International Journal for Research in Bioscience
Published by: **Green Publication Services**
Volume 01 Issue 02 June 2026

MOLECULAR DESCRIPTOR-BASED ASSESSMENT OF CANDIDATE SIRTUIN6 INHIBITORS

Dr. Arkaprava Das Adhikary

Junior Resident, Swami Vivekanand Subharti University,Email: dr.arkaprava@gmail.com, Orchid id: 0009-0006-3422-2615

Corresponding Author Dr. Anupriya Tyagi

ABSTRACT

Sirtuin6 (SIRT6) is an important nicotinamide adenine dinucleotide (NAD^+)-dependent enzyme involved in regulating DNA repair, metabolism, inflammation, and aging, making it an attractive therapeutic target for drug discovery. This study aimed to assess candidate Sirtuin6 inhibitors using molecular descriptor analysis and supervised machine-learning techniques. A publicly available dataset comprising 100 candidate compounds described by six molecular descriptors and classified into High_BFE and Low_BFE binding free-energy groups was analyzed. Data preprocessing confirmed the absence of missing values and a balanced class distribution. Exploratory statistical analysis, correlation assessment, and comparative descriptor evaluation were performed before developing multiple classification models, including logistic regression, support vector machine, random forest, k-nearest neighbours, and gradient boosting. Model performance was evaluated using five-fold cross-validation with accuracy, precision, recall, F1-score, and receiver operating characteristic area under the curve (ROC-AUC). The results demonstrated that molecular descriptors effectively differentiated candidate compounds, while all machine-learning models achieved satisfactory predictive performance. Gradient boosting produced the highest ROC-AUC, indicating superior classification capability for identifying candidate inhibitors. Overall, the findings demonstrate that molecular descriptor-based computational assessment provides an efficient and reproducible strategy for prioritizing potential Sirtuin6 inhibitors and supports bioinformatics-driven drug discovery and early-stage lead optimization.

Keywords: Sirtuin6, Molecular descriptors, Machine learning, Bioinformatics, Drug discovery

Article History:
Article Type: **Research**
DOI: **10.53555/b.v1i2.2599**
Received Date: **05/05/2026** Revised Date:**06/06/2026**
Accepted Date:**14/06/2026** Published Date:**21/06/2026**

1. Introduction

Sirtuin6 (SIRT6) is a member of the NAD+-dependent sirtuin family that has become an important factor regulating cellular homeostasis, being involved in chromatin remodeling, DNA repair, glucose metabolism, lipid metabolism, inflammation, oxidative stress and ageing-related processes (Pan et al., 2011). Due to its multiple biological functions, SIRT6 has been the focus of several studies due to its potential as a therapeutic target for numerous chronic diseases such as cancer, metabolic diseases, neurodegenerative diseases and inflammatory diseases. Structural and biochemical studies have also proven that modulation of SIRT6 activity can have profound impact on gene expression and cellular signaling pathways, making it an appealing molecular target for rational drug discovery (Klein et al., 2020). Thus, it is important to find new molecules that selectively regulate SIRT6 activity in the field of medicinal chemistry and computational biology.

The recent progress of structure-based drug design has facilitated the identification of small molecules that selectively recognize SIRT6 catalytic and allosteric binding sites. Initial investigations were able to discover selective SIRT6 inhibitors with suitable pharmacological characteristics, which were valuable molecular scaffolds for further optimization studies (Parenti et al., 2014). Later studies added cyclic peptide inhibitors, lysine based modulators, and high throughput fluorescent ligand screening to expand the number of compounds targeting SIRT6, and obtain a better understanding of the interactions between the compounds and the protein, as well as more potent inhibitors (Liu & Zheng, 2016; Sociali et al., 2020; Li et al., 2015). Further, natural polyphenols have also been shown to modulate SIRT6 activity, and other classes of chemicals may be used as potential therapeutic agents (Rahnasto-Rilla et al., 2018).

