Exploring QSAR of FLAP inhibitors using kernel partial least squares modeling: Insights from molecular binary fingerprints

Mandeth Kodiyil Geetha Nambiar and Thaikadan Shameera Ahamed*

Department of Chemistry, Govt College Malappuram, **INDIA**

* Corresponding Author: shameeraahamed8@gmail.com
Received March 01st 2024; Revised May 13rd 2024; Accepted May 15th 2024

Cite this https://doi.org/10.24036/teknomekanik.v7i1.29772

Abstract: FLAP (5-Lipoxygenase Activating Protein) inhibitors, offering targeted intervention in leukotriene biosynthesis and holding therapeutic promise for inflammatory diseases like asthma, are hindered by current inhibitors' off-target effects, limited efficacy, safety concerns, potential drug interactions, and accessibility issues. Given these challenges, computational methods, particularly Quantitative Structure Activity Relationships (QSAR) modeling, are vital for developing novel FLAP inhibitors. This study specifically investigates the QSAR of FLAP inhibitors using Kernel Partial Least Squares (KPLS) modeling. Leveraging a dataset of FLAP inhibitors, we employ KPLS within the Schrödinger Canvas environment to correlate molecular descriptors with biological activity. Out of the eight models developed, the "atom pairs" fingerprint yielded a statistically significant 2D QSAR model with outstanding regression coefficient values (R^2=0.9624). The model demonstrated high predictive ability for external test set data (R^2pred = 0.7105), underscoring its robustness and reliability in accurately predicting biological activity based on molecular structure. Additionally, we tried to visualize the relative contributions of individual atoms within FLAP inhibitors, providing insights into their favorable and unfavorable characteristics. Through the analysis of atomic contributions, we identify key structural motifs crucial for predicting FLAP inhibitor activity. Our findings not only advance our understanding of FLAP inhibitor SAR but also demonstrate the utility of KPLS modeling and atomic contribution analysis in drug discovery efforts. Furthermore, this study contributes to the development of anti-inflammatory therapeutics by elucidating the structural determinants of FLAP inhibitor activity, with potential applications in the treatment of inflammatory disorders.

Keywords: 5-Lipoxygenase Activating Protein; Inflammatory disorders; Kernel partial least squares regression; Visualization of atomic

1. Introduction

FLAP inhibitors represent a crucial class of pharmaceutical compounds extensively studied for their therapeutic potential in inflammatory and respiratory diseases [1]. Particularly, their role in modulating the biosynthesis of leukotrienes, potent mediators of inflammation, underscores their significance in drug development for conditions such as asthma, chronic obstructive pulmonary disease (COPD), and allergic rhinitis [2]. In recent years, the utilization of advanced computational techniques in pharmaceutical research has provided unprecedented insights into the QSAR of FLAP inhibitors [3]. Among these techniques, kernel-based partial least squares regression (KPLSR) models have emerged as powerful tools for analysing complex molecular datasets, facilitating the design and optimization of novel drug candidates.

QSAR models, which provide information on various aspects of the chemical structure that contribute favourably and adversely to the model for predicting biological activity, are important for the rational design of drugs through chemical modifications of existing lead [4], [5]. This goal has been achieved by a variety of approaches, including CoMFA, CoMSIA, 4D QSAR, Phase QSAR,

Hologram QSAR, and StarDrop, although only the latter two are independent of 3D structure [6]. But we need to take care of too many adjustable parameters such as 3D structural alignment, orientation, lattice size, probe atom, cut-off limit, variable selection procedure, etc. and these are the big challenges we faced there. Instead of 3D structural parameters, the implementation of 2D fingerprints is the perfect way to avoid the problem due to the orientation and alignment of 3D conformers. Fingerprints are a set of descriptors that can be easily assembled into a "string" that characterizes a compound. These descriptors can be either binary, numeric or categorical. 2D fingerprints descriptors are the most reliable and common technique used in the discovery of new drugs and are commonly used for the retrieval of hit compounds compared to 3D descriptors [7], [8], [9], [10]. Analysis of similarity based on fingerprints offers a wide range of possibilities in terms of fragment-based approaches, atom/bond typing schemes, rules for bit scaling, and similarity indices. Fingerprints are therefore very useful in the study of the characteristics responsible for the biological activity of molecules.

