



Article

Multi-Criteria Decision Analysis in Drug Discovery

Rafał A. Bachorz ^{*} , Michael S. Lawless, David W. Miller and Jeremy O. Jones

Simulations Plus, Inc., Research Triangle Park, P.O. Box 12317, Durham, NC 27709, USA;
michael.lawless@simulations-plus.com (M.S.L.); david.miller@simulations-plus.com (D.W.M.);
jeremy.jones@simulations-plus.com (J.O.J.)

* Correspondence: rafal.bachorz@simulations-plus.com

Abstract: Drug discovery is inherently a multi-criteria optimization problem. In the first instance, it involves a tremendously large chemical space, where each compound can be characterized by multiple molecular and biological properties. Modern computational approaches try to efficiently explore the chemical space in search of molecules with the desired combination of properties. For example, Pareto optimizers identify a so-called “Pareto front”, a set of non-dominated solutions. From a qualitative perspective, all solutions on the front are potentially equally desirable, each expressing a trade-off between the goals. However, often there is a need to weight the objectives differently, depending on their perceived importance. To address this, we recently implemented a new Multi-Criteria Decision Analysis (MCDA) method as part of the AI-powered Drug Design (AIDDTM) technology initiative. This allows the user to weight various objective functions differently, which, in turn, efficiently directs the generative chemistry process toward the desired areas in chemical space.

Keywords: drug design; multi-criteria decision analysis (MCDA); generative chemistry



Academic Editor: Nikolaos Kourkoumelis

Received: 15 November 2024

Revised: 11 December 2024

Accepted: 20 December 2024

Published: 6 January 2025

Citation: Bachorz, R.A.; Lawless, M.S.; Miller, D.W.; Jones, J.O. Multi-Criteria Decision Analysis in Drug Discovery. *Appl. Biosci.* **2025**, *4*, 2.

<https://doi.org/10.3390/applbiosci4010002>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Drug discovery and design are complex, multifaceted processes requiring the integration of vast amounts of information to make decisions that balance efficacy, safety, and economic feasibility. The early stages of drug design typically involve the identification and optimization of molecules with potential therapeutic effects, necessitating the consideration of multiple criteria, including pharmacokinetics, pharmacodynamics, toxicity, and synthesis feasibility. The task of balancing these often-competing criteria is critical, as suboptimal trade-offs can lead to delays, increased costs, or even failures in the later stages of drug development.

Multi-Criteria Decision Analysis (MCDA) offers a structured approach to managing this complexity by providing a framework for evaluating multiple, sometimes conflicting, objectives. MCDA methodologies enable decision makers to systematically prioritize, compare, and integrate various drug properties based on their relative importance [1]. By doing so, MCDA helps streamline the decision-making process in drug design by identifying optimal candidates or eliminating compounds unlikely to meet therapeutic and safety thresholds early on.

In recent years, the advent of generative chemistry—particularly through the use of artificial intelligence (AI) and machine learning (ML)—has revolutionized molecular design. Generative models based on deep neural networks (DNNs) [2–11], Evolutionary Algorithms (EAs) [12–16], and recently, large language models (LLMs) [17–20] offer the capability to autonomously generate novel molecular structures with desired properties

by learning patterns from large chemical databases. For example, deep neural networks, including Variational Autoencoders (VAEs) and Generative Adversarial Networks (GANs), have shown promising results in generating diverse and synthetically accessible molecules that adhere to desired pharmacokinetic and pharmacodynamic profiles [21–23]. The above techniques form an efficient toolset capable of exploring vast chemical space in search of drug candidates with desired characteristics. Additionally, recent developments in the area of LLMs, which model the chemical language of molecular structures, allow for direct generation and manipulation of SMILES strings, enhancing the rapid design of new compounds based on textual representations of chemical entities. An up-to-date survey about the application of various generative approaches can be found in [24–26].

The integration of Multi-Criteria Decision Analysis (MCDA) with generative chemistry capabilities significantly enhances the design and selection process by providing a systematic framework for evaluating multiple molecular properties simultaneously. As generative models often produce thousands of candidate molecules, MCDA techniques, such as the Analytic Hierarchy Process (AHP) [27,28], ELimination Et Choix Traduisant la REalité (ELECTRE) [29], UTilités Additives (UTA) [30], Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) [31,32], Višekriterijumsko KOmpromisno Rangiranje (VIKOR) [33–35], or other methods, could be applied to rank these candidates based on criteria such as synthetic feasibility, predicted biological activity, toxicity, and novelty. This approach enables a more strategic and targeted selection from the pool of generated molecules, prioritizing those that best balance multiple critical attributes. By combining the exploratory power of the generative models with the structured evaluation of MCDA, researchers can streamline the discovery of optimal drug candidates and reduce the risk of late-stage attrition. This fusion of generative chemistry and MCDA holds transformative potential for drug design, promising both efficiency in exploring chemical space and rigor in candidate evaluation. There are already applications of MCDA methods supporting selection in the drug design process. For instance, Peng et al. [36] applied the SMAA-TOPSIS approach to rank candidate therapeutics, specifically antibodies for Alzheimer's disease, based on multiple uncertain screening criteria. Jinmiao et al. [37] proposed the TOPSIS method to support the selection of generic antibiotics. There have also been multiple applications of MCDA techniques in the area of Health Technology Assessment (HTA). Baran-Kooiker et al. provided a systematic review of Multi-Criteria Decision Analysis (MCDA) models for orphan drugs [38]. Angelis and Kanavos [39] proposed the Advance Value Framework, a structured MCDA-based approach that evaluates drugs on multiple dimensions beyond cost-effectiveness, such as therapeutic impact, safety, innovation, disease burden, and socioeconomic impact.

This paper examines the role of MCDA methods in advancing the exploration of chemical space within the generative chemistry process. Specifically, we discuss the incorporation of the VIKOR approach, an MCDA technique, into the AIDDTM drug discovery pipeline to guide chemical space exploration toward candidates that exhibit desired properties. This integrated method facilitates the seamless incorporation of user preferences by allowing weight assignments to objective functions, enabling the algorithm to account for multiple objectives and their relative importance. As a result, this approach allows for efficient compound ranking, directing the generative chemistry algorithm to explore a focused, user-defined segment of chemical space. By narrowing the search, this method increases efficiency and reduces overall costs in the drug discovery process.

2. Materials and Methods

2.1. Integration of MCDA in AIDDTM

AIDDTM is a generative chemistry engine coupled with Multi-Parameter Optimization (MPO) algorithms and exists as a module within ADMET Predictor® [16]. It applies a set of chemical transformation rules expressed in the SMIRKS language. New compounds, created within the evolutionary loop, are subject to quality estimation with certain objective functions. The best trade-off molecules chosen within the pruning procedure are systematically improved as the optimization evolves.

As a default pruning component in AIDDTM, we have utilized the methodology based on Pareto pruning, which focuses on filtering generated compounds by eliminating those that are Pareto-dominated, that is, those that are outperformed by others across all criteria. This process identifies a set of “Pareto-optimal” alternatives (here and throughout, the term “alternative” should be understood as a molecule), where no single molecule is better in the context of a certain objective without making sacrifices for others. For two alternatives, x and y , we say that x Pareto dominates y if x is at least as good as y across all criteria, and at the same time, strictly better in at least one criterion. The formal expression for Pareto dominance is

$$x \succ_{\text{Pareto}} y \iff (\forall i \in \{1, \dots, n\}, f_i(x) \leq f_i(y)) \wedge (\exists j \in \{1, \dots, n\}, f_j(x) < f_j(y)) \quad (1)$$

with the assumption that the multi-criteria problem is expressed such that all objective functions are of cost type, i.e., they are intended to be minimized. Pareto dominance is a very useful concept that helps filter out all dominated alternatives, leaving a set of non-dominated compounds that are optimal in some sense across multiple criteria. However, the main drawback of the Pareto procedure is that it treats all compounds on the Pareto front as equally optimal (which, in the literature, is called Pareto-optimal) without any further qualitative distinction. Since each compound on this front is not dominated by any other, they are all considered “optimal” in the sense of Pareto efficiency (see Figure 1). This can be limiting because it does not provide a way to differentiate between these alternatives, even though decision makers may still have preferences or priorities within this set. Multi-Criteria Decision Analysis (MCDA) approaches offer valuable tools to address this limitation. Unlike the Pareto procedure, which identifies only non-dominated solutions, MCDA methods provide a framework to rank and differentiate between the Pareto-optimal compounds based on the combination of objective function values and decision-maker preferences. One of many MCDA techniques is the VIKOR approach, which is well suited for helping rank alternatives based on a combination of ideal solutions and decision-maker preferences [33]. Unlike the Pareto method, which only identifies non-dominated solutions, VIKOR aims to find a compromise solution that balances the trade-offs between multiple criteria, considering both the ideal (best possible) and anti-ideal (worst possible) values for each criterion. In VIKOR, each alternative is evaluated by calculating two measures (utility (S) and regret (R)), which together capture how close each alternative is to the ideal solution. The calculations begin by determining the ideal f_i^* and anti-ideal f_i^- values for each criterion i , which are defined as