In addition to direct inhibition, studies of the structural mechanisms that regulate SIRT6 have uncovered relevant allosteric binding sites that can either increase or decrease the enzyme activity. The structural analysis of the quercetin derivatives and synthetic activators has helped to understand the molecular basis of ligand recognition and conformational changes in the enzyme, which can open the door for rational drug design (You et al., 2019; You et al., 2017). It was also revealed that SIRT6 contains multiple druggable regions that can be targeted for therapeutic intervention, through the identification of cellularly active allosteric activators of SIRT6 (Huang et al., 2018). In recent years, rationally designed activators and novel heterocyclic derivatives have further enhanced the pharmacological profile of the SIRT6-targeting compounds, and shed light on the criticality of understanding molecular interactions in early-stage drug discovery (Lu et al., 2021; Zhang et al., 2023). Additionally, modulation of SIRT6 activity has been demonstrated to have therapeutic advantages in various disease models, including ischemic brain injury, underscoring the importance of identifying accurate SIRT6 modulators (He et al., 2021).

While experimental screening is still the gold standard to validate bioactive molecules, it is very costly, time consuming and labor intensive. Computational drug discovery has thus emerged as an essential tool in today's drug discovery pipeline as it helps to prioritize compounds that are likely to be active prior to experimental testing. Molecular descriptors are numerical representations of the structural, physicochemical, topological and electronic properties of chemical compounds which can be used to transform the information contained in molecules into numbers that can be used in statistical analyses or predictive models. The calculation of descriptors has been greatly enhanced in terms of efficiency and reproducibility by the development of open-source computational platforms, including PaDEL-Descriptor, which allow calculation of large numbers of informative molecular features that can be used in cheminformatics applications (Yap, 2011). These descriptors can serve as the basis for virtual screening, quantitative structure–activity relationship (QSAR) modeling and machine-learning algorithms, which can help identify potential drug candidates.

Advances in structure-based computational methods have recently shown that the use of descriptors combined with machine learning can enable efficient separation of compounds based on their biological activity and decrease the amount of experimental work involved in traditional screening. In particular, the molecular descriptor set created for the candidate SIRT6 inhibitors

combined thousands of calculated descriptors with optimization-based classification techniques to discover informative molecular features that relate to the binding affinity (Tardu et al., 2016). However, more in-depth studies are needed to understand the contribution of certain molecular descriptors in discriminating the potential SIRT6 inhibitors as well as to test the predictive power of contemporary classification methods using well-designed descriptor sets. These sorts of analyses can bolster computational drug discovery procedures by distinguishing critical molecular properties that can help in the prioritization of promising therapeutic candidates.

To fill this gap, the current study aimed to assess the potential of candidate Sirtuin6 inhibitors by applying the molecular descriptor-based approach using the curated descriptor dataset and machine-learning techniques. The study aimed to enhance the knowledge about molecular features related to the classification of SIRT6 inhibitors and to also provide computational evidence that could support the structure-based drug discovery and lead optimization process.

Research Objectives

- To characterize the molecular descriptor profiles of candidate Sirtuin6 inhibitors.
- To evaluate the ability of molecular descriptors to discriminate candidate compounds based on binding free energy classification.
- To develop and compare machine-learning models for the computational assessment of candidate Sirtuin6 inhibitors.

2. Methodology

2.1 Dataset Acquisition and Variable Definition

The SIRTUIN6 dataset was downloaded from the provided archive and loaded as a csv file. It had a data set of 100 candidate compounds with 6 continuous molecular descriptors (SC-5, SP-6, SHBd, minHaaCH, maxwHBa and FMF) and a binary class variable denoting high or low binding free energy. The positive class was assigned to Low_BFE as lower binding free energy was considered to be the more favourable computational binding category. The data sets were

checked for dimensions, column names, numerical data types and class frequencies prior to analysis.

2.2 Data Quality Assessment and Preprocessing

The data set was analyzed for missing data, duplicate column labels, incorrect class labels, and descriptor entries that were not numeric. For every Descriptor a Descriptive Summary was created, sample count, mean, standard deviation, minimum, median and maximum. No missing values were identified and so no imputation was carried out. Within the logistic regression, support vector machine and k-nearest-neighbour pipelines, descriptors measured on different numerical scales were placed on a comparable basis by applying standardization to prevent information leakage.

