



COMPARATIVE PHARMACOINFORMATIC ANALYSIS OF PHYSICOCHEMICAL PROPERTIES AND DRUG- LIKENESS IN DRD2- AND HDAC1-TARGETING COMPOUNDS

Dr. Humair Bin Md Omer^{1*}, Dr. Abu Tayab Moin², Dr. Sharika Hossain³

¹*School of Pharmacy, BRAC University, Dhaka, Bangladesh

²Department of Genetic Engineering and Biotechnology, Faculty of Biological Sciences, University of Chittagong, Chattogram, Bangladesh

³Department of Pharmacy, Jahangirnagar University, Savar, Dhaka, Bangladesh

**Corresponding Author: dr.humairbinmdomer@gmail.com*

Received: 02/02/2026 Revised: 14/04/2026 Acceptance: 17/05/2026 Published: 30/05/2026

ABSTRACT

A pharmacoinformatic approach was used to comparatively analyse the physicochemical and drug-likeness properties of the bioactive compounds that modulate the dopamine D2 receptor (DRD2) and histone deacetylase 1 (HDAC1). 1,331 compounds (1,273 DRD2-targeting and 58 HDAC1-targeting) were analysed. Descriptive statistics, Mann–Whitney U tests (with false discovery rate correction), Spearman correlation, Lipinski Rule-of-Five and principal component analysis (PCA) were used in conjunction with hierarchical clustering to analyse molecular weight, LogP, hydrogen-bond acceptors (HBA), hydrogen-bond donors (HBD), and topological polar surface area (TPSA). The differences in molecular weight, LogP, and HBA between the two target groups were not significant, while the differences in HBD and TPSA were significant. Compounds that target HDAC1 had a higher H-bond donor number and higher molecular polarity. However, overall drug-likeness rates were higher in most compounds of both classes, with higher rates in the DRD2 ligands. The first two PCs accounted for 81.5% of the total descriptor variance, and hierarchical cluster analysis indicated a partial grouping of the HDAC1 ligands according to their physicochemical properties. The overall results suggest that hydrogen-bond donor capacity and polar surface area are important physicochemical properties that separate the DRD2- and HDAC1-targeting molecules.

Keywords: DRD2; HDAC1; pharmacoinformatic; physicochemical properties; drug-likeness



1. Introduction

The dopamine D2 receptor (DRD2) has been one of the most thoroughly studied G-protein-coupled receptors, due to its central role in dopaminergic neurotransmission and its therapeutic importance in many neurological and psychiatric diseases. Structural research of dopamine receptor complexes has significantly increased the knowledge of ligand recognition, activation and downstream signalling mechanisms, which has laid the foundation for rational ligand design (Zhuang et al., 2021). Changes to the expression and signalling of DRD2 in the striatum may also impact the function of striatal circuits, motivational behaviour, and reward processes, adding to the biological relevance of this receptor as a pharmacological target (Simpson et al., 2022). DRD2 is also involved in the autoreceptor-mediated regulation of dopamine synthesis and release, and disruption of such a regulatory pathway has been associated with several neurobehavioral conditions (Chen et al., 2020).

Research in medicinal chemistry has resulted in a series of chemically distinct classes of DRD2 ligands to be developed for modulation of receptor activity, which also demonstrate enhanced affinity, selectivity and pharmacokinetic properties. Fluorescent and structurally optimised ligands have been developed to aid in the characterisation of receptors and add to a suite of tools for studying the pharmacology of dopamine receptors (Allikalt et al., 2020). Optimising ligands that target receptors is a crucial therapeutic approach to dopaminergic modulation that remains relevant to the treatment of Parkinsonian syndromes and other dopamine-signalling related disorders (Mao et al., 2020). However, this broad interest is accompanied by several considerations regarding the physicochemical properties of DRD2 ligands, as they must be designed to possess well-balanced drug-like behaviour, membrane permeability and systemic exposure to the receptor.

Another mechanistically different, but equally significant therapeutic target is histone deacetylase 1 (HDAC1). HDAC1 is a class I histone deacetylase which plays a role in chromatin remodelling, transcriptional regulation, cell-cycle control, and epigenetic modulation. Because of its role in malignant transformation and other disease processes, HDAC inhibition has become a crucial strategy in developing anticancer drugs and chemoprevention (Ruzic et al., 2022). In recent years, new efforts in HDAC-targeted drug discovery have been directed towards the development of more isoform-selective inhibitors, decreasing off-target effects and optimising the physicochemical properties of the inhibitors under consideration (Ru et al., 2023). The capacity of the HDAC family to perform diverse structural and regulatory functions underscores the need for understanding the role of molecular properties in their ability to recognise their target and their pharmacological activity (Han et al., 2024). DRD2 and HDAC1 have opposing biological roles and are valuable targets for comparison in pharmacoinformatic analysis. DRD2 is a membrane-associated receptor, for which the ligands often have to meet certain physicochemical properties, such as receptor access and, in certain therapeutic applications, central nervous system penetration properties. In contrast, HDAC1 is an intracellular zinc-dependent enzyme for which the design of ligands often includes the incorporation of polar functionalities that can interact with the enzyme in the process of recognition. The structural scaffolds of these chemical targets may not only vary greatly with respect to how they are distributed within "target-associated chemical space", but also with respect to "global molecular features", such as the size, lipophilicity, hydrogen-bonding capacity and polarity of the molecules.

The systematic study of these properties can be carried out by quantitative molecular descriptors, known as pharmacoinformatic. Molecular descriptors are numerical variables generated from the chemical structure that can be analysed statistically to find patterns of relationships between the structure and the biological target, chemical class or pharmacological property (Fernández-Torras et al., 2022). These are especially valuable for assessing chemical space as they enable large collections of chemicals to be compared by means of standardised physicochemical dimensions, and not just by a chemical structure inspection. In modern drug discovery, the analysis of chemical space has grown in significance given that bioactive compounds can reside in more than one co-overlapping region characterised by structural, chemical and biological properties (Medina-Franco et al., 2022).

Analysis of chemical space can also be used to assess if there are any similarities in chemical space



between molecules linked to different therapeutic targets or if there are significantly overlapping regions of space. The integration of molecular descriptors and multivariate and visualisation methods can enhance the interpretation of collections of compounds and enable more informed decisions in medicinal chemistry (Saldívar-González & Medina-Franco, 2022). This is especially true for comparing targets with different ligand polarity, different hydrogen-bonding behaviour and different sizes of molecules, such as DRD2 and HDAC1.

