file contained the top three SMILES showing the highest similarity to the reference molecule within the pre-selected subset of 250 SMILES, along with the reference SMILE itself. Additionally, eight more files were generated, each containing four randomly selected SMILES from the respective 250 SMILE datasets, ensuring diversity in our analysis and a broader exploration of the chemical space. This study employed two advanced chemical language models: "Molecular Design with Beam Search" and "REINVENT 4" to identify and optimize potential inhibitors targeting Salt-Inducible Kinases 2 and 3 (SIK2 and SIK3).

Phase 1: De Novo Molecular Generation

In the initial phase, we utilized the "Molecular Design with Beam Search" model to generate SMILES strings analogous to a reference molecule for SIK2/3 inhibitors [1]. This process involved several fine-tuning strategies:

Single-Step Fine-Tuning: The model was fine-tuned using four SMILES.
Two-Step Fine-Tuning: adjusted the model with 250 SMILES closely related to the reference.
Random Samples: Variations of fine-tuning processes using randomly selected SMILES.

Phase 2: Fragment-Based Drug Design

In Phase 2, fragment-based drug design (FBDD) was applied using both the "Molecular Design with Beam Search" and "REINVENT 4" models [3]. This phase involved fragment-based generation and optimization of molecules by employing fragment files(Appendix C) like "Enamine_MiniFrag.txt", "Enamine_Single_Pharmacophore.txt", and "Asinex_BioDesign.txt". The process involved two steps:

Fragment-Based Fine-Tuning: The model was further refined with fragment files.
Lead Optimization: The lead was used as a basis for exploring optimization potential.

The goal was to design a novel molecular entity distinct from known SIK2/3 inhibitors, leading to the identification of a compound (Figure 1), "O=C(c1ccc2ccoc2c1)N1CCC(C(=O)N2CCCC2)CC1". This compound, generated via two-step fine-tuning without including the reference molecule, showed close but slightly lower descriptor values compared to the reference compound. The molecular docking

analysis was performed using the X-ray crystal structure of the MARK2-SIK2 chimera (PDB ID: 8TXY) as the receptor. Receptor and ligand preparation involved processing the 3D structures with AutoDock Tools, removing non-essential molecules, adding polar hydrogens, and correcting charges.

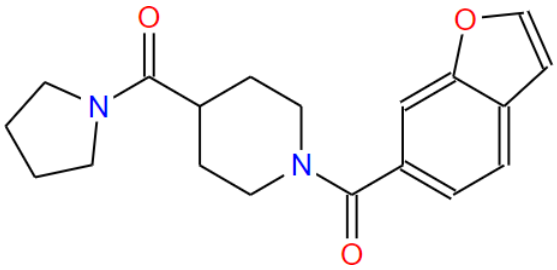

Figure 1. 2d representation of chosen lead.

Docking simulations employed both AutoDock 4 and AutoDock Vina, using a genetic algorithm and blind docking to explore potential binding pockets[2]. Key docking parameters included a grid size of 126 × 126 × 126 Å, energy range of 4, exhaustiveness of 300, and up to 20 binding modes. Following preliminary docking, the lead compound underwent iterative human-guided optimization to enhance its pharmacological profile. Structural modifications were introduced, such as adding an amino bridge between a six-membered ring and a carbonyl group, forming an amino carbonyl group. Additionally, the six-membered ring was modified into a benzene ring with repositioned nitrogen atoms, aiming to improve reactivity and structural integrity. The resulting initial SMILE for human optimization, "`O=C(CNc3cnc(C(=O) c2ccc1ccoc1c2)nc3)N4CCCC4`", led to the generation of 30 distinct optimized molecules. The optimization process involved dividing the molecule into five segments, color-coded for different functional groups (e.g., indole, carbonyl, pyrimidine) (Figure 2). Docking simulations validated the optimized molecules using AutoDock Vina. For covalent docking, we used Meeko (for ligand preparation) and AutoDock GPU. As an amino acid of interest, we selected Methionine 104. Fragmenstein was used to generate novel molecules based on the structure of the reference and lead compounds. For molecular dynamics, we used GROMACS 2024.1 with the AMBER force field for the receptor and OpenFF tools for ligand parameterization and SPC216 water model for 10 ns with a 2 fs time step. To analyze the result, we used the python library MDAnalysis[6].