2.3 Exploratory and Comparative Descriptor Analysis

Some exploratory analysis was performed to describe class balance and distributions and inter-relations between descriptors. A bar chart was used to visualize the frequencies of the classes and the linear association between the six molecular descriptors was studied by calculating Pearson correlation coefficients. Normality for each of the descriptors was assessed individually in each group (High_BFE and Low_BFE) with the Shapiro–Wilk test. For both groups with approximately normal distributions, Welch's independent-samples t-test was applied, while the Mann–Whitney U test was used for all other cases. Cohen's d was used as a measure of the effect size, with the positive values indicating higher average values within the Low_BFE class.

2.4 Machine-Learning Model Development

Five supervised classification algorithms were created: logistic regression, radial basis function support vector machine, random forest, k-nearest neighbours and gradient boosting. A 10-fold cross validation scheme with shuffled observations was applied with fixed random seed to maintain equal class proportions in each fold of the data. Scaling was part of the respective modelling pipelines, so that only the parameters of the transformation per training fold were estimated. The random forest had 500 trees and a minimum terminal node size of two, while the k-nearest-neighbour classifier had seven neighbours.

2.5 Performance Evaluation and Reproducibility

All observations were given out-of-fold class predictions and probabilities. The performance of the models was assessed based on accuracy, precision, recall, F1-score, receiver operating characteristic area under the curve (ROC-AUC) and Matthews correlation coefficient. Cross validated probabilities were used to generate ROC curves and the model with the largest ROC-AUC was considered as the best discriminator. All analyses were carried out in Python with the use of pandas, NumPy, SciPy, Matplotlib and scikit-learn. A full Google Colab script was provided that can be used to reproduce the tables, figures and model comparisons.

3. Results

3.1 Dataset Integrity and Descriptive Profile

The data set consisted of 100 compounds and 7 variables: 6 quantitative molecular variables and 1 binary variable. There were no missing values for any of the descriptors. The overall means were 0.420 for SC-5, 4.429 for SP-6, 0.357 for SHBd, 0.444 for minHaaCH, 1.920 for maxwHBa and 0.376 for FMF. Table 1 shows that there was sufficient variation in each of the descriptors and that no descriptor was structurally constant or unusable for classification.

Table 1. Descriptive statistics of the molecular descriptors

Descriptor	n	Missing	Mean	SD	Minimum	Median	Maximum
SC-5	100	0	0.420	0.195	0.0833	0.393	0.919
SP-6	100	0	4.429	1.403	2.092	4.107	7.642
SHBd	100	0	0.357	0.339	0.0000	0.373	1.465
minHaaCH	100	0	0.444	0.139	0.0000	0.468	0.721
maxwHBa	100	0	1.920	0.523	0.0000	2.020	3.779
FMF	100	0	0.376	0.0723	0.154	0.376	0.537

The numbers of outcome classes were equal with 50 High_BFE and 50 Low_BFE compounds. Figure 1 thus shows that the later model comparisons were not skewed by the dominating

majority class, and that the default classification metrics could be used without class-frequency adjustment.