When characterising bioactive small molecules, additional information can be provided by drug-likeness assessment. Molecular weight, lipophilicity, hydrogen-bond donor and acceptor numbers are still commonly used in the early evaluation of compounds using physicochemical properties. It is important to note that these are not absolute properties that can determine the suitability of a drug for a particular type of pharmacology, though they should be considered as guiding principles (Tinworth & Young, 2020). Target engagement, solubility, permeability, metabolic behaviour, and systemic exposure are important characteristics of small-molecule drug development that are all affected by lipophilicity, making it a key variable in small-molecule optimisation (Miller et al., 2020). The donor capacity of hydrogen bonds is also a crucial parameter in medicinal chemistry related to binding to receptors, desolvation, membrane transport, and general pharmacokinetic behaviour (Kenny, 2022).

In addition to molecular descriptors, a multivariate approach can be useful for describing the complex relationship in collections of compounds. Principal component analysis can map the correlated physicochemical variables in the molecular space, making it easier to visualise the molecular chemical space and identify the main descriptive patterns. The compounds can be clustered further by hierarchical clustering based on the physicochemical similarity, and by this method it can be determined whether target-associated molecules can be clustered in one or in multiple clusters. They are broadly applicable to pharmacoinformatics workflows as they can be used to explore the data without the need for prior assumptions regarding the separation between classes.

The emergence of pharmacoinformatics has focused more than ever on the use of molecular descriptors, molecular space analysis, statistical models and visualisation as tools for better understanding of bioactive compound collections (Ramírez et al., 2024). These are especially useful if the target is not a predictive classifier but understanding the differences in the physicochemical and drug-like properties of target-associated compounds. Comparative analysis of compounds targeting DRD2 and HDAC1, therefore, could help determine if there are different profiles in the molecular properties of the two sets of ligands, which are mechanistically different.

Thus, the present study was designed to find out the difference in the physicochemical and drug-likeness properties of DRD2- and HDAC1-targeting bioactive compounds by employing a pharmacoinformatic approach. Comparative statistical testing, correlation analysis, Lipinski Rule of Five analysis, principal component analysis, and hierarchical clustering were used to analyse molecular weight, LogP, hydrogen-bond acceptors and donors, and topological polar surface area. The study aimed to characterise the target-associated molecular property patterns and estimate the degree of physicochemical space overlap between DRD2 and HDAC1 ligands.

2. Materials and Methods

2.1 Study Design and Data Source

The secondary-data pharmacoinformatic design approach was used to identify and evaluate the physicochemical and drug-likeness properties of the bioactive molecules of the dopamine D2 receptor (DRD2) and histone deacetylase 1 (HDAC1). The analytical dataset comprised data on compound identifiers, Simplified Molecular Input Line Entry System (SMILES) representations, protein-target annotations, as well as physicochemical descriptors: molecular weight, octanol–water partition coefficient (LogP), hydrogen-bond acceptors (HBA), hydrogen-bond donors (HBD), and topological polar surface area (TPSA). The analyses were developed to describe differences between the targets, relationships between the molecular descriptors, conventional drug-likeness and multidimensional physicochemical space and natural molecular groupings. Each of the analytical components is the same as the comparative analysis, correlation analysis, Lipinski analysis, principal component analysis



and hierarchical clustering analysis reported.

2.2 Compound Selection and Eligibility Criteria

Any compound targeted to either DRD2 or HDAC1 was deemed suitable for the main comparative analysis. The records must include a valid protein-target annotation and have enough physicochemical data to be analysed in the record's analysis. Compounds that were assigned to BACE1 were not formally compared, as there was not enough representation to support reliable statistical inferences between targets. Only those analyses which required missing data for specific descriptors excluded records with missing data, not all analyses. This was done on a variable-by-variable basis to retain as much information as possible at the molecular level.

2.3 Data Cleaning and Preprocessing

All preprocessing was done in Python. The data was checked for multiple compound identifiers, multiple molecular representations, missing data, data of little value, and implausible physicochemical data. To remove variables not related to either chemical or biological information, variables were automatically generated and then removed prior to analysis. The ChEMBL identifiers remained as the main identifiers for the compounds, and the SMILES strings were also kept representing the molecules. Prior to statistical analysis, the data for molecular weight, LogP, HBA, HBD, and TPSA were converted to a numeric format appropriate to that variable and checked for completeness. Protein target was considered as the main categorical variable.

2.4 Descriptive Physicochemical Profiling

Each of the physicochemical properties of DRD2- and HDAC1-targeting compounds was summarised separately. The molecular weight, LogP, HBA, HBD and TPSA were determined by suitable measures of central tendency and dispersion such as mean, standard deviation, median, interquartile range, minimum and maximum. Histograms, boxplots, Q-Q plots, skewness and kurtosis were used to examine distributional characteristics. Due to the anticipated skewed distributions of some molecular descriptors and the presence of outlying values, which are typical for chemical data, a focus was placed on median-based summaries.

2.5 Comparative Statistical Analysis

The nonparametric comparative procedures were used to evaluate the differences in molecular weight, LogP, HBA, and HBD between targets. Nonparametric comparative procedures were used to evaluate differences between targets in molecular weight, LogP, HBA, and HBD. To compare DRD2-targeting compounds with HDAC1-targeting compounds, the Mann–Whitney U test was used as the main inferential test, given the lack of distributional assumptions that are necessary for conventional parametric testing. Two-sided statistical tests were used.

Multiple physicochemical descriptors were tested at the same time, and p-values were further adjusted with a false discovery rate (FDR) correction to minimise the risk of false positive results due to multiple tests. The adjusted probability values led to the decision to use 0.05 as the alpha level for statistical significance.

2.6 Effect-Size Estimation

To provide insight into the strength of the physicochemical differences between the two protein-target groups, a statistical significance assessment was supplemented by an effect size estimation for this purpose. Mann–Whitney comparisons were used to calculate rank-biserial correlation. Effect sizes were judged both by magnitude and direction, enabling the practical significance of descriptor differences from statistical significance to be differentiated.