$$f_i^* = \min_j f_i(x_j) \quad \text{and} \quad f_i^- = \max_j f_i(x_j) \quad (2)$$

where $f_i(x_j)$ represents the performance of alternative x_j on criterion i , and the indices i and j span the criteria and alternatives, respectively. The utility (S) for each alternative x_j is then computed as the weighted sum of the distances from the ideal value across all criteria:

$$S_j = \sum_{i=1}^n w_i \frac{f_i^* - f_i(x_j)}{f_i^* - f_i^-} \quad (3)$$

where w_i is the weight assigned to criterion i , which reflects its relative importance. The utility value, S_j , reflects the overall closeness of x_j to the ideal solution, capturing the alternatives' general performance across all criteria. The regret (R) measure, on the other hand, focuses on the maximum individual criterion regret, indicating the worst-case deviation from the ideal for any single criterion:

$$R_j = \max_i \left(w_i \frac{f_i^* - f_i(x_j)}{f_i^* - f_i^-} \right) \quad (4)$$

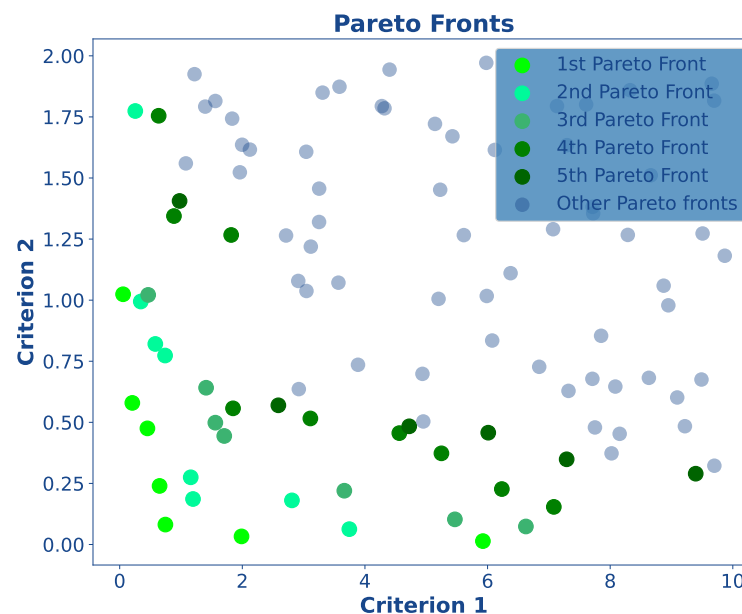


Figure 1. Illustration of multiple Pareto fronts based on two criteria. Each point represents an alternative in the two-dimensional space of both criteria, with the lowest Pareto front (1st Pareto front) containing non-dominated alternatives, and each subsequent front containing points that are dominated by all alternatives in the previous fronts. All alternatives on the same Pareto front are considered qualitatively equivalent among all the alternatives on that front. This highlights a limitation of the Pareto approach, i.e., the lack of the ability to distinguish between alternatives on a given front.

The regret measure R_j provides insight into the greatest weakness of the alternatives, offering an alternative perspective on utility by highlighting the criterion with the greatest shortfall relative to the ideal. VIKOR then combines these two measures into an aggregated Q score to find a compromise solution (see Figure S1 for a conceptual illustration of the VIKOR method). This is achieved by introducing a preference parameter v , which reflects the decision maker's tendency toward group benefit (if v is closer to 1, giving more weight to utility) or individual satisfaction (if v is closer to 0, emphasizing regret). The Q value for each alternative x_j is calculated as

$$Q_j = v \frac{S_j - S^*}{S^- - S^*} + (1 - v) \frac{R_j - R^*}{R^- - R^*} \quad (5)$$

where S^* and S^- are the minimum and maximum values of S_j across all alternatives, and R^* and R^- are the minimum and maximum values of R_j across all alternatives. The Q score provides a composite ranking by balancing the compromise between utility and regret. In this study, the preference parameter v was chosen to be 0.5. The VIKOR method described

above provides a convenient toolset for ranking alternatives, in the context of generative chemistry—compounds, within the Pareto fronts. This is achieved by calculating Q scores that reflect their closeness to the ideal solution, adjusted by decision-maker preferences for balance or minimization of regret. This clearly enables more refined decision making by highlighting compromise solutions and offering a way to differentiate between compounds of “equal quality” on the Pareto front (see Figure 2).

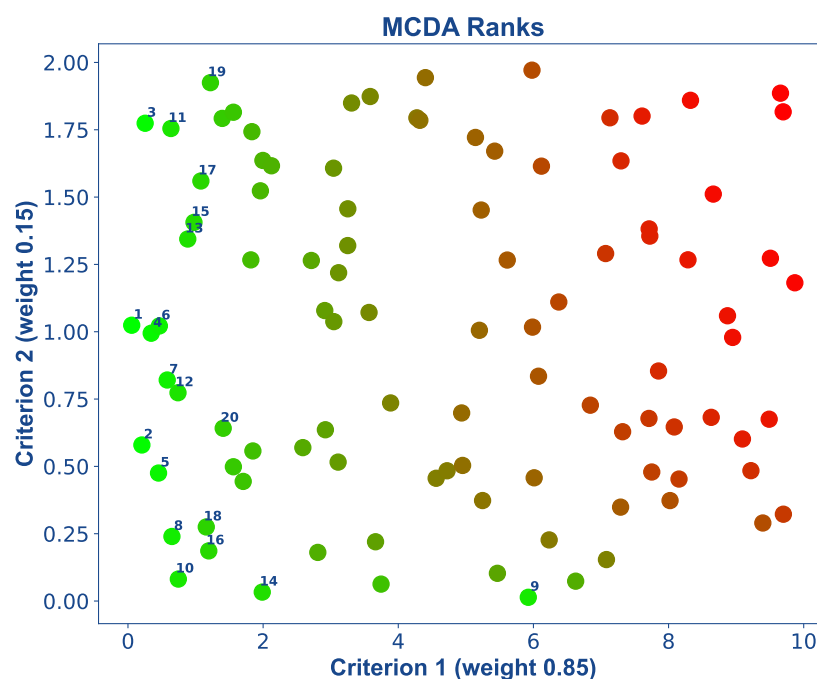


Figure 2. A set of alternatives evaluated across two differently weighted criteria: Criterion 1 (weight 0.85) and Criterion 2 (weight 0.15). Each numbered point represents an alternative uniquely ranked by the VIKOR method. By incorporating criterion weights and calculating a combined Q score, VIKOR enables differentiation among alternatives that might otherwise appear equally optimal on the same Pareto front, providing a more refined decision-making framework.

The method described above has been seamlessly integrated into the AIDDTM module [16]. As part of the generative chemistry task definition, along with other parameters, the user selects objective functions that guide the exploration of the chemical space. Here, the full range of ADMET (Absorption, Distribution, Metabolism, Excretion, and Toxicity) and HTPK (High-Throughput Pharmacokinetics) properties is available as potential goal functions. In each generation cycle, newly created compounds undergo quality assessment. All properties previously defined as objective functions are calculated, and a pruning procedure is applied to rank the compounds based on quality estimates, selecting the best ones to carry forward to the next generation.

The key innovation of this work introduces a new capability at the pruning stage: the user can assign a preference weight to each MPO objective function, thereby activating an alternative pruning scheme based on the VIKOR method discussed above. The ranking Q obtained with Equation (5) explicitly incorporates the objective functions and weights, as shown in Equations (3) and (4). This approach enables individual ranking for each compound, which is directly utilized as a compound selection method.