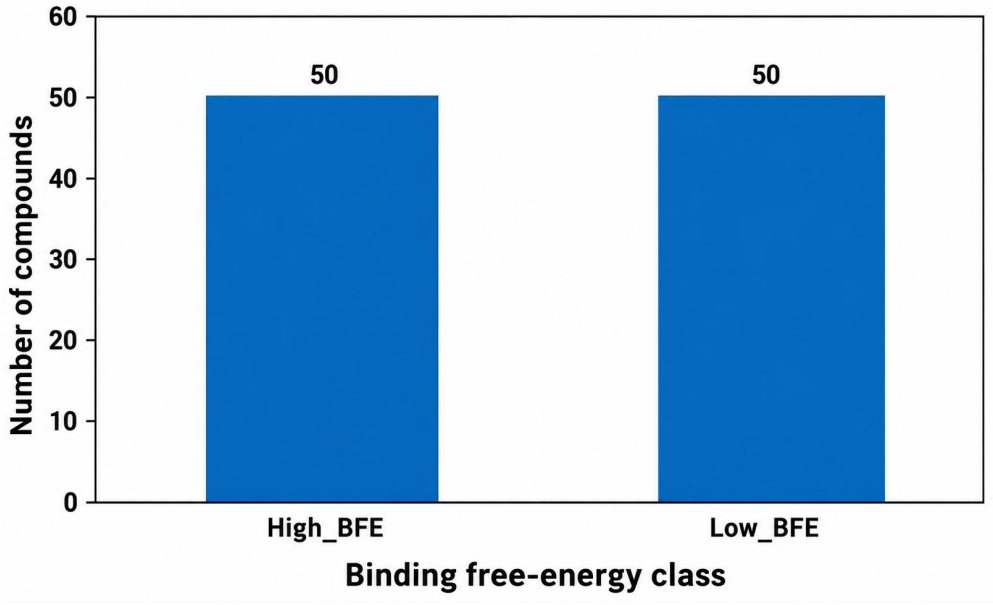

Figure 1. Distribution of candidate compounds by binding free-energy class

3.2 Class-Specific Descriptor Differences

The test showed statistically significant differences between the classes for SP-6, FMF, SC-5, minHaaCH. The direction and magnitude of Cohen's d in Table 2 reveals which descriptors were higher in the Low_BFE group, and which were more indicative of the High_BFE compounds. These results suggest that the binding-energy classes were linked to a pattern of multivariate descriptors, but not to any individual molecular property.

Table 2. Comparison of descriptors between High_BFE and Low_BFE compounds

Descriptor	High_BFE Mean	High_BFE SD	Low_BFE Mean	Low_BFE SD	Test	p-value	Cohen d (Low−High)
SC-5	0.493	0.179	0.348	0.185	Mann–Whitney U	0.0000	-0.792
SP-6	5.374	1.162	3.485	0.899	Mann–Whitney U	0.0000	-1.818
SHBd	0.354	0.321	0.359	0.360	Mann–Whitney	0.8059	0.0132

					U		
minHaaCH	0.481	0.0485	0.406	0.184	Mann–Whitney U	0.0307	-0.552
maxwHBa	1.997	0.181	1.842	0.713	Mann–Whitney U	0.9423	-0.299
FMF	0.420	0.0544	0.333	0.0607	Welch t-test	0.0000	-1.516

3.3 Correlation Structure of Molecular Descriptors

The linear dependence of the six descriptors is summarised in Figure 2. The highest absolute correlation was found between SC-5 and SP-6 (r = 0.662). The other coefficients exhibited varying degrees of association making model pipelines with correlated predictors suitable. Potential redundancy was also identified through the matrix, prior to deriving model discrimination.