Eight different types of binary fingerprinting methods have been used in this study to predict structural criteria for selective inhibition of FLAP on three molecular datasets. Here, the appropriateness of the specific fingerprint for the selected molecules is determined by kernel based partial least squire value. To obtain highly predictive and interpretable 2D QSAR models, we have used the KPLS combined with the 2D fingerprint definitions implemented in the Schrodinger software suit's Canvas module. It is well known that the partial least square method has been popular modelling and regression technique in chemometrics. KPLS is an extension of partial least-square regression that adds some non-linearity through a "kernel" in the scalar products of the X variables used in regression [11]. It has a more predictive ability.

This paper aims to delve into the application of KPLSR models in the study of FLAP inhibitors, elucidating their efficacy in predicting the biological activity of novel compounds based on their chemical structure. By integrating molecular descriptors and biological data, KPLSR offers a holistic approach to QSAR analysis, enabling researchers to discern key molecular features contributing to the pharmacological profile of FLAP inhibitors.

2. Material and Methods

2.1 Data set selection

A set of 173 5-(5-cyclobutylpyridin-2-yl) pyrimidin-2-amine derivative were assessed for the ability to bind to FLAP in a binding assay that measures compound-specific displacement of an iodinated (125I) FLAP inhibitor via a scintillation proximity assay format and the result were reported in PubChem [12] with a Assay ID (AID) 1257599 [13]. Compounds are optimized and energy minimized using SE/PM6 method implemented in gaussian 09 software [14]. Then all compounds are converted to Mol2 format using open babel utility. The dataset was then divided into a training set comprising 139 compounds and a test set consisting of 34 compounds. The structural motif of the compound dataset is given in Fig. 1.

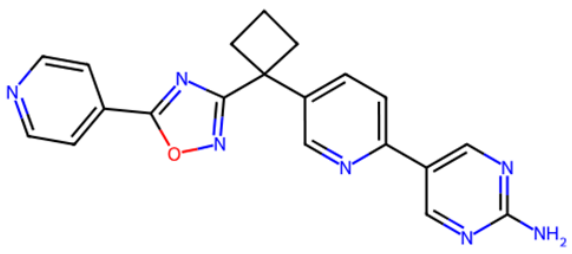

Figure 1. The structural motif of 5-(5-cyclobutylpyridin-2-yl) pyrimidin-2-amine derivatives

2.2 Binary fingerprints

The 2D fingerprints including seven hashed type fingerprints such as Linear, Dendritic, Radial, MOLPRINT2D, Atom pairs and Atom triplets, Topological as well as one structural key type of fingerprint 'MACCS' are calculated for the dataset [8], [15]. In total, eight fingerprints are used to generate eight QSAR models. Linear fingerprints capture substructures formed by linear paths of up to seven bonds. Each path is hashed to produce an integer bit address. For rings, the path length extends to fourteen bonds. Radial fingerprints, derived from the Morgan algorithm, assign identifiers to each atom based on its neighbors, resulting in a 2048-bit vector after removing duplicates. Dendritic Fingerprint incorporates branched features along with linear ones, with a maximum of five bonds per path.

MOLPRINT2D encrypts atom environments into varying-sized strings, finding applications in molecular similarity studies. Atom pairs represent pairs of atoms and their associated distance, useful for pharmacophore analysis and QSAR modeling. Atom triplets fingerprints encode structural information based on the arrangement of groups of three atoms within a molecule, valuable for QSAR modeling. Topological fingerprints encode molecular structure information based on the topology of the molecule, commonly used for similarity searching and machine learning-based modeling. MACCS fingerprints encode structural information based on predefined substructures, widely used in drug discovery for virtual screening and QSAR modeling [16]. These fingerprinting methods are essential for molecular representation and analysis in drug discovery, cheminformatics, and molecular modeling [17].