2.2. Application Use Cases

The crystal structures of the benzimidazole sulfonamide complex with the PXR receptor, and ciprofloxacin complex with 24-mer of DNA and DNA gyrase, were downloaded from the PDB data base (PDB ID: 5a86 and 5btc, respectively) and processed with the ProDy

library [40] in the Python 3.9 environment. The generative chemistry calculations were carried out with the AIDDTM module in ADMET Predictor® 12. The key technical settings of the AIDDTM calculations were as follows:

- Number of optimization generations: 30–50;
- Number of candidate molecules per generation: 1000–2500;
- Size of initial population: 2000–10,000;
- In the case of PXR use cases, the following objective function capping values were applied:
 1. PXR activity: 3.5;
 2. Sim3D: 0.88;
 3. SynthDiff+: 2;
 4. ADMET Risk: 1.

The choice of the objective functions and their MCDA weights was specific to each use case and is discussed in dedicated sections. The 3D visualization of the compounds was carried out in the BIOVIA Discovery Studio [41], and for the 2D visualization, the RDKit library was used [42].

3. Results and Discussion

The Pregnane X Receptor (PXR) is a nuclear receptor that regulates the expression of genes involved in drug metabolism and transport. It is important to target PXR because it plays a critical role in protecting the body from toxic substances by activating the expression of detoxifying enzymes and transporters upon binding to various endogenous and exogenous ligands. We chose this receptor as a target for the validation and illustration of MCDA pruning in the AIDDTM module. Two independent use cases were considered: the PXR crystal structure rediscovery and the PXR ligand optimization. Additionally, to more thoroughly assess the performance of the VIKOR method, we considered an additional use case, where the oral bioavailability of a known antibiotic—ciprofloxacin—was the key objective. Three calculations involving different pruning schemes were carried out: Pareto, TOPSIS, and VIKOR.

3.1. PXR Crystal Structure Rediscovery

The study began by downloading the 5a86 receptor structure, which served as the basis for our virtual screening experiments. This crystal structure represents the benzimidazole sulfonamide bound to the Pregnane X Receptor [43]. From this protein–ligand complex, we extracted the 3D coordinates of the existing ligand bound within the receptor's binding site. These coordinates provided a crucial structural reference, allowing us to incorporate 3D similarity as a key objective in our compound generation and optimization process.

The AIDDTM calculations were configured with three objective functions: 3D similarity to the reference crystal structure (Sim3D), synthetic difficulty (SynthDiff+), and ADMET risk (ADMET_Risk). Benzimidazole was chosen as the seed structure for this generation process due to its relevance to the target receptor and its potential as a pharmacophore scaffold (see Figure 3). The goal of the calculations was mainly to identify compounds that maintain a high degree of structural similarity to the reference ligand and are also synthetically feasible and exhibit favorable ADMET profiles. Two variants of the AIDD calculation were conducted to evaluate the effectiveness of different pruning strategies: (1) traditional Pareto pruning and (2) Multi-Criteria Decision Analysis (MCDA) pruning. For the MCDA pruning approach, specific weights were applied to the objective functions to align with the study's primary objective—recovering the 3D structure of the reference ligand. Therefore, the 3D similarity objective was assigned a weight of 70%, emphasizing its importance. Synthetic difficulty and ADMET risk were each assigned a weight of 15%, balancing synthetic feasibility and pharmacokinetic considerations without overshadowing the focus on structural similarity.

This weighting scheme within MCDA pruning allowed us to prioritize candidates that closely resembled the reference structure, thus increasing the likelihood of generating relevant bioactive compounds. The MCDA pruning approach effectively channeled the exploration of chemical space toward compounds with higher structural similarity to the reference crystal structure. This is evident in Figure 4, where the distribution of the Sim3D objective function obtained with MCDA pruning is clearly skewed toward higher similarity scores, reflecting the MCDA framework's prioritization of structural resemblance. This outcome underscores the method's capability to target specific molecular characteristics in generative chemistry. It is worth explaining why, despite relatively high values, the similarity measures remain noticeably below 1. A significant part of this phenomenon arises from the fact that the geometry of the reference crystal structure is influenced to some extent by the protein interacting with the ligand in the binding pocket. In contrast, the calculation of the Sim3D similarity only compares the reference structure to candidate structures, which are generated without any information about the protein's impact on the candidate molecule. The high similarity of rediscovered structures to the target is illustrated in Figure 3. Here, it can be seen that the resulting compounds exhibit minor structural differences; for example, in the upper structure, a methyl group is missing in the methylene bridge connecting the benzimidazole and phenylsulfonamide moieties. The efficiency of the MCDA pruning approach proposed here is also evident in Figure 5, which shows the best and worst cases in terms of the Sim3D similarity. In both examples, the candidate molecule is superimposed with the reference crystal geometry, with the latter shown in green. In the top part, we observe almost perfect structural alignment; the minor differences in the 3D fit are primarily due to the effect described above. In contrast, in the bottom part of the figure, substantial structural mismatches are noticeable in both terminal components of the compound. The trifluoromethyl group has been replaced with a fluorine atom, and the methoxy group has taken the place of a chlorine atom in the benzimidazole and phenylsulfonamide moieties, respectively.

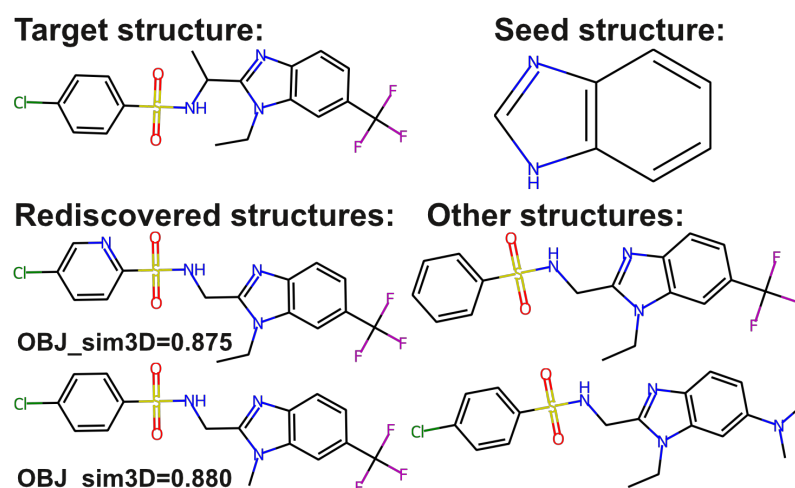


Figure 3. The seed structure (benzimidazole), target, and rediscovered structures resulting from the calculation in which MCDA pruning was applied to the 5a86 ligand rediscovery.

In summary, the application of MCDA pruning enabled a more directed exploration of the chemical space. The strong emphasis on the Sim3D objective during the iterative generative chemistry procedure led to an almost perfect rediscovery of the target compound. It is worth noting that, while the benzimidazole seed compound was a reasonable starting guess, it represented only a fraction of the final compound. Only the combination of an appropriate objective function setup, efficient generative chemistry algorithms, and MCDA

pruning—all available in the AIDDTM module—made the successful crystal structure rediscovery possible.

One might ask how to quantitatively describe the benefits of applying MCDA pruning in the generative chemistry process discussed above. As shown in Figure 4, the maximum values of the Sim3D objective function were achieved in both the Pareto and MCDA cases. However, when using Pareto pruning as a reference, the number of generated compounds above the 95th percentile for this objective is 81, whereas with MCDA pruning, this number increases to 236. This indicates that the number of generated compounds in this crucial regime is nearly three times higher when MCDA pruning is applied, which subsequently translates to increased chances of hit finding, provided the compounds meet the expected quality assumptions.

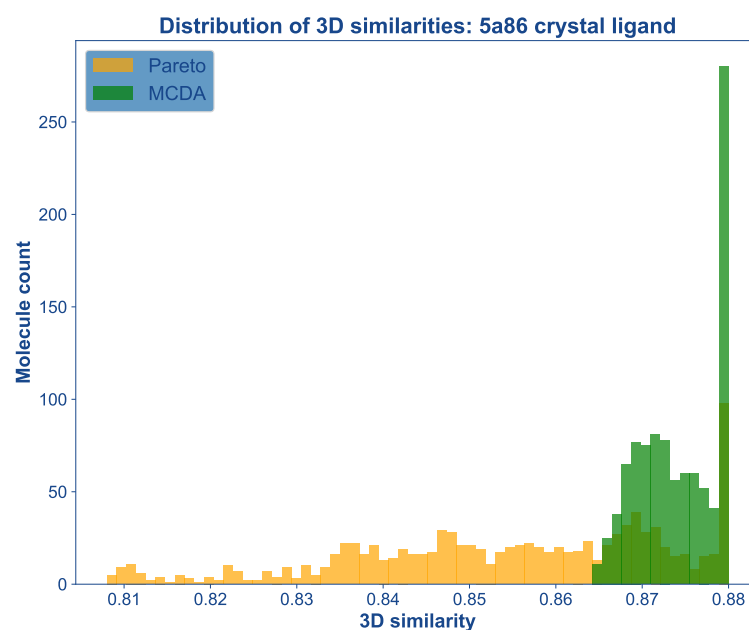
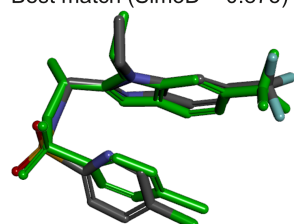


Figure 4. Histogram showing the distribution of 3D similarity scores for compounds generated using MCDA and Pareto pruning applied within the AIDDTM module.