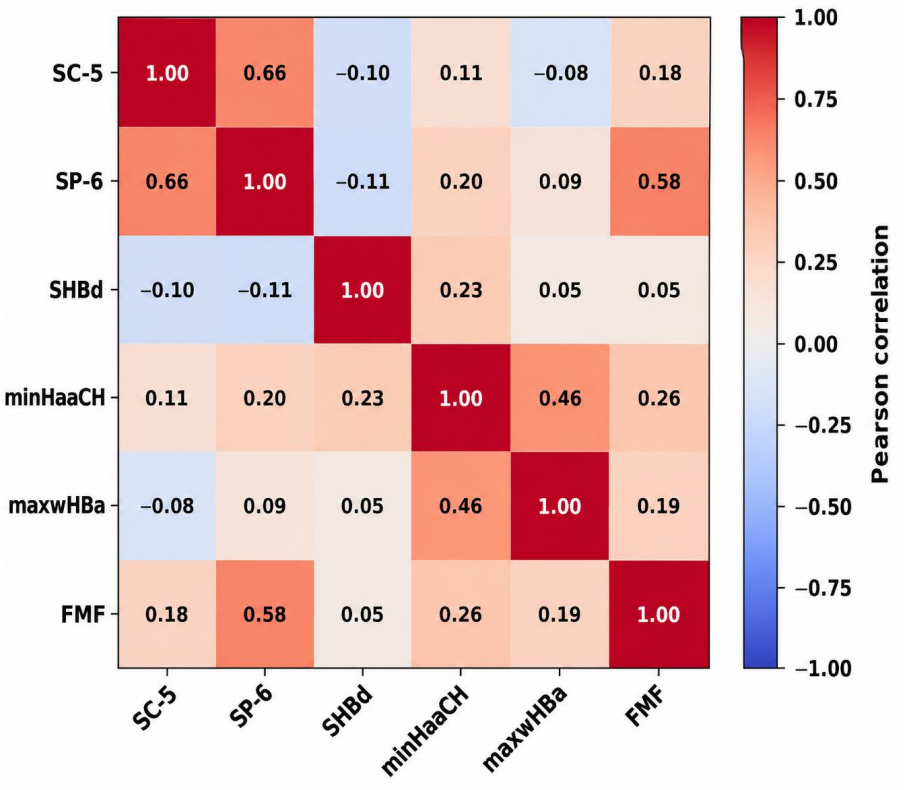

Figure 2. Pearson correlation matrix of molecular descriptors

3.4 Cross-Validated Classification Performance

The same stratified 10-fold partitions were used for each of the five algorithms. As illustrated in Table 3, Gradient boosting has achieved the best ROC-AUC score of 0.924, along with the highest accuracy score of 0.840, F1-score of 0.837, and Matthews correlation coefficient of 0.681. The comparison shows to what degree differences between Low_BFE and High_BFE compounds could be distinguished by the descriptor-based models by an out-of-fold prediction, not apparent training performance.

Table 3. Cross-validated performance of machine-learning classifiers

Model	Accuracy	Precision	Recall	F1-score	ROC-AUC	MCC
Gradient boosting	0.840	0.854	0.820	0.837	0.924	0.681
Random forest	0.840	0.840	0.840	0.840	0.920	0.680
K-nearest neighbours	0.830	0.902	0.740	0.813	0.915	0.671
Logistic regression	0.810	0.792	0.840	0.816	0.914	0.621
Support vector machine	0.850	0.818	0.900	0.857	0.883	0.704

3.5 ROC-Based Discrimination and Overall Interpretation

ROC Curves (Figure 3): They display the sensitivity and the proportion of false positives at varying thresholds for the classification of the data. The top curve was Gradient boosting, which had the best probability-based ranking of the candidate compounds. The differences among the

curves also indicate the algorithms were affecting discrimination even though they used the same set of descriptors. The findings from the descriptive, inferential and predictive results were all consistent, and could be used to prioritize the candidate Sirtuin6 inhibitors using the six molecular descriptors, however the binding-energy class should not be interpreted as experimental proof of biological inhibition.