2.7 Correlation Analysis

Spearman's rank-order correlation coefficient (ρ) was used to investigate the correlation between



molecular weight, LogP, HBA, HBD, and TPSA. Spearman correlation was chosen as it does not require multivariate normality and can be used for continuous or ordinal molecular variables that are skewed. Pairwise correlations were computed for each set of the physicochemical descriptors. The result was the direction and size of the correlation coefficient and a diverging correlation heatmap that represented the complete correlation matrix. The analysis was designed to help establish patterns of association between size, lipophilicity, hydrogen-bonding capability, and molecular polarity, which correlates with the correlation analysis.

2.8 Drug-Likeness Assessment

The available criteria derived from Lipinski's Rule of Five were used for the evaluation of drug-likeness. The four physicochemical thresholds for each compound were: molecular weight ≤ 500 Da, $\log P \leq 5$, no more than five hydrogen-bond donors, and no more than ten hydrogen-bond acceptors. A compound was deemed to meet the evaluated Rule-of-Five profile if all four available criteria were met at the same time.

Compliance was computed for each criterion and overall, separately for DRD2- and HDAC1-targeting compounds. TPSA was not added as an official Lipinski Rule but was added as an extra molecular polarity and potential permeability criterion. The number of compounds that meet the individual and combined criteria was summarised numerically and displayed on grouped bar charts.

2.9 Principal Component Analysis

A Principal component analysis (PCA) was carried out to explore the multidimensional physicochemical space defined by the molecular weight, LogP, HBA, HBD, and TPSA. All the variables were z-score transformed before PCA due to different scales on which the descriptors were measured.

The standardised descriptor matrix was subjected to PCA, and the eigenvalues, explained proportions of variance and loadings of the descriptors for each component were analysed. The first two principal components were used for primary graphical interpretation as they accounted for the two-dimensional representation of the major variance structure. PCA scores were plotted by protein-target annotation to see how much overlap or separation exists between the DRD2- and HDAC1-targeting molecules. To find out which physicochemical variables had the greatest contribution to the principal components, the loadings of the descriptors were used. PCA was not used as a classification procedure but as an exploratory dimensionality reduction technique.

2.10 Hierarchical Cluster Analysis

Hierarchical clustering was performed to find molecular groupings that occur in nature based on their physicochemical similarity. The same molecular descriptors, as used for the PCA, were employed for the clustering procedure. Dissimilarity between compounds was calculated by Euclidean distance, and successive cluster formation was carried out using Ward's linkage method to minimise the variance within a cluster.

This hierarchical structure was then visualised using a dendrogram. A three-cluster solution was used to explore the clusters for possible interpretation; compounds targeting DRD2 and HDAC1 were then characterised in each of the three clusters. Due to the fact that the target classes were rather unbalanced in size, the clustering was considered to be an exploratory physicochemical grouping procedure rather than a proof of predictive target classification. The clustering analysis of the manuscript was also done with the use of standardised descriptors and Ward linkage.

2.11 Data Visualisation

For the statistical analyses and to aid in the interpretation of the molecular property patterns, graphical outputs were produced. Distributions of molecular weight, LogP, HBA, HBD and TPSA were shown using comparative boxplots, presented for the two groups of protein targets. The Spearman correlation matrix was used to create the correlation heatmap. The results for compliance with single drug-



likeness criteria and for the combined drug-likeness criteria were presented in grouped bar charts. PCA score plots were used for visualising multivariate structure and a Ward-linkage dendrogram for visualising the hierarchical relationships among compounds. All figures were presented with descriptive and user-friendly axes, consistent axis labels and publication style.

2.12 Statistical and Computational Software

In Python, all data preprocessing, statistical analyses, pharmacoinformatic calculations, multivariate procedures and graphical visualisation were carried out. Data manipulation was performed using pandas and NumPy, statistical testing and correlation analysis were done with SciPy, and related statistical libraries and multivariate analysis were done with scikit-learn and SciPy clustering functions. For graphical output, Matplotlib and compatible packages for visualisation in Python were used. All inferential analyses were performed on a 2-tailed basis and statistical significance was considered $p < 0.05$; false discovery rate adjustment was performed for multiple physicochemical comparisons.

2.13 Handling of Target-Class Imbalance

Analysis considers the unequal presence of DRD2- and HDAC1-targeting compounds. No oversampling, undersampling or artificial balancing process was done in the study, as it was to describe the observed distributions of the physicochemical properties rather than to build a supervised classification model. Nonparametric statistical procedures and appropriate effect-size measures were thus used when the groups were unequal in number. PCA and hierarchical clustering were not used to determine classification accuracy but rather descriptively/exploratorily.

2.14 Ethical Considerations

This study was purely a secondary molecular study and did not require the recruitment of humans, collection of identifiable personal data, experimental interventions or the use of live animals. Therefore, no institutional ethical approval and informed consent was needed for the computational secondary analysis.

3. Results

3.1 Dataset Composition and Physicochemical Characteristics

The final analysis-ready data included 1,331 bioactive compounds, of which 1,273 compounds were assigned to the dopamine D2 receptor (DRD2) and 58 compounds were assigned to histone deacetylase 1 (HDAC1). The physicochemical properties of the two target groups were determined in terms of molecular weight, LogP value, hydrogen-bond acceptors (HBA), hydrogen-bond donors (HBD) and topological polar surface area (TPSA). Table 1 summarises the molecular weight of the compounds targeted at DRD2 and HDAC1, with a median MW of 399.51 Da for the DRD2-targeting compounds and 419.98 Da for the HDAC1-targeting compounds. The median LogP values were also comparable for DRD2 and HDAC1 molecules, 3.56 and 3.72, respectively. The median value for the number of hydrogen-bond acceptors was the same for both groups: five.

However, significant differences were found for HBD and TPSA. The median HBD number for DRD2-targeting compounds was 1.00, while it was 3.00 for HDAC1-targeting compounds. Likewise, the median TPSA was 51.96 Å² for DRD2 compounds, and 92.39 Å² for HDAC1 compounds. The results of these analyses suggest that HDAC1-targeting molecules were more polar and had a higher hydrogen-bond donation potential than DRD2-targeting molecules.



Table 1. Descriptive physicochemical characteristics of DRD2- and HDAC1-targeting compounds

Physicochemical descriptor	DRD2 median	HDAC1 median	Observed pattern
Molecular weight (Da)	399.51	419.98	Similar distributions
LogP	3.56	3.72	Similar distributions
H-bond acceptors	5.00	5.00	Similar distributions
H-bond donors	1.00	3.00	Higher in HDAC1
TPSA ($\text{\AA}^2$)	51.96	92.39	Substantially higher in HDAC1

The distributional differences between the two protein-target groups are visually presented in Figure 1.