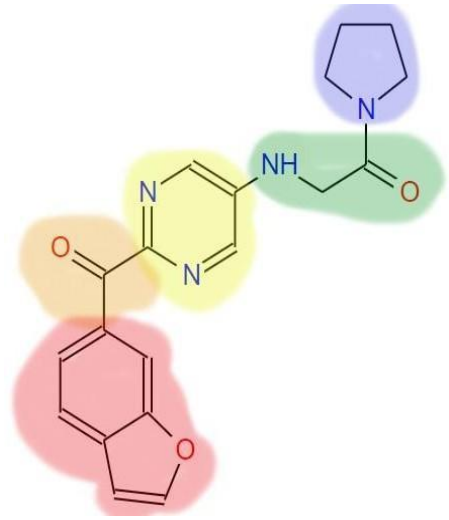

Figure 2. initial SMILE of human optimization.

Part 3: Experimental Results

Phase 1 results from the generative AI-driven approach for De Novo Drug Design, included two models. Model 1, Molecular Design with Beam Search, was run 42 times. Each run generated four files, including a text file with 15 SMILES, producing the best results with reasonable similarity and sufficient diversity. Model 2, REINVENT4, generated 628 SMILES strings without fine-tuning, followed by a two-step fine-tuning run. After transfer learning, 154 SMILES strings were generated. In Phase 2, Fragment-Based Drug Design also utilized generative AI, with Model 1, Molecular Design Using Beam Search, executed 16 times, each producing 15 SMILES strings, though with low similarity to the reference molecule. Model 2, REINVENT4 Implementation, focused on Lead SMILE exploration, generating 12 CSV files with a median output of 157 SMILES strings per run. These generated SMILES had high to reasonable similarity but lacked diversity (Appendix C).

A notable molecule, "`O=C(c1ccc2ccoc2c1)N1CCC (C(=O)N2CCCC2)CC1`," generated in Phase 1 using Model 1, was selected for further optimization. This molecule displayed simplicity yet enough similarity. Using AutoDock Vina, the reference molecule exhibited a minimum binding energy of −9.2 kcal/mol. The lead had a minimum binding energy of −8.2 kcal/mol, indicating the need for further optimization (Table 1). SMILE10 of the Human Optimization phase had the highest binding affinity, followed closely by other derivatives (Table 2).

Table 1. Reference and Lead docking results select covalent variants of SMILES 1 and SMILE 10 were subjected to this process. Despite the limited number of variants tested, significant enhancements in binding affinity were observed (Table 3).

SMILE	Mode	Affinity (kcal/mol)	Dist from (rmsd l.b.)
Reference	1	−9.2	0
	2	−9.2	1.264
Lead	1	−8.2	0
	2	−7.9	2.367

Table 2. Human Optimization variants.

SMILE Number	Affinity (kcal/mol)
SMILE 10	−8.788
SMILE 29	−8.786
SMILE 30	−8.763
SMILE 8	−8.719

Table 3. Binding Energies of Covalent Inhibitors.

SMILE	Warhead Variant	Variant	RMSD (kcal/mol)	Log (kcal/mol)
SMILE 1	Benzene sulfonyl Fluoride	1.4	−9.63	−15.15
SMILE 10		1.3	−9.27	−16.83
SMILE 1		1.1	−8.99	−15.78

The data shows that the binding affinities for the first two variants surpassed those of the reference. The Fragmenstein algorithm facilitated the creation of structurally unique molecules, which were docked against the target protein using AutoDock Vina, resulting in highly promising binding affinities (Table 4). As a result, SMILE 13 from Step 1 exhibited the lowest binding energy at −10.7 kcal/mol, outperforming the best results from the human optimization phase (−8.788 kcal/mol) which surpasses the reference's lowest binding energy of −9.2 kcal/mol (Table 1).

Table 4. Fragmenestein docking results.