Best match (Sim3D = 0.875)



Worst match (Sim3D = 0.81)

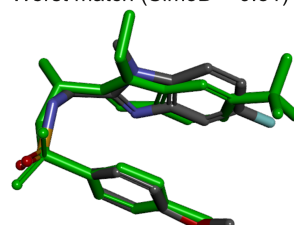


Figure 5. Overlay of candidate and reference (green) compounds, illustrating the structural similarity achieved. The upper image shows the best match obtained with MCDA pruning, with a high 3D similarity score (Sim3D = 0.875). The lower image displays one of the worst matches, with a lower Sim3D score (Sim3D = 0.81), reflecting a less favorable alignment between the candidate and reference compounds.

3.2. PXR Ligand Optimization

The second use case involved optimizing the ligand to reduce its PXR biological activity. Hennessy et al. [43] identified a potent Sphingosine 1-Phosphate Receptor antagonist with favorable oral bioavailability, which was also the subject of the crystal structure rediscovery discussed in the previous section. Unfortunately, the discovered compound, shown in Figure 6, turned out to be a potent CYP3A4 inducer. We decided to apply the AIDDTM module, coupled with MCDA pruning, to optimize the ligand, with the primary objective of reducing its PXR agonistic activity while maintaining structural similarity to the crystal reference compound.

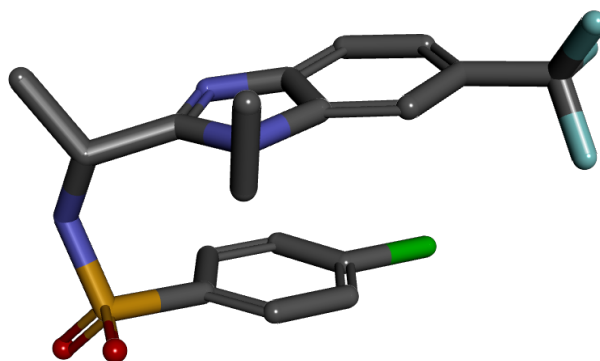


Figure 6. The reference 5a86 ligand structure in its native experimental conformation, serving as a baseline for 3D similarity assessments in compound optimization.

The first step involved creating a QSAR activity model, reflecting the agonistic biological activity with respect to the PXR receptor. To achieve this, we extracted a dataset of 516 compounds from the ChEMBL database [44]. The data underwent manual curation, focusing primarily on the analysis of applied biological assays. For compounds with more than one EC₅₀ value, the mean value was calculated and used for further analysis. In cases of conflicting values, the literature sources were manually reviewed, and less credible values were excluded. To ensure that the test set was a representative sample of the training data, the split into test and training sets was performed using self-organizing Kohonen maps. The final distribution of data points for both the training and test sets in terms of pEC₅₀ values is shown in Figure 7. The QSAR model was developed using the ADMET Modeler program, a component of the ADMET Predictor® suite. An ensemble of linearly boosted artificial neural networks was employed for this purpose, providing a robust approach to predict PXR agonist activity. The model demonstrated satisfactory performance, achieving RMSE and MAE values of 0.468 and 0.380, respectively, on the test set, as shown in Figure 8. The AIDDTM calculation was configured with three primary objective functions: pEC₅₀ with respect to the PXR receptor, 3D similarity (Sim3D) to the reference crystal structure, and synthesis difficulty (SynthDiff+). The pEC₅₀ objective function was set for minimization to reduce PXR receptor activation, which is crucial for avoiding CYP3A4 induction. In contrast, the Sim3D objective was aimed at maximization to ensure high structural similarity with the reference ligand, while SynthDiff+ was also minimized to simplify the synthesis of candidate compounds. MCDA weights were assigned to these objectives as follows: pEC₅₀ at 50%, Sim3D at 30%, and SynthDiff+ at 20%, reflecting their relative importance in the optimization.

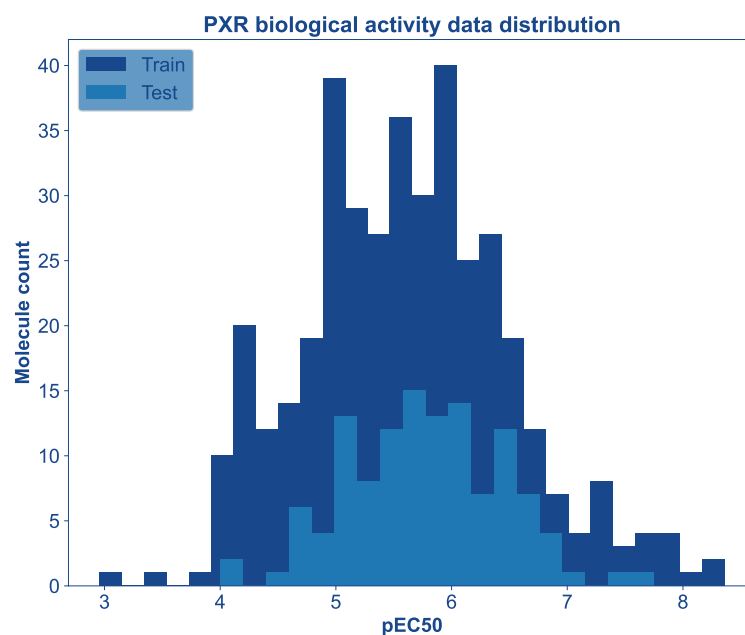


Figure 7. Distribution of pEC₅₀ biological activity values for the PXR receptor, obtained from the ChEMBL database. The training data are shown in dark blue, while the test data are represented in light blue.

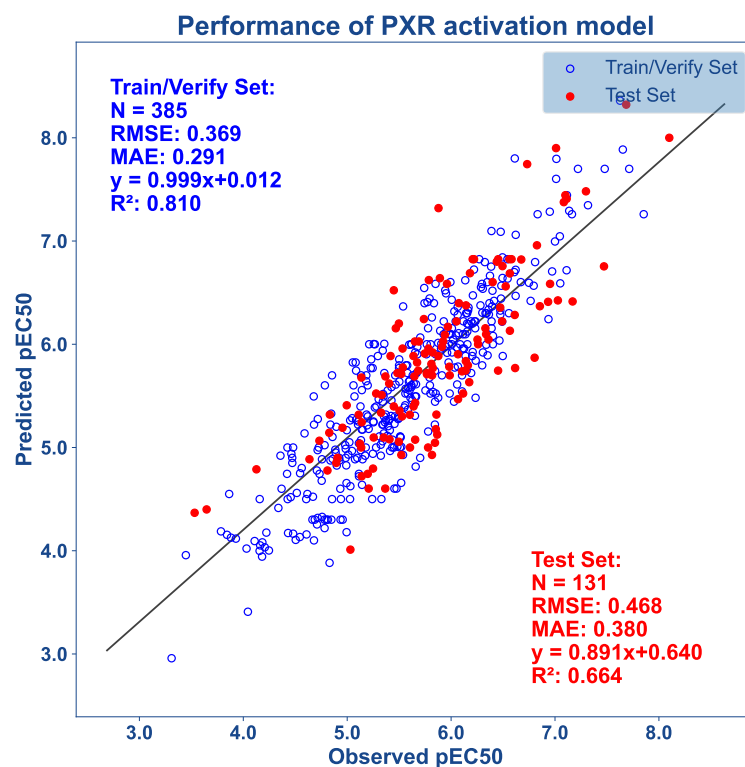


Figure 8. Plot of predicted vs. observed pEC₅₀ values, illustrating the quality of the QSAR regression model for agonistic biological activity toward the PXR receptor. On the test set, the model achieved an RMSE of 0.468 and an MAE of 0.380, indicating satisfactory predictive capability.