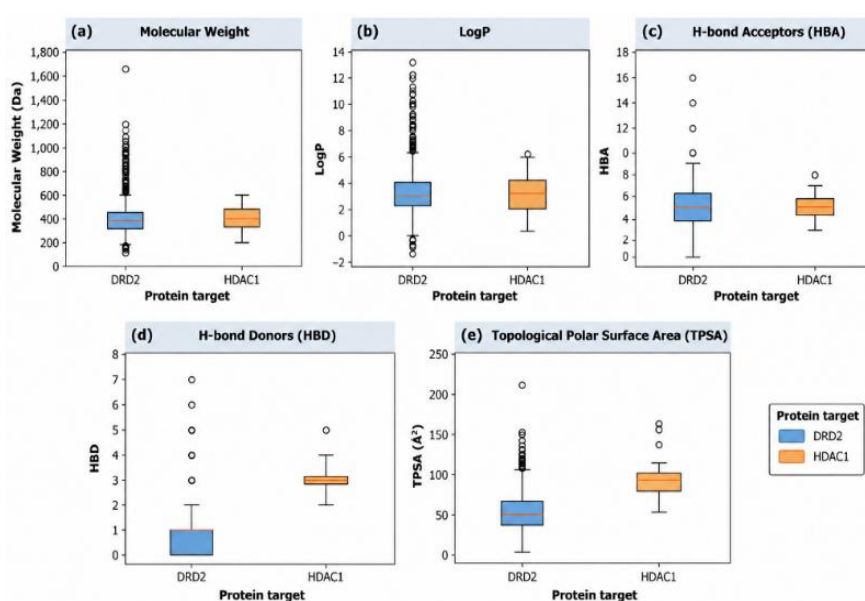


Figure 1. Comparative distributions of molecular weight, LogP, hydrogen-bond acceptors, hydrogen-bond donors, and topological polar surface area in DRD2- and HDAC1-targeting compounds

3.2 Comparative Statistical Analysis of DRD2 and HDAC1 Ligands

Because the physicochemical variables showed non-normal distributions in at least one of the target groups, Mann-Whitney U tests were used as the primary inferential procedure. To account for multiple comparisons, false discovery rate-adjusted p-values were also calculated. The results are reported in Table 2.

No statistically significant difference was detected between DRD2 and HDAC1 compounds for molecular weight (adjusted $p = 0.588$), LogP (adjusted $p = 0.989$), or HBA (adjusted $p = 0.989$). The associated rank-biserial effect sizes were also small, indicating minimal between-group separation for these descriptors.

In contrast, very big differences were identified for HBD and TPSA. HBD differed significantly between the two target groups ($p = 3.17 \times 10^{-30}$; adjusted $p = 1.58 \times 10^{-29}$), with a rank-biserial correlation of -0.835 . TPSA also demonstrated a highly significant between-group difference ($p = 2.10 \times 10^{-23}$; adjusted $p = 5.24 \times 10^{-23}$), with a rank-biserial correlation of -0.773 . The magnitude of these effect sizes indicates that HBD and TPSA were the most strongly differentiated physicochemical properties between DRD2- and HDAC1-targeting compounds.



Table 2. Mann-Whitney U comparison of physicochemical descriptors between DRD2- and HDAC1-targeting compounds

Descriptor	p-value	FDR-adjusted p-value	Rank-biserial correlation	Interpretation
Molecular weight	0.353	0.588	-0.072	Not significant
LogP	0.989	0.989	0.001	Not significant
H-bond acceptors	0.959	0.989	0.004	Not significant
H-bond donors	3.17×10^{-30}	1.58×10^{-29}	-0.835	Very large difference
TPSA	2.10×10^{-23}	5.24×10^{-23}	-0.773	Very large difference

These results support the interpretation that DRD2 and HDAC1 ligand sets differ primarily in hydrogen-bond donation and molecular polarity rather than in molecular size, overall lipophilicity, or hydrogen-bond accepting capacity.

3.3 Correlation Structure Among Molecular Descriptors

Spearman correlation analysis was performed to characterise relationships among the physicochemical descriptors. As shown in Table 3, the strongest positive association was observed between HBD and TPSA ($\rho = 0.660$), followed by HBA and TPSA ($\rho = 0.642$). These associations indicate that increased hydrogen-bonding capacity was generally accompanied by increased molecular polar surface area. Molecular weight showed a moderate positive correlation with LogP ($\rho = 0.537$) and HBA ($\rho = 0.474$), while its relationship with TPSA was somewhat weaker but still positive ($\rho = 0.403$). By contrast, LogP exhibited a modest inverse association with TPSA ($\rho = -0.217$), suggesting that compounds with greater polar surface area tended to display lower lipophilicity.

Table 3. Selected Spearman correlations among physicochemical descriptors

Descriptor pair	Spearman ρ	Direction
HBD – TPSA	0.660	Strong positive
HBA – TPSA	0.642	Strong positive
Molecular weight – LogP	0.537	Moderate positive
Molecular weight – HBA	0.474	Moderate positive
Molecular weight – TPSA	0.403	Moderate positive
LogP – TPSA	-0.217	Weak-to-moderate negative

The overall correlation structure is displayed in Figure 2. The heatmap highlights the clustering of polarity-related variables, particularly TPSA, HBD, and HBA, while LogP exhibits a distinct inverse relationship with TPSA.