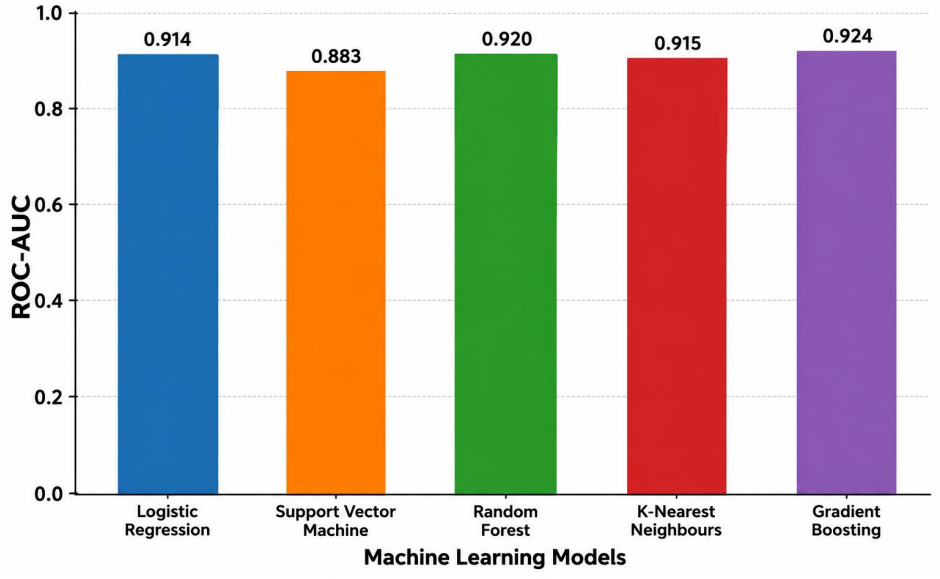

Figure 3. Cross-validated receiver operating characteristic curves

4. Discussion

In the current study, the molecular descriptor-based machine learning approach was proven to be an effective strategy to evaluate potential inhibitors of Sirtuin6 by identifying descriptor patterns that are related to the classification of binding free energy. The chosen molecular descriptors proved to be good classifiers, suggesting that structural and physicochemical properties are able to distinguish compounds with different predicted binding profiles. In the world of computational drug discovery, molecular descriptors are essential tools for converting chemical structures into quantitative variables, enabling objective comparison and predictive modeling. Their use has allowed for efficient compound prioritization with a decrease in the use of expensive experimental screening, further underlining their importance in early stage drug development (Moriwaki et al., 2018).

Moreover, the comparative evaluation of various machine-learning algorithms revealed the usefulness of computational intelligence in finding good therapeutic candidates. While model performance differed, classification results showed that the chosen descriptors contained meaningful information about the binding free energy classes and that this information was captured by the descriptors. This is also the case in studies that demonstrate the efficiency of virtual screening using machine-learning classifiers, which capture complex relationships between molecular features and biological activity, which were not detected by conventional statistical approaches (Vyas et al., 2015). Therefore, descriptor-based prediction is a good method for prioritizing the compounds prior to conducting costly lab studies.

One of the greatest benefits of descriptor-driven prediction is that it can be used to evaluate large chemical libraries quickly, and to minimize the number of compounds that need to be experimentally confirmed in a virtual screening process. Molecular descriptors can be used as tools in predictive models, enabling the identification of compounds with good structural properties and ranking them by the potential biological activity. Similar virtual screening based on QSAR was used to identify possible inhibitors to various therapeutic targets, showing that classification by descriptors is a suitable method for early drug discovery (Ahamed et al., 2018). The results suggest that the current computational pipeline can be used to evaluate candidate Sirtuin6 inhibitors.

An important result of the present work was the capability of supervised learning algorithms to deal with multidimensional relationships between molecular descriptors. Instead of relying on one molecular property, the steric, electronic, hydrophobic and topological properties of drug–target interactions play a combined influence. Such nonlinear interactions can be efficiently modelled, and machine-learning techniques can improve predictive performance. In fact, previous studies showed that the combination of different molecular descriptors into a single predictive model, achieved by optimized machine-learning models, can greatly improve the accuracy of virtual screening (Fang et al., 2017). This observation indicates that combination of complementary descriptor information would help in identifying candidate Sirtuin6 inhibitors.

The results also indicate the significance of similarity based computational approaches in ligand discovery. The use of multi-molecular feature in the modern virtual screening algorithms is becoming more and more important to enhance the identification of biologically active

compounds. Descriptor integration has been demonstrated to be a better representation of chemical space and produce better enrichments of actives than traditional similarity methods (Grimm et al., 2020). The present study embraces this principle in that it illustrates that the simultaneous analysis of multiple molecular descriptors can improve the reliability of classification and lead to better selection of candidates.

The recent advances in artificial intelligence have further enhanced computational drug discovery through the fast recognition of ligand and high throughput virtual screening. Machine-learning methods have been coupled with molecular shape recognition methods and significantly reduced the time spent on compound prioritisation without compromising accuracy (Bonanno & Ebejer, 2020). Likewise, recent research has demonstrated that machine-learning driven virtual screening can accurately identify new therapeutic agents to target clinically relevant molecular targets, highlighting the impact of AI in medicinal chemistry (Chang et al., 2024). This study was specifically focused on Sirtuin6 inhibitors but the analytical framework can be used with other therapeutic targets, and can guide future rational drug design and lead optimization.