The resulting compounds are shown in Figure 9. At the top, the initial structure is presented, which was the subject of optimization and served as a reference for 3D shape matching. This structure is characterized by a relatively high predicted pEC₅₀ of approximately 6.2 and, as experimentally measured [43], is a potent CYP3A4 inducer. This finding was also confirmed by our in-house CYP3A4 induction classifier. The optimized

structures are displayed in the bottom part of Figure 9. Each has a substantially lower pEC₅₀, ranging from 3.0 to 4.0, clearly indicating a 35–50% reduction in agonistic activity with respect to the PXR receptor. As expected, this low PXR activation potential translates to the absence of predicted CYP3A4 induction. At the same time, all four compounds demonstrate high 3D similarity to the reference structure (see Figure 10), suggesting that they remain effective antagonists of the Sphingosine 1-Phosphate Receptor. Interestingly, in all proposed structures, the benzimidazole motif was excluded in favor of the furo[2,3-c]pyridine moiety, which turned out to be a key structural modification.

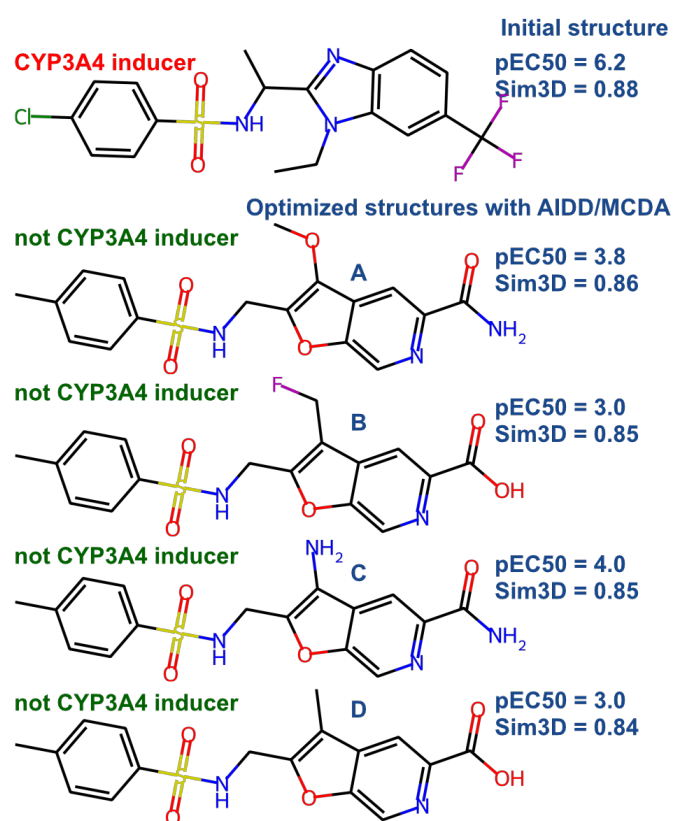


Figure 9. The optimized ligands aim to reduce activation of the PXR receptor while maintaining 3D similarity to the crystal structure reference. At the top is the initial structure from the 5a86 crystal structure, and at the bottom are four proposed analogs exhibiting low PXR activity and high 3D similarity.

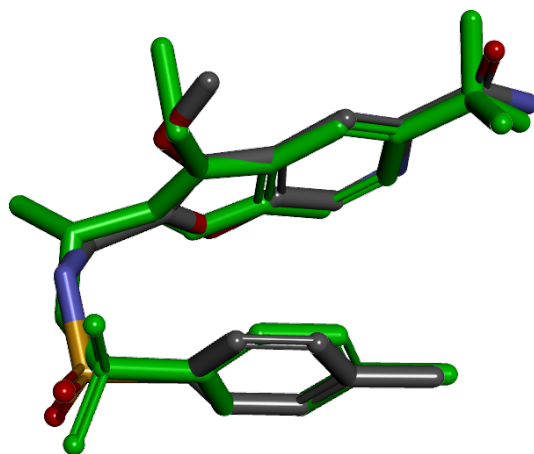


Figure 10. Overlay of the reference ligand reference structure (in green) with the selected optimized compound.

In summary, the application of the AIDDTM technology, coupled with the proposed MCDA pruning strategy, allowed for achieving key goals: mitigating the CYP3A4 induction effect and keeping the compounds structurally similar to the reference molecule.

3.3. Optimization of Oral Bioavailability of Ciprofloxacin

Ciprofloxacin is a broad-spectrum antibiotic that belongs to the fluoroquinolone class of drugs. It is widely used to treat a variety of bacterial infections, including those affecting the respiratory tract, urinary tract, gastrointestinal system, skin, and soft tissues. Ciprofloxacin's mode of action involves the inhibition of bacterial DNA gyrase and topoisomerase IV, enzymes critical for bacterial DNA replication, transcription, repair, and recombination, which ultimately leads to bacterial cell death [45]. The measured oral bioavailability of ciprofloxacin, defined as the ratio of the area under the curve from 0 h to infinity for the oral and intravenous doses, was estimated to be $69 \pm 7\%$ [46]. The absorption occurs primarily in the small intestine, and the bioavailability is minimally affected by food intake, allowing for flexible dosing schedules [47]. However, ciprofloxacin's moderately high oral bioavailability can be significantly reduced by several factors, e.g., co-administration with di- and trivalent cations [48], various dietary factors [49], and others. Therefore, we aimed to optimize the molecular structure of ciprofloxacin to enhance its oral bioavailability. As a key objective function, we chose fraction bioavailable (%Fb), a predicted in vivo endpoint available in the HTPK module. Here, we used human physiology parameters and a dose of 200 mg. All necessary properties required to carry out the HTPK simulation, such as liver microsome clearance, octanol-water partition coefficient (logP), solubility, permeability, and others, were calculated within ADMET Predictor 12. Additionally, we incorporated 3D similarity to the known crystal structure complex (PDB: 5btc) formed between DNA 24-mer, DNA gyrase subunit B, and ciprofloxacin [50]. This objective ensured that the optimization maintained high structural and pharmacophoric similarity, which is crucial for retaining the inhibitory efficacy of the resulting compounds. As the third objective, we included synthesis difficulty, which helped to avoid overly complex structures. To qualitatively and quantitatively assess the performance of VIKOR pruning, we applied three pruning schemes based on different ranking procedures: Pareto (default), VIKOR, and TOPSIS. The TOPSIS method is another widely applied MCDA approach [51], and similar to the VIKOR approach, it is based on the distance from the ideal solution. Although the methods are conceptually similar, there are several technical differences [51–53]. The key difference is related to how the distances from the ideal and anti-ideal solutions are expressed. In the VIKOR method, the regret and utility measures are introduced (see Equations (3) and (4)), whereas in the TOPSIS method, these distances are expressed as Euclidean metrics. For a more in-depth comparative study of both approaches, see the study by Shekhovtsov et al. [52]. The AIDD calculations were carried out with the following assumed MCDA weights: 3D similarity (Sim3D) 60%, fraction bioavailable (%Fb) 30%, and synthetic difficulty (SynthDiff+) 10%. The histograms of the key objective function values of the resulting populations are presented in Figures 11 and 12. A single-objective analysis leads to the conclusion that the VIKOR method achieves the best performance in terms of 3D similarity, whereas the highest %Fb values are obtained with Pareto pruning. However, if both objective functions are considered simultaneously, a different conclusion can be drawn. Figure 13 presents the distribution of the 3D similarity objective function for the resulting population from Pareto, VIKOR, and TOPSIS molecular optimization calculations. The presented data were limited to the top 500 compounds (which amount to half of the entire population) with respect to the %Fb objective. It is evident that a significant number of compounds with high %Fb values obtained through Pareto pruning, as shown in Figure 12, are poorly represented in terms of 3D similarity.

From this perspective, both VIKOR and TOPSIS pruning schemes are significantly superior to the Pareto approach. Additionally, it is evident that the VIKOR method provides a better trade-off between the considered key objective functions. Similarly, if we calculate the upper 90th percentiles of the Sim3D and %Fb objective functions and select the compounds for which both objectives exceed these values, in the case of Pareto pruning, no such compound exists, whereas for the VIKOR and TOPSIS methods, there are three and seven such compounds, respectively. As the final result of the successful molecular optimization, the plasma concentration curves are presented in Figure 14. These curves were generated based on the HTPK simulation using the same settings as those applied during the AIDD molecular optimization. Clearly, the new compound, while retaining the same quinolone scaffold, exhibits significantly improved pharmacokinetic (PK) properties. The structural modification resulted in a substantial change in hydrophilicity and permeability, ultimately resulting in an enhancement of oral bioavailability by more than 40%.