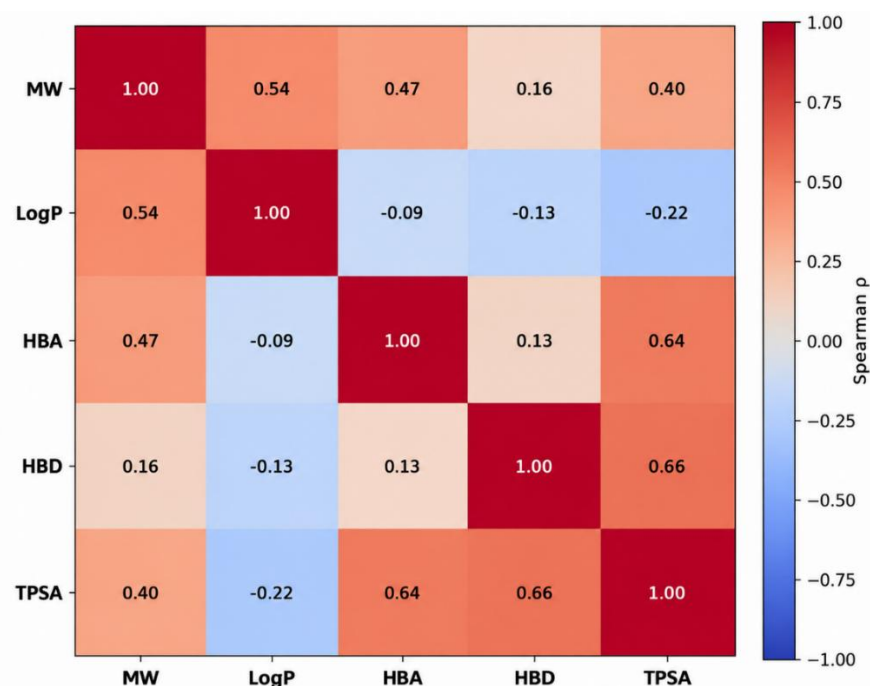


Figure 2. Spearman correlation heatmap showing associations among molecular weight, LogP, hydrogen-bond acceptors, hydrogen-bond donors, and topological polar surface area

3.4 Drug-Likeness Assessment

Drug-likeness was evaluated using available Lipinski Rule-of-Five criteria. As presented in Table 4, 83.69% of complete DRD2 compounds satisfied all four evaluated criteria, compared with 72.41% of HDAC1 compounds. DRD2-targeting compounds therefore showed a somewhat higher overall proportion of conventional drug-like profiles.

For individual criteria, 89.23% of DRD2 compounds and 72.41% of HDAC1 compounds had molecular weights not exceeding 500 Da. LogP values of 5 or lower were observed in 88.99% of DRD2 compounds and 81.03% of HDAC1 compounds. Compliance with HBD and HBA criteria was very high in both groups, with more than 99% of DRD2 compounds and all HDAC1 compounds satisfying the respective limits.

Table 4. Lipinski Rule-of-Five compliance among DRD2- and HDAC1-targeting compounds

Drug-likeness criterion	DRD2 (%)	HDAC1 (%)
Molecular weight $\leq$ 500 Da	89.23	72.41
LogP $\leq$ 5	88.99	81.03
HBD $\leq$ 5	99.76	100.00
HBA $\leq$ 10	99.76	100.00
All evaluated criteria satisfied	83.69	72.41

The comparative compliance pattern is shown in Figure 3, demonstrating that both target groups predominantly occupied conventional drug-like physicochemical space, although HDAC1 compounds showed a relatively lower overall proportion satisfying all criteria simultaneously.

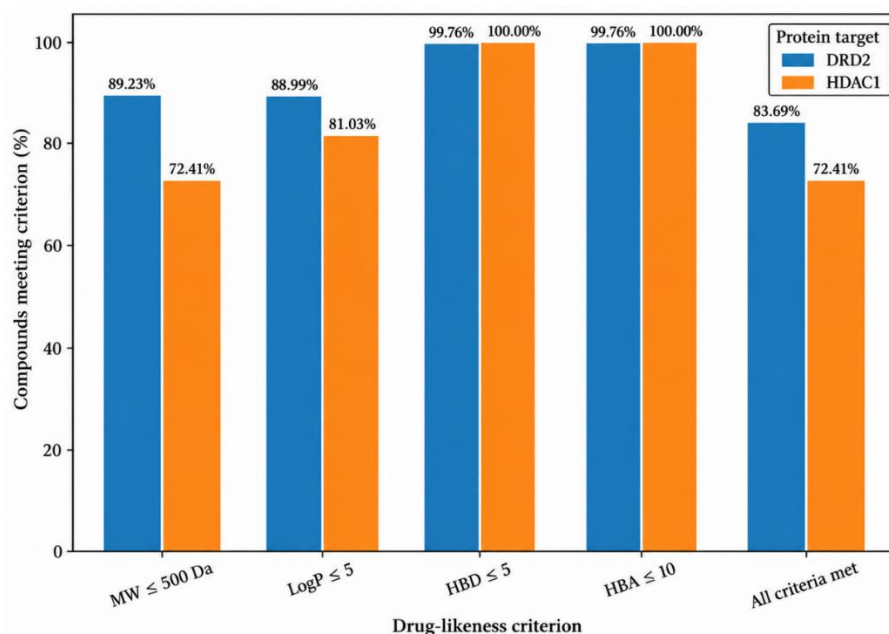


Figure 3. Comparative Lipinski Rule-of-Five compliance among DRD2- and HDAC1-targeting compounds

3.5 Principal Component Analysis of Physicochemical Space

Principal component analysis was conducted using standardised values of molecular weight, LogP, HBA, HBD, and TPSA. The first principal component accounted for 53.0% of the total variance, while the second principal component explained an additional 28.5%. Together, PC1 and PC2 captured 81.5% of the total physicochemical variance in the dataset.

As shown in Table 5, PC1 was driven most strongly by TPSA, molecular weight, HBA, and HBD, whereas PC2 was dominated by LogP and molecular weight, with negative contributions from HBD and TPSA. These loading patterns suggest that PC1 primarily represented a combined size-polarity-hydrogen-bonding dimension, whereas PC2 reflected a contrast between lipophilicity and molecular polarity.

Table 5. Principal component loadings for physicochemical descriptors

Descriptor	PC1 loading	PC2 loading
Molecular weight	0.503	0.427
LogP	—	0.728
H-bond acceptors	0.492	—
H-bond donors	0.400	-0.412
TPSA	0.542	-0.338