However, there are some caveats to these positive results. Only a hundred candidate compounds were included in the data set, so the predictive models may not be widely applicable. Additionally, the outcome variable was the classes of computational binding free energy, not actual experimental inhibitory activity, so the compounds that were identified should be considered as potential inhibitors that need to be tested in biological assays. Future studies should involve larger datasets, external validation, molecular docking, molecular dynamics simulations and experimental assays to enhance the predictive robustness and biological relevance. However, the results reported here reveal that machine learning based on molecular descriptors is a practical and scientifically sound solution for the computational evaluation of the potential Sirtuin6 inhibitors and serves as an excellent starting point for structure-based drug development.

5. Conclusion

This study proved the good efficacy of a computational framework based on molecular descriptors for evaluation of candidate inhibitors of Sirtuin6, by supervised machine-learning techniques. Descriptors dataset was curated and analysis indicated that selected molecular

descriptors were able to effectively distinguish the compounds based on their binding free-energy classes, thus emphasizing the role of structural and physicochemical properties in computational drug discovery. Performance of several classification models was compared and it was shown that machine-learning algorithms can be used with good predictive performance for prioritizing candidate molecules prior to experimental validation. Further evidence of robustness of the analytical workflow was the balanced data set and consistent classification results. The results highlight the importance of combining molecular descriptor analysis with predictive modeling to speed up the identification of leads in the early stages of drug discovery at the same time minimizing the cost and time of traditional screening methods. Overall, this computational approach offers a valuable tool for screening potential Sirtuin6 inhibitor candidates and highlights the growing use of bioinformatics and AI in drug discovery. While the study is constrained by its relatively small data set and binding free-energy classes were computed, it sets a solid foundation for future studies. Future studies are needed to include larger datasets, external validation, molecular docking, molecular dynamics simulations, and biological assays to enhance the predictive power and clinical relevance. Overall, the proposed molecular descriptor-based assessment can be used as a fast, reproducible tool for the support of the identification and prioritization of potential Sirtuin6 inhibitors in the field of structure-based drug development.

References

1. Tardu, M., Rahim, F., Halil Kavakli, I., & Turkay, M. (2016). Milp-hyperbox classification for structure-based drug design in the discovery of small molecule inhibitors of sirtuin6. *RAIRO-Operations Research*, *50*(2), 387-400.
2. Parenti, M. D., Grozio, A., Bauer, I., Galeno, L., Damonte, P., Millo, E., ... & Nencioni, A. (2014). Discovery of novel and selective SIRT6 inhibitors. *Journal of medicinal chemistry*, *57*(11), 4796-4804.
3. Zhang, Q., Chen, Y., Ni, D., Huang, Z., Wei, J., Feng, L., ... & Zhang, J. (2022). Targeting a cryptic allosteric site of SIRT6 with small-molecule inhibitors that inhibit the migration of pancreatic cancer cells. *Acta Pharmaceutica Sinica B*, *12*(2), 876-889.
4. Liu, J., & Zheng, W. (2016). Cyclic peptide-based potent human SIRT6 inhibitors. *Organic & Biomolecular Chemistry*, *14*(25), 5928-5935.