STEP	SMILE	RMSD (kcal/mol)
Step 1	SMILE13	−10.7
	SMILE2	−9.4
Step 2	SMILE8	−9.3
	SMILE15	−9.2

The molecular dynamics (MD) simulations were carried out on selected ligands to evaluate their dynamic behavior within the target binding sites, representing various stages of optimization. The chosen ligands included Human Optimization Ligand 1 Pose 5, Ligand 2 Pose 18, and Ligand 10 Pose 2, Fragmenstein Ligand 2 Pose 12 Step 1, and Covalent Inhibitors Ligand 10 Variant 1.3 with a Benzenesulfonyl Fluoride warhead. These selections reflect a broad cross-section of optimization efforts, from human-driven refinements to algorithmic and covalent inhibitor design. The MD trajectory analysis provided insights into the stability and behavior of specific residues involved in ligand binding. Key residues analyzed included MET 104, ILE 119, GLU 120, ARG 102, and HOH 204:

Ligand 10 BSF Variant 1.3: This covalent inhibitor demonstrated consistent hydrogen bonding with critical residues, particularly MET 104 and GLU 120, throughout the simulation. The heatmap (Figure 4) visualizes these dynamics, showing stable interactions that reinforce the ligand's strong binding affinity.

Fragmenstein Ligand 2 Step 1: This ligand exhibited dynamic hydrogen bonding patterns, particularly with residues ARG 102 and ILE 119. Despite some fluctuation, the heatmap indicates overall stable interactions, suggesting the ligand maintains effective binding within the active site over the course of the simulation.

These results highlight the stability of both ligands in their respective binding pockets, suggesting strong potential as inhibitors targeting SIK2/ SIK3. The Root Mean Square Deviation (RMSD) analysis offers insight into the stability of ligands within the binding pocket:

SMILE 10 BSF Variant 1.3: Displayed relatively stable binding, with RMSD values stabilizing around 0.225 nm (Figure 3).

Fragmenstein Ligand 2 Step 1: Exhibited larger conformational changes with RMSD peaks around 0.45 nm, suggesting a more flexible binding mode.

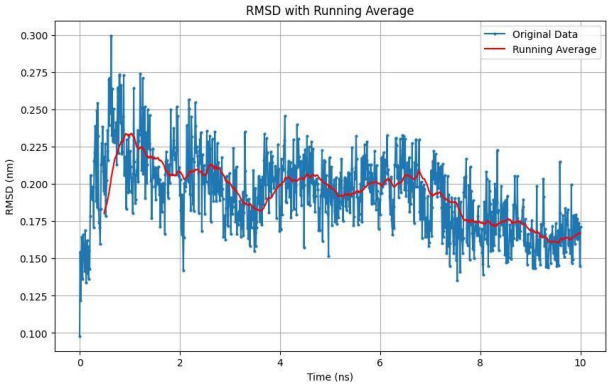

Figure 3. SMILE 10 BSF warhead variant 1.3 RMSD.

The interaction energy analysis evaluated the energetic stability of the ligands:

SMILE 10 BSF Variant 1.3: Exhibited fluctuations in interaction energy between 4650 kJ/mol and 5000 kJ/mol, with a running average around 4850 kJ/mol, indicating stable binding.

Fragmenstein Ligand 2 Step 1: Showed a wider range of energy fluctuations (4700–5050 kJ/mol) but maintained an overall stable interaction, with the running average around

4850 kJ/mol. The interaction energy profiles of both ligands indicate generally stable interactions with the target protein, despite some transient destabilization.

In this final phase, extensive exploration of chemical space and molecular dynamics (MD) simulations focused on enhancing the stability and binding affinity of Ligand 10 variant 1.3, a covalent inhibitor. A total of 78 variants were generated across two trials (Appendix C). Altering the grid box size in docking had a notable impact on results, as in variant "2-oh" which went from +349.35 to −9.67 kcal/mol. Seven variants were selected for further MD simulations. Among them, variants 5 and 48 demonstrated the best performance (Table 5)(Figure 6).

However, neither surpassed the parent ligand -Ligand 10 variant 1.3- in stability and binding interactions:

Variant 48: intermittent hydrogen bonding, with RMSD values stabilizing around 0.2 nm
Variant 5: Displayed better hydrogen bond stability. RMSD fluctuated more, reaching 0.6 nm (Figure 5).