3.6 Hierarchical Cluster Analysis

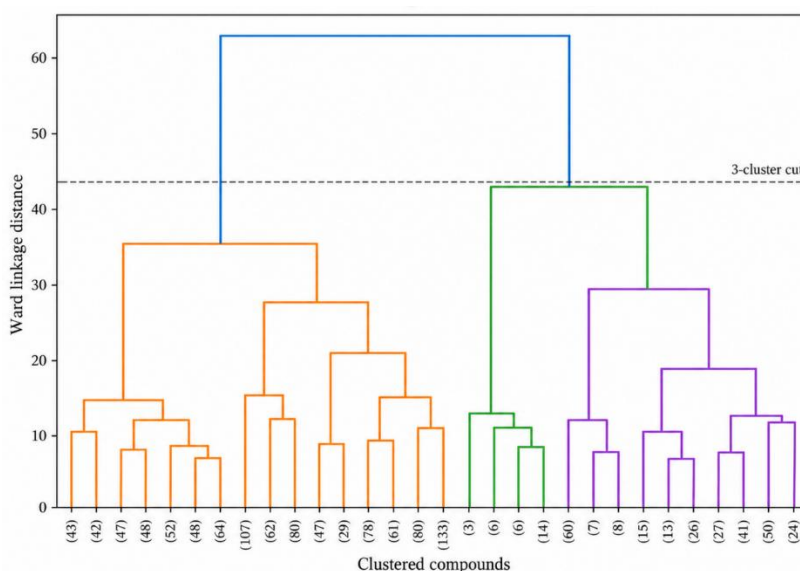
Hierarchical clustering was performed using standardised physicochemical descriptors and Ward's linkage method. The analysis identified three principal exploratory clusters. As summarised in Table 6, Cluster 1 contained 1,021 compounds, including 1,009 DRD2 compounds and 12 HDAC1 compounds. Cluster 2 consisted entirely of 29 DRD2 compounds. Cluster 3 contained 271 compounds, comprising 225 DRD2 compounds and 46 HDAC1 compounds.

Notably, 46 of the 58 HDAC1-targeting compounds were in Cluster 3. This concentration indicates that a substantial proportion of HDAC1 ligands shared a related physicochemical profile that differed from the dominant DRD2-containing cluster. However, because of the marked imbalance between the two target groups, these clustering results were interpreted as exploratory rather than as evidence of predictive target classification.

**Table 6. Distribution of DRD2- and HDAC1-targeting compounds across hierarchical clusters**

Cluster	DRD2	HDAC1	Total
Cluster 1	1,009	12	1,021
Cluster 2	29	0	29
Cluster 3	225	46	271

The hierarchical structure of the compounds is illustrated in Figure 4. The dendrogram demonstrates the presence of multiple physicochemical groupings, with a substantial proportion of HDAC1 ligands concentrated within one major cluster while most DRD2 compounds remained distributed within the largest cluster.

**Figure 4. Ward's hierarchical clustering dendrogram of DRD2- and HDAC1-targeting compounds based on standardised physicochemical descriptors**

3.7 Overall Physicochemical Differentiation Between Target Groups

Collectively, the analyses showed that DRD2- and HDAC1-targeting compounds shared broadly comparable molecular weight, LogP, and hydrogen-bond acceptor profiles but differed substantially in hydrogen-bond donor capacity and topological polar surface area. The large effect sizes associated with HBD and TPSA were consistent with the observed multivariate patterns. Correlation analysis further demonstrated that HBD and HBA were strongly associated with TPSA, reinforcing the importance of molecular polarity and hydrogen-bonding characteristics within the dataset.

The drug-likeness analysis showed that most compounds in both target groups remained within conventional Lipinski-compatible physicochemical space. PCA explained 81.5% of total descriptor variance within the first two dimensions and revealed partial, but not complete, differentiation between DRD2 and HDAC1 ligands. Hierarchical clustering similarly identified a physicochemical subgroup containing the majority of HDAC1 compounds. Overall, the findings indicate that hydrogen-bond donor capacity and molecular polar surface area represent the principal physicochemical features distinguishing HDAC1-targeting compounds from DRD2-targeting compounds in the present dataset.

4. Discussion

The present pharmacoinformatic analysis revealed that there is a shared, but distinct, physicochemical space for DRD2- and HDAC1-targeting compounds. The most significant differences between targets were the hydrogen-bond donor capacity, and the topological polar surface area, while the molecular weight, LogP and the number of hydrogen-bond acceptors showed a similar profile. It appears that there is a polarity & hydrogen-bond donation effect in mind for distinguishing the two ligand groups. The target-dependent variation is in line with the general opinion that drug-like properties should be



read in the context of the biological target and not on a single universal optimum (Leeson et al., 2021). Chemically, the higher polarity of HDAC1 targeting compounds is plausible since many HDAC inhibitors feature polar groups that bind to the zinc in HDAC1 and are attached to surface-recognition groups. Target binding activity, however, must be coupled with tolerable physicochemical properties when developing class I HDAC inhibitors (Schäker-Hübner et al., 2022). The higher HBD and TPSA values found here are also in line with medicinal-chemistry principles that focus on improving zinc-binding groups, linker regions, and cap structures to attain potency and selectivity (He et al., 2022).

There is no significant difference in the molecular weight or overall Lipophilicity (LogP) which indicates that HDAC1 and DRD2 ligands cannot be separated on the basis of molecular size alone or overall Lipophilicity (LogP). Rather, a discrimination seems to be exhibited by differences in polar functionality. Various HDAC1 inhibitors have been reported that exhibit broad structural diversity, with the arrangement of the functional groups that are crucial for showing activity (Patel et al., 2022). This is also in line with the current design strategies for HDAC inhibitors, where a pharmacophoric framework is not significantly altered in structurally different compounds (Daško et al., 2022).

This was confirmed by a correlation analysis. A positive correlation between HBD and HBA with TPSA was observed, which suggested that there is a correlation between the number of hydrogen-bonding functionalities and the molecular polarity. The rational development of HDAC inhibitors requires tuning these characteristics without compromising the necessary permeability and drug-like behaviour (Qiu et al., 2021). The small negative correlation between LogP and TPSA found here also points to the desired compromise between the lipophilicity and the polarity in the small molecule optimisation.

There was a higher number of compounds that meet the evaluated Lipinski criteria in both target groups, but less in the HDAC1 group. This should not be equated with reduced drug quality from a pharmacological perspective as Rule-of-Five descriptors are not definitive measures of drug suitability, but rather a screening heuristic. Drug-likeness can be defined as a multi-dimensional property space as opposed to a binary classification (Lee et al., 2022). Optimisation of the physicochemical properties can also enhance the drug-likeness without compromising the biological activity when done intentionally (Varela et al., 2022).

The ligands of the DRD2 receptor exhibited lower HBD and TPSA values, but had similar LogP values. These properties could apply to aminergic G-protein coupled receptor ligands, for which receptor affinity, membrane accessibility and central nervous system exposure must be carefully controlled by maintaining the control of polarity. Many reports have been published about structural optimisation of dopamine D2 receptor ligands for various neuropsychiatric applications (Juza et al., 2023). Small structural variations have also been demonstrated to significantly affect the recognition and selectivity of dopamine receptors in experiments and calculations (Zell et al., 2022).