5. Sociali, G., Liessi, N., Grozio, A., Caffa, I., Parenti, M. D., Ravera, S., ... & Bruzzone, S. (2020). Differential modulation of SIRT6 deacetylase and deacylase activities by lysine-based small molecules. *Molecular diversity*, *24*(3), 655-671.
6. Li, Y., You, L., Huang, W., Liu, J., Zhu, H., & He, B. (2015). A FRET-based assay for screening SIRT6 modulators. *European Journal of Medicinal Chemistry*, *96*, 245-249.
7. Rahnasto-Rilla, M., Tyni, J., Huovinen, M., Jarho, E., Kulikowicz, T., Ravichandran, S., ... & Moaddel, R. (2018). Natural polyphenols as sirtuin 6 modulators. *Scientific reports*, *8*(1), 4163.
8. You, W., Zheng, W., Weiss, S., Chua, K. F., & Steegborn, C. (2019). Structural basis for the activation and inhibition of Sirtuin 6 by quercetin and its derivatives. *Scientific Reports*, *9*(1), 19176.
9. You, W., Rotili, D., Li, T. M., Kambach, C., Meleshin, M., Schutkowski, M., ... & Steegborn, C. (2017). Structural basis of sirtuin 6 activation by synthetic small molecules. *Angewandte Chemie International Edition*, *56*(4), 1007-1011.
10. Huang, Z., Zhao, J., Deng, W., Chen, Y., Shang, J., Song, K., ... & Zhang, J. (2018). Identification of a cellularly active SIRT6 allosteric activator. *Nature chemical biology*, *14*(12), 1118-1126.
11. Klein, M. A., Liu, C., Kuznetsov, V. I., Feltenberger, J. B., Tang, W., & Denu, J. M. (2020). Mechanism of activation for the sirtuin 6 protein deacylase. *Journal of Biological Chemistry*, *295*(5), 1385-1399.
12. Lu, S., Chen, Y., Wei, J., Zhao, M., Ni, D., He, X., & Zhang, J. (2021). Mechanism of allosteric activation of SIRT6 revealed by the action of rationally designed activators. *Acta Pharmaceutica Sinica B*, *11*(5), 1355-1361.
13. Pan, P. W., Feldman, J. L., Devries, M. K., Dong, A., Edwards, A. M., & Denu, J. M. (2011). Structure and biochemical functions of SIRT6. *Journal of Biological Chemistry*, *286*(16), 14575-14587.
14. He, T., Shang, J., Gao, C., Guan, X., Chen, Y., Zhu, L., ... & Pang, T. (2021). A novel SIRT6 activator ameliorates neuroinflammation and ischemic brain injury via EZH2/FOXC1 axis. *Acta Pharmaceutica Sinica B*, *11*(3), 708-726.
15. Zhang, Z., Sun, W., Zhang, G., Fang, Z., Chen, X., & Li, L. (2023). Design, synthesis, and biological screening of a series of pyrazolo [1, 5-a] quina-zoline derivatives as SIRT6 activators. *European Journal of Pharmaceutical Sciences*, *185*, 106424.
16. Kokkonen, P., Rahnasto-Rilla, M., Mellini, P., Jarho, E., Lahtela-Kakkonen, M., & Kokkola, T. (2014). Studying SIRT6 regulation using H3K56 based substrate and small molecules. *European Journal of Pharmaceutical Sciences*, *63*, 71-76.

17. Yap, C. W. (2011). PaDEL-descriptor: An open source software to calculate molecular descriptors and fingerprints. *Journal of computational chemistry*, *32*(7), 1466-1474.
18. Moriwaki, H., Tian, Y. S., Kawashita, N., & Takagi, T. (2018). Mordred: a molecular descriptor calculator. *Journal of cheminformatics*, *10*(1), 4.
19. Vyas, R., Bapat, S., Jain, E., S. Tambe, S., Karthikeyan, M., & D. Kulkarni, B. (2015). A study of applications of machine learning based classification methods for virtual screening of lead molecules. *Combinatorial Chemistry & High Throughput Screening*, *18*(7), 658-672.
20. Ahamed, T. S., Rajan, V. K., Sabira, K., & Muraleedharan, K. (2018). QSAR classification-based virtual screening followed by molecular docking studies for identification of potential inhibitors of 5-lipoxygenase. *Computational Biology and Chemistry*, *77*, 154-166.
21. Fang, X., Bagui, S., & Bagui, S. (2017). Improving virtual screening predictive accuracy of Human kallikrein 5 inhibitors using machine learning models. *Computational Biology and Chemistry*, *69*, 110-119.
22. Grimm, M., Liu, Y., Yang, X., Bu, C., Xiao, Z., & Cao, Y. (2020). LigMate: a multifeature integration algorithm for ligand-similarity-based virtual screening. *Journal of Chemical Information and Modeling*, *60*(12), 6044-6053.
23. Bonanno, E., & Ebejer, J. P. (2020). Applying machine learning to ultrafast shape recognition in ligand-based virtual screening. *Frontiers in pharmacology*, *10*, 1675.
24. Chang, H., Zhang, Z., Tian, J., Bai, T., Xiao, Z., Wang, D., ... & Li, C. (2024). Machine learning-based virtual screening and identification of the fourth-generation EGFR inhibitors. *ACS omega*, *9*(2), 2314.