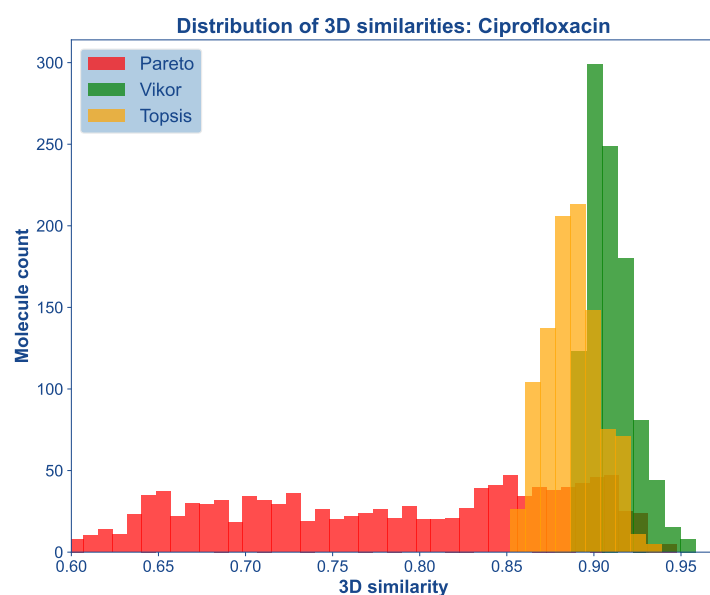


Figure 11. Histogram showing the distribution of 3D similarity scores for compounds generated using VIKOR, TOPSIS, and Pareto pruning in the ciprofloxacin molecular optimization.

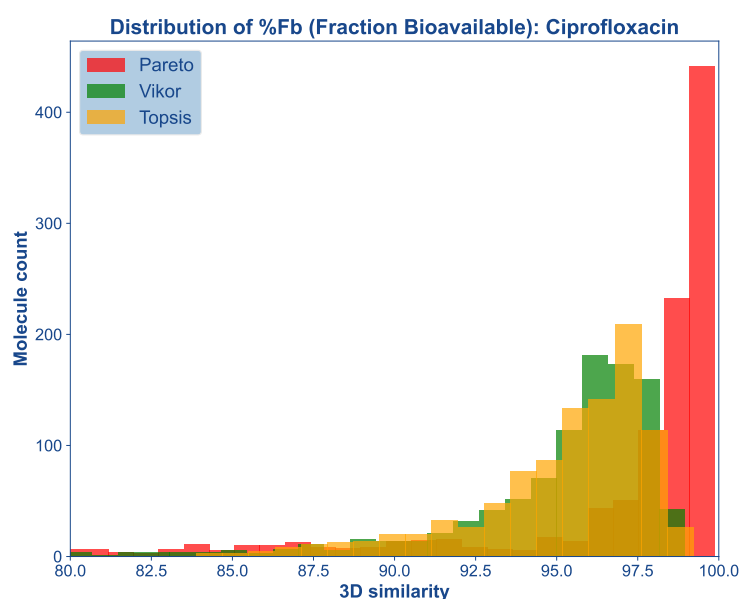


Figure 12. Histogram showing the distribution of fraction bioavailable (%Fb) for compounds generated using VIKOR, TOPSIS, and Pareto pruning in the ciprofloxacin molecular optimization.

Distribution of Sim3D for top 500 (with respect to %Fb) compounds: Ciprofloxacin

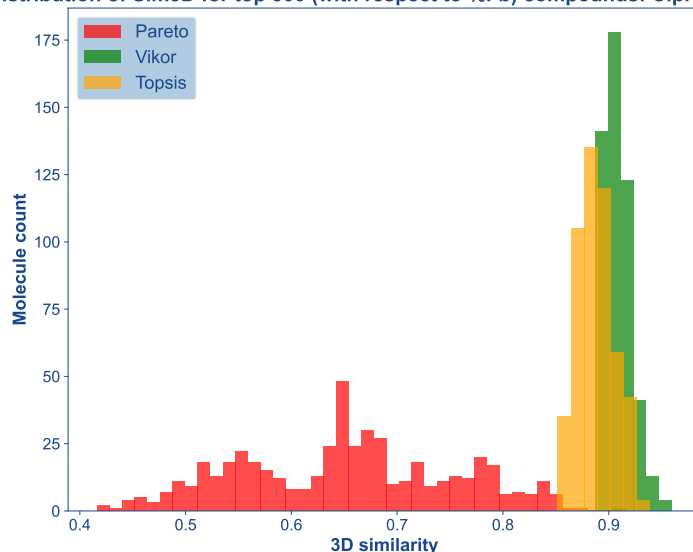


Figure 13. Histogram showing the distribution of 3D similarity (Sim3D) for compounds generated using VIKOR, TOPSIS, and Pareto pruning in ciprofloxacin molecular optimization. For each pruning scheme, the data were narrowed down to the top 500 compounds with respect to the %Fb objective. This enables a quantitative estimation of the quality of the trade-offs made by each pruning scheme.

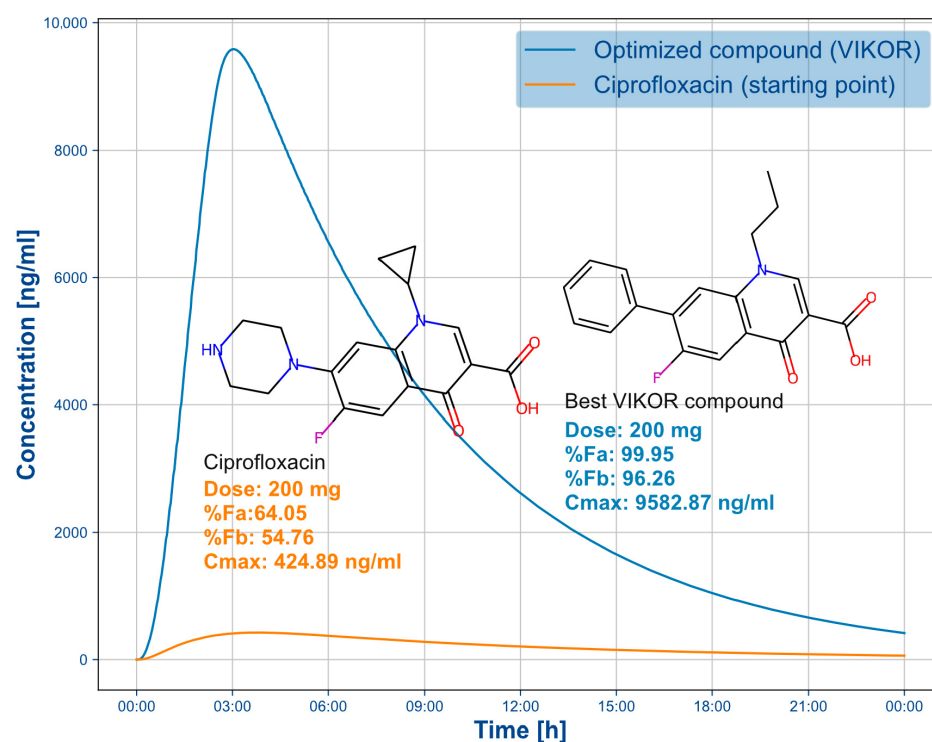


Figure 14. Predicted plasma concentration curves for ciprofloxacin (orange) and the optimized compound (blue). The molecular optimization, supported by the VIKOR pruning scheme, resulted in a structure with significantly improved pharmacokinetic (PK) characteristics while retaining the key structural features of ciprofloxacin.

The ability to weight the objective function during molecular optimization strongly influences the resulting population of compounds. This is, of course, the intended purpose; however, if MCDA pruning is applied incorrectly, the algorithm's focus may shift to a region of the chemical space that does not align with the user's preferences. Conversely, when applied correctly, MCDA pruning—particularly in the form of the VIKOR method—can

act as a highly efficient optimization driver, guiding the chemical space search toward relevant areas.