The first two principal components of principal component analysis removed the majority of variance in the descriptors and there was a significant overlap between the DRD2 and HDAC1 compounds. So, the chosen global descriptors are meaningful physicochemical differences but cannot be used to separate the target classes. Dopamine receptor ligand selectivity has also been found to be dependent on the subtleties in molecular recognition that go beyond just bulk descriptors (Zell et al., 2023). This was confirmed by hierarchical clustering as a significant amount of the HDAC1 ligands clustered together, and there was still some overlap with DRD2 compounds. Structure-activity analyses of D2/D3 ligands also reveal that targeting of the receptor needs to rely heavily on the scaffold and on the arrangement of substituents (Shaik et al., 2021).

The value of pharmacoinformatics in the characterisation of target-associated chemical space can be shown through the combined use of the above-mentioned methods, based on descriptive statistics, correlation analysis, drug-likeness assessment, PCA and clustering. ChEMBL would be suitable for this use case as it allows for the ability to combine the standardised structure of compounds, target annotation, and bioactivity data from big collections (Zdravil et al., 2024). Advanced analyses, such as quantitative potency, fingerprinting, scaffold decomposition and target-specific bioactivity modelling, could be included in more advanced analyses. Structural descriptors in addition to computational



classification has already been demonstrated to offer additional insights into HDAC1 structure-activity relationships (Li et al., 2023).

A few drawbacks should be considered. A significant imbalance in the sample sizes of DRD2 and HDAC1 precludes direct comparisons, and experimental potency, selectivity, pharmacokinetic, and toxicity endpoints were not analysed. The physicochemical differences, therefore, are not necessarily indicative of improved binding or therapeutic effects. Furthermore, the information of the molecule relevant for target recognition is only partially captured by five descriptors. An integrated strategy of chemical design and kinetic and biological assays is useful for discovering selective HDAC1/2 inhibitors (Itoh et al., 2023). The incorporation of more molecular descriptors with experimental activity and selectivity should therefore be considered in future studies. These restrictions are relevant for the interpretation of the patterns of variation found in the exploratory multivariate analysis, especially about the composition of the clusters; for this reason, the results might be affected by unequal target representation. However, the convergence between univariate and multivariate analyses supports the observation that there is a difference between the two sets of ligands with respect to their polarity.

5. Conclusion

The present study showed that the bioactive compounds targeting DRD2 and HDAC have some common features among them in terms of physicochemical properties, although they are significantly different in terms of molecular polarity and hydrogen-bond donor capacity. The molecular weight, LogP and number of hydrogen-bond acceptors were relatively similar for the two sets of ligands, while hydrogen-bond donors and topological polar surface area were the most unique descriptors. These differences indicate that polarity-related properties of molecules could be among the most critical ones defining the physicochemical space of HDAC1 ligands in comparison to DRD2 targeting molecules. Most compounds in both groups met the evaluated Lipinski Rule-of-Five criteria, with a higher percentage of the DRD2 compounds meeting all the criteria at once. The results of the correlation analysis also confirmed the strong association between TPSA and hydrogen-bonding descriptors, indicating the relevance of the polarity-related descriptors. Most of the total variation of the descriptors was found in the two main dimensions of the PCA, with partial but not complete separation between the target groups. In addition, the hierarchical clustering identified another physicochemical subgroup, which was also enriched with HDAC1 ligands. The study illustrates overall the usefulness of pharmacoinformatic analysis for comparisons of chemical spaces related to targets. Future research should integrate quantitative bioactivity, structural fingerprint, scaffold information, selectivity data and other pharmacokinetic descriptors to gain a more complete understanding of molecular factors that govern DRD2 and HDAC1 ligand behaviour.

References

1. Leeson, P. D., Bento, A. P., Gaulton, A., Hersey, A., Manners, E. J., Radoux, C. J., & Leach, A. R. (2021). Target-based evaluation of “drug-like” properties and ligand efficiencies. *Journal of Medicinal Chemistry*, 64(11), 7210–7230.
2. Zdrazil, B., Felix, E., Hunter, F., Manners, E. J., Blackshaw, J., Corbett, S., de Veij, M., Ioannidis, H., Mendez Lopez, D., Mosquera, J. F., Magarinos, M. P., Bosc, N., Arcila, R., Kizilören, T., Gaulton, A., Bento, A. P., Adasme, M. F., Monecke, P., Landrum, G. A., & Leach, A. R. (2024). The ChEMBL database in 2023: A drug discovery platform spanning multiple bioactivity data types and time periods. *Nucleic Acids Research*, 52(D1), D1180–D1192.
3. Lee, K., Jang, J., Seo, S., Lim, J., & Kim, W. Y. (2022). Drug-likeness scoring based on unsupervised learning. *Chemical Science*, 13(2), 554–565.
4. Varela, M. T., Amaral, M., Romanelli, M. M., Levatti, E. V. de C., Tempone, A. G., & Fernandes, J. P. S. (2022). Optimization of physicochemical properties is a strategy to improve drug-likeness associated with activity: Novel active and selective compounds against *Trypanosoma cruzi*. *European Journal of Pharmaceutical Sciences*, 171, 106114.