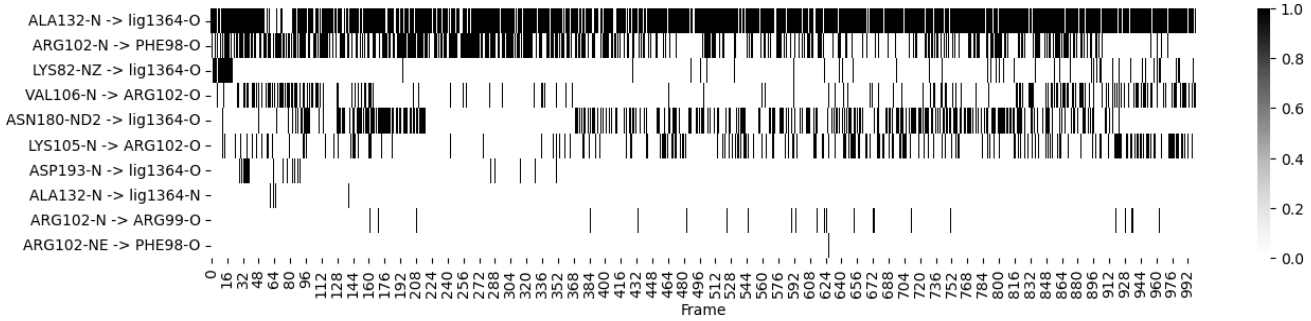

Figure 4. Heatmap H-Bond interactions Ligand 10 BSF Variant 1.3.

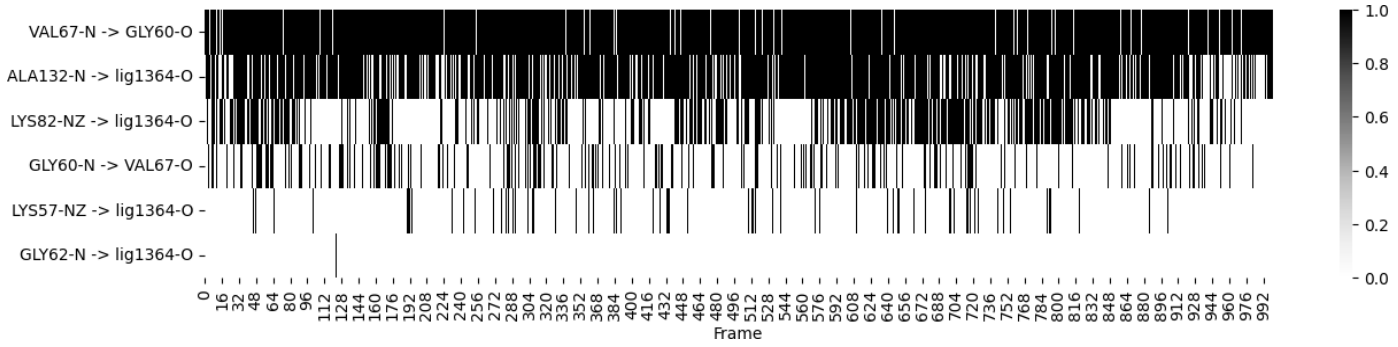

Figure 5. Heat map of variant 5 h-bond interaction.

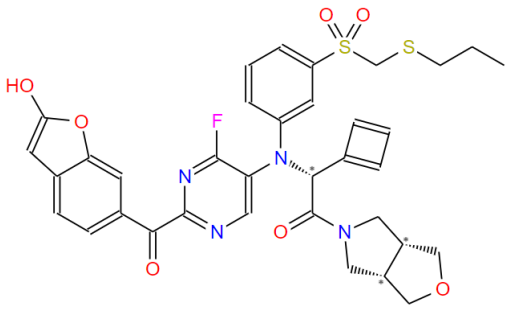

Figure 6. illustrates the structure of variant 48.

Table 5. top-performing variants of Ligand 10 V 1.3.

Trial	Variant	Binding Energy (kcal/mol)	SMILE
1	18	−10.42	CCSCS(=O)(=O)c1cccc(N(c2cnc(C(=O)c3ccc4cc(O)oc4c3)nc2F)C@HN2CCCC2)c1
1	5	−9.95	CCCSCS(=O)(=O)c1cccc(N(CC(=O)N2C[C@H]3COC[C@H]3C2)c2cnc(C(=O)c3ccc4cc(O)oc4c3)nc2F)c1
1	44	−10.39	CCSCS(=O)(=O)c1cccc(N(CC(=O)N2C[C@H]3COC[C@H]3C2)c2cnc(C(=O)c3ccc4cc(N5CCCC5)oc4c3)nc2F)c1
1	48	−10.79	CCSCS(=O)(=O)c1cccc(N(c2cnc(C(=O)c3ccc4cc(O)oc4c3)nc2F)C@@HC2=CC=C2)c1