4. Summary

In this work, the VIKOR technique was integrated with the AIDD generative chemistry module [16]. The theoretical background of the method, as well as some practical aspects of the implementation, were thoroughly discussed in Section 2.1. This new capability provides an additional tool for supporting directed chemical space exploration. Simple user preference elicitation, implemented through the assignment of weights to objective functions, ultimately translates into compound designs that more closely align with the expectations and intuition of medicinal chemists. The application use cases presented—crystal structure rediscovery and ligand optimization for low PXR activity (see Sections 3.1 and 3.2), as well as the oral bioavailability optimization of ciprofloxacin (see Section 3.3)—demonstrate the method's effectiveness. In all cases, the stated goals were achieved, and the properties of the newly created compounds were consistent with the defined MCDA objective weights. In the last use case, the VIKOR method was thoroughly compared to another widely applied MCDA technique—TOPSIS. Both approaches provide the objective weighting capability; thus, from this perspective, they outperform traditional Pareto pruning. Moreover, in the case of ciprofloxacin molecular optimization, the VIKOR method proved to be more effective than TOPSIS. The MCDA component proved effective for property-driven exploration of chemical space, making it a useful tool in the challenging drug design process. As a direction for future work, we plan to apply the presented method to various hit-discovery problems and to integrate alternative MCDA techniques into the AIDDTM generative chemistry workflow.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/applbiosci4010002/s1>, Figure S1: Vikor method visualization.

Author Contributions: Conceptualization, R.A.B.; methodology, R.A.B., M.S.L., D.W.M. and J.O.J.; software, R.A.B. and D.W.M.; validation, R.A.B., M.S.L., D.W.M. and J.O.J.; formal analysis, R.A.B.; investigation, R.A.B.; data curation, R.A.B.; writing—original draft preparation, R.A.B.; writing—review and editing, R.A.B., M.S.L., D.W.M. and J.O.J.; visualization, R.A.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research was internally funded by Simulations Plus, Inc.

Data Availability Statement: All data used and generated in the course of this work will be shared upon request.

Conflicts of Interest: The authors are employed by the company Simulations Plus, Inc., which was a funder of this study. As employees of Simulations Plus, the authors have potential conflicts of interest in writing this article as it discusses products and services directly related to the company's business operations. While all efforts have been made to maintain objectivity and transparency, readers should be aware of these potential conflicts.

Abbreviations

The following abbreviations are used in this manuscript:

MCDA	Multi-Criteria Decision Analysis
AIDD TM	Artificial Intelligence-Driven Drug Design
AI	Artificial Intelligence
ML	Machine Learning
LLM	Large Language Model
VAE	Variational Autoencoder
GAN	Generative Adversarial Network

AHP	Analytic Hierarchy Process
ELECTRE	Elimination and Choice Expressing Reality
UTA	UTilités Additives (French)
TOPSIS	Technique for Order of Preference by Similarity to Ideal Solution
VIKOR	VIšekriterijumsko KOmpromisno Rangiranje (Serbian)
TOPSIS	Technique of Order Preference Similarity
HTPK	High-Throughput Pharmacokinetics
MPO	Multi-Parameter Optimization

References

- Greco, S.; Słowiński, R.; Wallenius, J. Fifty Years of Multiple Criteria Decision Analysis: From Classical Methods to Robust Ordinal Regression. *Eur. J. Oper. Res.* **2024**, *in press*. [\[CrossRef\]](#)
- Grisoni, F.; Moret, M.; Lingwood, R.; Schneider, G. Bidirectional Molecule Generation with Recurrent Neural Networks. *J. Chem. Inf. Model.* **2020**, *60*, 1175–1183. [\[CrossRef\]](#)
- Kadurin, A.; Nikolenko, S.; Khrabrov, K.; Aliper, A.; Zhavoronkov, A. druGAN: An Advanced Generative Adversarial Autoencoder Model for de Novo Generation of New Molecules with Desired Molecular Properties in Silico. *Mol. Pharm.* **2017**, *14*, 3098–3104. [\[CrossRef\]](#)
- 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. [\[CrossRef\]](#)
- Bian, Y.; Wang, J.; Jun, J.J.; Xie, X.Q. Deep Convolutional Generative Adversarial Network (dcGAN) Models for Screening and Design of Small Molecules Targeting Cannabinoid Receptors. *Mol. Pharm.* **2019**, *16*, 4451–4460. [\[CrossRef\]](#)
- Yoshikai, Y.; Mizuno, T.; Nemoto, S.; Kusahara, H. A Novel Molecule Generative Model of VAE Combined with Transformer for Unseen Structure Generation. *arXiv* **2024**, arXiv:2402.11950.
- Wang, Y.; Zhao, H.; Sciabola, S.; Wang, W. cMolGPT: A Conditional Generative Pre-Trained Transformer for Target-Specific De Novo Molecular Generation. *Molecules* **2023**, *28*, 4430. [\[CrossRef\]](#)
- Kusner, M.J.; Paige, B.; Hernández-Lobato, J.M. Grammar Variational Autoencoder. In Proceedings of the 34th International Conference on Machine Learning, Sydney, Australia, 6–11 August 2017.
- Dai, H.; Tian, Y.; Dai, B.; Skiena, S.; Song, L. Syntax-Directed Variational Autoencoder for Structured Data. *arXiv* **2018**, arXiv:1802.08786. [\[CrossRef\]](#)
- Jin, W.; Barzilay, R.; Jaakkola, T. Junction Tree Variational Autoencoder for Molecular Graph Generation. *arXiv* **2019**, arXiv:1802.04364. [\[CrossRef\]](#)
- Munson, B.P.; Chen, M.; Bogosian, A.; Kreisberg, J.F.; Licon, K.; Abagyan, R.; Kuenzi, B.M.; Ideker, T. De Novo Generation of Multi-Target Compounds Using Deep Generative Chemistry. *Nat. Commun.* **2024**, *15*, 3636. [\[CrossRef\]](#)
- Jensen, J.H. A Graph-Based Genetic Algorithm and Generative Model/Monte Carlo Tree Search for the Exploration of Chemical Space. *Chem. Sci.* **2019**, *10*, 3567–3572. [\[CrossRef\]](#)
- Steinmann, C.; Jensen, J.H. Using a Genetic Algorithm to Find Molecules with Good Docking Scores. *PeerJ Phys. Chem.* **2021**, *3*, e18. [\[CrossRef\]](#)
- Greenstein, B.L.; Elsey, D.C.; Hutchison, G.R. Determining Best Practices for Using Genetic Algorithms in Molecular Discovery. *J. Chem. Phys.* **2023**, *159*, 091501. [\[CrossRef\]](#)
- Hiener, D.C.; Hutchison, G.R. Pareto Optimization of Oligomer Polarizability and Dipole Moment Using a Genetic Algorithm. *J. Phys. Chem. A* **2022**, *126*, 2750–2760. [\[CrossRef\]](#)
- Jones, J.; Clark, R.D.; Lawless, M.S.; Miller, D.W.; Waldman, M. The AI-driven Drug Design (AIDD) Platform: An Interactive Multi-Parameter Optimization System Integrating Molecular Evolution with Physiologically Based Pharmacokinetic Simulations. *J. Comput.-Aided Mol. Des.* **2024**, *38*, 14. [\[CrossRef\]](#)
- Cavanagh, J.M.; Sun, K.; Gritsevskiy, A.; Bagni, D.; Bannister, T.D.; Head-Gordon, T. SmileyLlama: Modifying Large Language Models for Directed Chemical Space Exploration. *arXiv* **2024**, arXiv:2409.02231. [\[CrossRef\]](#)
- Wang, H.; Skreta, M.; Ser, C.T.; Gao, W.; Kong, L.; Strieth-Kalthoff, F.; Duan, C.; Zhuang, Y.; Yu, Y.; Zhu, Y.; et al. Efficient Evolutionary Search over Chemical Space with Large Language Models. *arXiv* **2024**, arXiv:2406.16976. [\[CrossRef\]](#)
- Ahmed, S.J.; Elattar, M.A. Improving Targeted Molecule Generation through Language Model Fine-Tuning Via Reinforcement Learning. *arXiv* **2024**, arXiv:2405.06836. [\[CrossRef\]](#)
- Edwards, C.; Lai, T.; Ros, K.; Honke, G.; Cho, K.; Ji, H. Translation between Molecules and Natural Language. *arXiv* **2022**, arXiv:2204.11817. [\[CrossRef\]](#)