5. Schäker-Hübner, L., Haschemi, R., Büch, T., Kraft, F. B., Brumme, B., Schöler, A., Jenke, R., Meiler, J., Aigner, A., Bendas, G., & Hansen, F. K. (2022). Balancing histone deacetylase (HDAC) inhibition and drug-likeness: Biological and physicochemical evaluation of class I selective HDAC inhibitors. *ChemMedChem*, 17(9), e202100755.
6. He, X., Hui, Z., Xu, L., Bai, R., Gao, Y., Wang, Z., Xie, T., & Ye, X.-Y. (2022). Medicinal chemistry updates of novel HDACs inhibitors (2020 to present). *European Journal of Medicinal Chemistry*, 227, 113946.
7. Patel, P., Wahan, S. K., Vishakha, S., Kurmi, B. D., Gupta, G. D., Rajak, H., & Asati, V. (2022). Recent progress in histone deacetylase (HDAC) 1 inhibitors as anticancer agent. *Current Cancer Drug Targets*, 23(1), 47–70.
8. Daško, M., de Pascual-Teresa, B., Ortín, I., & Ramos, A. (2022). HDAC inhibitors: Innovative strategies for their design and applications. *Molecules*, 27(3), 715.
9. Qiu, X., Zhu, L., Wang, H., Tan, Y., Yang, Z., Yang, L., & Wan, L. (2021). From natural products to HDAC inhibitors: An overview of drug discovery and design strategy. *Bioorganic & Medicinal Chemistry*, 52, 116510.
10. Li, R., Tian, Y., Yang, Z., Ji, Y., Ding, J., & Yan, A. (2023). Classification models and SAR analysis on HDAC1 inhibitors using machine learning methods. *Molecular Diversity*, 27(3), 1037–1051.
11. Itoh, Y., Zhan, P., Tojo, T., Jaikhan, P., Ota, Y., Suzuki, M., Li, Y., Hui, Z., Moriyama, Y., Takada, Y., Yamashita, Y., Oba, M., Uchida, S., Masuda, M., Ito, S., Sowa, Y., Sakai, T., & Suzuki, T. (2023). Discovery of selective histone deacetylase 1 and 2 inhibitors: Screening of a focused library constructed by click chemistry, kinetic binding analysis, and biological evaluation. *Journal of Medicinal Chemistry*, 66(22), 15171–15188.
12. Juza, R., Musilek, K., Mezeiova, E., Soukup, O., & Korabecny, J. (2023). Recent advances in dopamine D2 receptor ligands in the treatment of neuropsychiatric disorders. *Medicinal Research Reviews*, 43(1), 55–211.
13. Zell, L., Lainer, C., Kollár, J., Temml, V., & Schuster, D. (2022). Identification of novel dopamine D2 receptor ligands—A combined in silico/in vitro approach. *Molecules*, 27(14), 4435.
14. Zell, L., Bretl, A., Temml, V., & Schuster, D. (2023). Dopamine receptor ligand selectivity—An in silico/in vitro insight. *Biomedicines*, 11(5), 1468.
15. Shaik, A. B., Boateng, C. A., Battiti, F. O., Bonifazi, A., Cao, J., Chen, L., Chitsazi, R., Ravi, S., Lee, K. H., Shi, L., & Newman, A. H. (2021). Structure activity relationships for a series of eticlopride-based dopamine D2/D3 receptor bitopic ligands. *Journal of Medicinal Chemistry*, 64(20), 15313–15333.
16. Zhuang, Y., Xu, P., Mao, C., Wang, L., Krumm, B., Zhou, X. E., Huang, S., Liu, H., Cheng, X., Huang, X.-P., Shen, D.-D., Xu, T., Liu, Y.-F., Wang, Y., Guo, J., Jiang, Y., Jiang, H., Melcher, K., Roth, B. L., & Xu, H. E. (2021). Structural insights into the human D1 and D2 dopamine receptor signaling complexes. *Cell*, 184(4), 931–942.e18.
17. Simpson, E. H., Gallo, E. F., Balsam, P. D., Javitch, J. A., & Kellendonk, C. (2022). How changes in dopamine D2 receptor levels alter striatal circuit function and motivation. *Molecular Psychiatry*, 27(1), 436–444.
18. Chen, R., Ferris, M. J., & Wang, S. (2020). Dopamine D2 autoreceptor interactome: Targeting the receptor complex as a strategy for treatment of substance use disorder. *Pharmacology & Therapeutics*, 213, 107583.
19. Allikalt, A., Purkayastha, N., Flad, K., Schmidt, M. F., Tabor, A., Gmeiner, P., Hübner, H., & Weikert, D. (2020). Fluorescent ligands for dopamine D2/D3 receptors. *Scientific Reports*, 10, 21842.
20. Mao, Q., Qin, W.-Z., Zhang, A., & Ye, N. (2020). Recent advances in dopaminergic strategies for the treatment of Parkinson's disease. *Acta Pharmacologica Sinica*, 41, 471–482.
21. Ruzic, D., Djoković, N., Srdić-Rajić, T., Echeverria, C., Nikolic, K., & Santibanez, J. F. (2022).



- Targeting histone deacetylases: Opportunities for cancer treatment and chemoprevention. *Pharmaceutics*, 14(1), 209.
22. Ru, J., Wang, Y., Li, Z., Wang, J., Ren, C., & Zhang, J. (2023). Technologies of targeting histone deacetylase in drug discovery: Current progress and emerging prospects. *European Journal of Medicinal Chemistry*, 261, 115800.
 23. Han, H., Feng, X., He, T., Wu, Y., He, T., Yue, Z., & Zhou, W. (2024). Discussion on structure classification and regulation function of histone deacetylase and their inhibitor. *Chemical Biology & Drug Design*, 103, e14366.
 24. Fernández-Torras, A., Comajuncosa-Creus, A., Duran-Frigola, M., & Aloy, P. (2022). Connecting chemistry and biology through molecular descriptors. *Current Opinion in Chemical Biology*, 66, 102090.
 25. Medina-Franco, J. L., Chávez-Hernández, A. L., López-López, E., & Saldívar-González, F. I. (2022). Chemical multiverse: An expanded view of chemical space. *Molecular Informatics*, 41(11), e2200116.
 26. Saldívar-González, F. I., & Medina-Franco, J. L. (2022). Approaches for enhancing the analysis of chemical space for drug discovery. *Expert Opinion on Drug Discovery*, 17(7), 789–798.
 27. Tinworth, C. P., & Young, R. J. (2020). Facts, patterns, and principles in drug discovery: Appraising the Rule of 5 with measured physicochemical data. *Journal of Medicinal Chemistry*, 63(18), 10091–10108.
 28. Miller, R. R., Madeira, M., Wood, H. B., Geissler, W. M., Raab, C. E., & Martin, I. J. (2020). Integrating the impact of lipophilicity on potency and pharmacokinetic parameters enables the use of diverse chemical space during small molecule drug optimization. *Journal of Medicinal Chemistry*, 63(21), 12156–12170.
 29. Kenny, P. W. (2022). Hydrogen-bond donors in drug design. *Journal of Medicinal Chemistry*, 65(21), 14261–14275.
 30. Ramírez, D., Peláez, R., Lagos, C. F., & Medina-Franco, J. L. (2024). Editorial: Pharmacoinformatics: New developments and challenges in drug design. *Frontiers in Pharmacology*, 15, 1462070.