Variant 48 achieved a binding energy of −10.79 kcal/mol, which was a significant improvement compared to the reference and initial lead compounds. Optimization efforts extended to other covalent inhibitors, with changes to the bonding of the BSF warhead to Ligand 1 variants. Ligand 1 variant 1.4 new 2 achieving a binding energy of −10.17 kcal/mol.MD simulations revealed the following (Table 6):

Ligand 1 Variant 1.1 New: Displayed intermittent hydrogen bonding with residues, with RMSD values fluctuating between 0.22 nm and 0.3 nm.
Ligand 1 Variant 1.4: Demonstrated significant H-bond stability. RMSD stabilizing 0.22 nm.
Ligand 1 Variant 1.4 New 2: stable hydrogen bonding, with RMSD stabilizing around 0.25 nm.

Table 6. top-performing variants of covalent inhibitors.

Ligand	RMSD (kcal/mol)	SMILE
Ligand 1 variant 1.4 new 2	−10.17	CCSC=S(=O)(ON(CC(=O)N1CCCC1)c4cnc(C(=O)c3ccc2cc[nH]c2c3)nc4)c5ccccc5
Ligand 10 variant 1.3 new	−9.95	CCSC=S(=O)(ON(CC(=O)N2C[C@@H]1COC[C@@H]1C2)c5cnc(C(=O)c4ccc3ccoc3c4)nc5)c6ccccc6
Ligand 10 variant 1.3 new 2	−10.14	CCS(C)=S(=O)(ON(CC(=O)N2C[C@H]1COC[C@H]1C2)c5cnc(C(=O)c4ccc3ccoc3c4)nc5)c6ccccc6

Part 4: Conclusions

This research illustrates the power of combining computational methods with human expertise to design potent inhibitors targeting SIK2/3 kinases. The optimized lead compounds demonstrated promising interactions with the protein targets, with stable binding profiles and consistent hydrogen bonding. These findings pave the way for future validation in vitro and in vivo, with the goal of developing new therapeutic drugs.

Author Contributions:

Funding:

Institutional Review Board Statement:

Informed Consent Statement:

Data Availability Statement:

Conflicts of Interest:

Appendix A

Appendix A: National Center for Biotechnology Information. (n.d.). PubChem Database. PubChem. https://pubchem.ncbi.nlm.nih.gov.

Appendix B

Supplementary computational notebooks. (n.d.). GitHub. https://github.com/EddieJosef/De_novo_design_of_sik2_sik3_inhibitors/tree/main/Notebooks.

Appendix C

Supplementary data files. (n.d.). Google Drive. https://drive.google.com/drive/folders/1OaXKp3LlCEb0P6aiJI5Ao1TMmHlPpCol.

Appendix D

Rashed, E.Y.E. (2024). Integrated Computational Approach to Rational Drug Design Targeting SIK2/3: From Theory to Practice. https://drive.google.com/file/d/1-uCXCF0vj1tVx-_PkQbDpOcKx-l6ozIe/view.

References

1. 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. https://doi.org/10.1021/acs.jmedchem.3c00485.
2. Eberhardt, J.; Santos-Martins, D.; Tillack, A.F.; Forli, S. AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings. *J. Chem. Inf. Model.* **2021**, *61*, 3891–3898. https://doi.org/10.1021/acs.jcim.1c00203.
3. Loeffler, H.H.; He, J.; Tibo, A.; Janet, J.P.; Voronov, A.; Mervin, L.H.; Engkvist, O. REINVENT4: Modern AI–Driven Generative Molecule Design. *ChemRxiv* **2023**. https://doi.org/10.26434/chemrxiv-2023-xt65x.
4. Forli, S. Meeko: Preparation of Small Molecules for AutoDock [Software]. Available online: https://github.com/forlilab/Meeko (accessed on).
5. Abraham, M.J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J.C.; Hess, B.; Lindahl, E. GROMACS: High Performance Molecular Simulations Through Multi-Level Parallelism from Laptops to Supercomputers. GROMACS. Available online: https://manual.gromacs.org/current/index.html (accessed on).

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