21. 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.; et al. Deep Learning Enables Rapid Identification of Potent DDR1 Kinase Inhibitors. *Nat. Biotechnol.* **2019**, *37*, 1038–1040. [CrossRef]
22. Merk, D.; Friedrich, L.; Grisoni, F.; Schneider, G. De Novo Design of Bioactive Small Molecules by Artificial Intelligence. *Mol. Inform.* **2018**, *37*, 1700153. [CrossRef]
23. Li, Y.; Zhang, L.; Wang, Y.; Zou, J.; Yang, R.; Luo, X.; Wu, C.; Yang, W.; Tian, C.; Xu, H.; et al. Generative Deep Learning Enables the Discovery of a Potent and Selective RIPK1 Inhibitor. *Nat. Commun.* **2022**, *13*, 6891. [CrossRef]
24. Tang, X.; Dai, H.; Knight, E.; Wu, F.; Li, Y.; Li, T.; Gerstein, M. A Survey of Generative AI for De Novo Drug Design: New Frontiers in Molecule and Protein Generation. *Briefings Bioinform.* **2024**, *25*, bbae338. [CrossRef]
25. Ivanenkov, Y.; Zagribelnyy, B.; Malyshev, A.; Evteev, S.; Terentiev, V.; Kamya, P.; Bezrukov, D.; Aliper, A.; Ren, F.; Zhavoronkov, A. The Hitchhiker's Guide to Deep Learning Driven Generative Chemistry. *ACS Med. Chem. Lett.* **2023**, *14*, 901–915. [CrossRef]
26. Liu, Y.; Xu, C.; Yang, X.; Zhang, Y.; Chen, Y.; Liu, H. Application Progress of Deep Generative Models in de Novo Drug Design. *Mol. Divers.* **2024**, *28*, 2411–2427. [CrossRef]
27. Saaty, T. *The Analytic Hierarchy Process: Planning, Priority Setting, Resource Allocation*; Advanced Book Program; McGraw-Hill International Book Company: New York, NY, USA, 1980.
28. Forman, E.H.; Gass, S.I. The Analytic Hierarchy Process—An Exposition. *Oper. Res.* **2001**, *49*, 469–486. [CrossRef]
29. Figueira, J.R.; Greco, S.; Roy, B.; Słowiński, R. ELECTRE Methods: Main Features and Recent Developments. In *Handbook of Multicriteria Analysis*; Zopounidis, C., Pardalos, P.M., Eds.; Springer: Berlin/Heidelberg, Germany, 2010; pp. 51–89. [CrossRef]
30. Jacquet-Lagrange, E.; Siskos, J. Assessing a Set of Additive Utility Functions for Multicriteria Decision-Making, the UTA Method. *Eur. J. Oper. Res.* **1982**, *10*, 151–164. [CrossRef]
31. Hwang, C.L.; Yoon, K. *Multiple Attribute Decision Making; Lecture Notes in Economics and Mathematical Systems*; Springer: Berlin/Heidelberg, Germany, 1981; Volume 186. [CrossRef]
32. Yoon, K. A Reconciliation Among Discrete Compromise Solutions. *J. Oper. Res. Soc.* **1987**, *38*, 277–286. [CrossRef]
33. Opricovic, S. A Compromise Solution in Water Resources Planning. *Water Resour. Manag.* **2009**, *23*, 1549–1561. [CrossRef]
34. Duckstein, L.; Opricovic, S. Multiobjective Optimization in River Basin Development. *Water Resour. Res.* **1980**, *16*, 14–20. [CrossRef]
35. Huang, J.J.; Tzeng, G.H.; Liu, H.H. A Revised VIKOR Model for Multiple Criteria Decision Making—The Perspective of Regret Theory. In *Cutting-Edge Research Topics on Multiple Criteria Decision Making*; Shi, Y., Wang, S., Peng, Y., Li, J., Zeng, Y., Eds.; Springer: Berlin/Heidelberg, Germany, 2009; Volume 35, pp. 761–768. [CrossRef]
36. Peng, X.; Gibbs, E.; Silverman, J.M.; Cashman, N.R.; Plotkin, S.S. A Method for Systematically Ranking Therapeutic Drug Candidates Using Multiple Uncertain Screening Criteria. *Stat. Methods Med. Res.* **2021**, *30*, 1502–1522. [CrossRef]
37. Lu, J.; Wang, G.; Ying, X.; Li, Z. A Novel Drug Selection Decision Support Model Based on Real-World Medical Data by the Hybrid Entropic Weight TOPSIS Method. *Technol. Health Care* **2023**, *31*, 691–703. [CrossRef]
38. Baran-Kooiker, A.; Czech, M.; Kooiker, C. Multi-Criteria Decision Analysis (MCDA) Models in Health Technology Assessment of Orphan Drugs—A Systematic Literature Review. Next Steps in Methodology Development? *Front. Public Health* **2018**, *6*, 287. [CrossRef]
39. Angelis, A.; Kanavos, P. Multiple Criteria Decision Analysis (MCDA) for Evaluating New Medicines in Health Technology Assessment and beyond: The Advance Value Framework. *Soc. Sci. Med.* **2017**, *188*, 137–156. [CrossRef]
40. Bakan, A.; Meireles, L.M.; Bahar, I. ProDy : Protein Dynamics Inferred from Theory and Experiments. *Bioinformatics* **2011**, *27*, 1575–1577. [CrossRef]
41. Dassault Systèmes BIOVIA. *Discovery Studio Modeling Environment*, version v21.1.0.20298; Dassault Systèmes BIOVIA: Paris, France. 2024.
42. Landrum, G. RDKit: Open-Source Cheminformatics Software, version 2024.03.05. 2024. Available online: <https://github.com/rdkit/rdkit> (accessed on 19 December 2024).
43. Hennessy, E.J.; Oza, V.; Adam, A.; Byth, K.; Castriotta, L.; Grewal, G.; Hamilton, G.A.; Kamhi, V.M.; Lewis, P.; Li, D.; et al. Identification and Optimization of Benzimidazole Sulfonamides as Orally Bioavailable Sphingosine 1-Phosphate Receptor 1 Antagonists with In Vivo Activity. *J. Med. Chem.* **2015**, *58*, 7057–7075. [CrossRef]
44. Gaulton, A.; Bellis, L.J.; Bento, A.P.; Chambers, J.; Davies, M.; Hersey, A.; Light, Y.; McGlinchey, S.; Michalovich, D.; Al-Lazikani, B.; et al. ChEMBL: A Large-Scale Bioactivity Database for Drug Discovery. *Nucleic Acids Res.* **2012**, *40*, D1100–D1107. [CrossRef]
45. Hooper, D.C. Mechanisms of Action of Antimicrobials: Focus on Fluoroquinolones. *Clin. Infect. Dis.* **2001**, *32*, S9–S15. [CrossRef]
46. Drusano, G.L.; Standiford, H.C.; Plaisance, K.; Forrest, A.; Leslie, J.; Caldwell, J. Absolute Oral Bioavailability of Ciprofloxacin. *Antimicrob. Agents Chemother.* **1986**, *30*, 444–446. [CrossRef]
47. Andriole, V.T. The Quinolones: Past, Present, and Future. *Clin. Infect. Dis.* **2005**, *41*, S113–S119. [CrossRef]
48. Lehto, P.; Kivisto, K.; Neuvonen, P. The Effect of Ferrous Sulphate on the Absorption of Norfloxacin, Ciprofloxacin and Ofloxacin. *Br. J. Clin. Pharmacol.* **1994**, *37*, 82–85. [CrossRef]

49. Neuvonen, P.J.; Kivistö, K.T.; Lehto, P. Interference of Dairy Products with the Absorption of Ciprofloxacin. *Clin. Pharmacol. Ther.* **1991**, *50*, 498–502. [[CrossRef](#)]
50. Blower, T.R.; Williamson, B.H.; Kerns, R.J.; Berger, J.M. Crystal Structure and Stability of Gyrase–Fluoroquinolone Cleaved Complexes from *Mycobacterium tuberculosis*. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 1706–1713. [[CrossRef](#)]
51. Behzadian, M.; Khanmohammadi Otaghsara, S.; Yazdani, M.; Ignatius, J. A State-of-the-Art Survey of TOPSIS Applications. *Expert Syst. Appl.* **2012**, *39*, 13051–13069. [[CrossRef](#)]
52. Shekhovtsov, A.; Sałabun, W. A Comparative Case Study of the VIKOR and TOPSIS Rankings Similarity. *Procedia Comput. Sci.* **2020**, *176*, 3730–3740. [[CrossRef](#)]
53. Sari, F. Comparison of Topsis and Vikor Multi Criteria Decision Analysis Techniques. *Selcuk. Univ. J. Eng. Sci. Technol.* **2018**, *6*, 825–831. [[CrossRef](#)]

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.