2.3 Kernel Partial Least Squares (KPLS) method

The study utilized the KPLS method to investigate the correlation between molecular structural properties and biological activity. KPLS amalgamates the robustness of Partial Least Squares (PLS) regression with the adaptability of kernel methods, enabling it to effectively capture nonlinear relationships within the data [18]. Renowned for its versatility, KPLS finds extensive application across disciplines such as chemometrics, bioinformatics, and computational biology. Its utility spans various tasks including regression modeling, classification, and dimensionality reduction. Leveraging a KPLS model within the Schrödinger Canvas environment [19], our study was able to conduct comprehensive analyses with enhanced computational efficiency and streamlined workflow integration. This study has developed eight KPLS QSAR models, each based on a distinct fingerprint.

2.4 Validation of KPLSR Model

The predictive power of the generated models was evaluated by using the R-squared (R^2) value, Root Mean Square Error (RMSE) and predictive correlation coefficient (R^2pred) of an external test set. The R^2 and R^2pred were determined according to the following equations [20].

$$R^2 = 1 - \frac{RSS}{TSS} \quad (1)$$

$$R^2_{Pred} = \frac{SD - PRESS}{SD} \quad (2)$$

where RSS is the sum of squares of residual and TSS is the total sum of squares. Standard Deviation (SD) is defined as the sum of the square deviation between the experimentally observed activity of the test set compounds and the mean activity of the training set molecules. PRESS stands for "Predicted Residual Error Sum of Squares". The sum of squares of the predicted residual errors over all individuals is the PRESS.

3. Results and discussion

The results from our study, which employed KPLS modeling within the Schrödinger Canvas environment, reveal intriguing insights into the correlation between molecular structural properties and biological activity. Utilizing various fingerprint representations, including Linear, Radial, Dendritic, MOLPRINT2D, Atom pairs, Atom triplets, Topological, and MACCS, we conducted a comprehensive analysis to elucidate their efficacy in predicting biological activity.

The statistical qualities of these models are shown in Table 1. Remarkably, the atom pairs fingerprint emerged as a standout performer, demonstrating exceptional correlation with biological activity, as evidenced by its remarkably high R^2 value and low RMSE. Out of eight models developed, the fingerprint atom pairs gave a statistically significant 2D QSAR model with excellent regression coefficient values (R^2=0.9624). High predictive ability of the models for the external test set data (R^2pred = 0.7105) as well as low uncertainties. Therefore, the model with atom pair fingerprints is considered the best model. This finding underscores the importance of selecting an appropriate fingerprint representation in cheminformatics studies, with the atom pairs fingerprint showing promising potential for accurately predicting biological activity based on molecular structural properties.

Among the fingerprint representations, we observed consistent performance metrics. The Linear, Radial, Dendritic, MOLPRINT2D, and Topological fingerprints exhibited low coefficients of determination (R^2), indicating weak correlations with biological activity. However, the predictive accuracy, as measured by the RMSE and R^2pred values, remained stable across these representations, suggesting similar abilities to predict biological activity on unseen data.

Conversely, the Atom triplets and MACCS fingerprints exhibited relatively lower correlations with biological activity, indicating potential limitations in capturing the underlying structural features that influence biological responses.

Table 1. Statistical quality parameters of KPLS models

Model No.	Fingerprint	SD	R^2	RMSE	R^2_{pred}
1	Linear	0.1584	0.9457	0.4182	0.5939
2	Radial	0.1846	0.9263	0.4494	0.5311
3	Dendritic	0.1585	0.9457	0.4368	0.5570
4	MOLPRINT2D	0.2551	0.8592	0.4619	0.5047
5	Atom pairs	0.1318	0.9624	0.3532	0.7105
6	Atom triplets	0.0603	0.9921	0.5294	0.3493
7	Topological	0.2576	0.8564	0.4401	0.5503
8	MACCS	0.3660	0.7102	0.5350	0.3355

The scatter plot in Fig. 2 displays observed versus predicted pIC_{50} values for both the training (black dots) and test sets (red dots) of FLAP inhibitors across various QSAR models. It is evident that experimental and predicted activities are closely aligned, particularly with the atom pairs fingerprint based QSAR, where most molecules exhibit residuals of less than 0.4. These results further affirm the exceptional predictive capability of the established model.

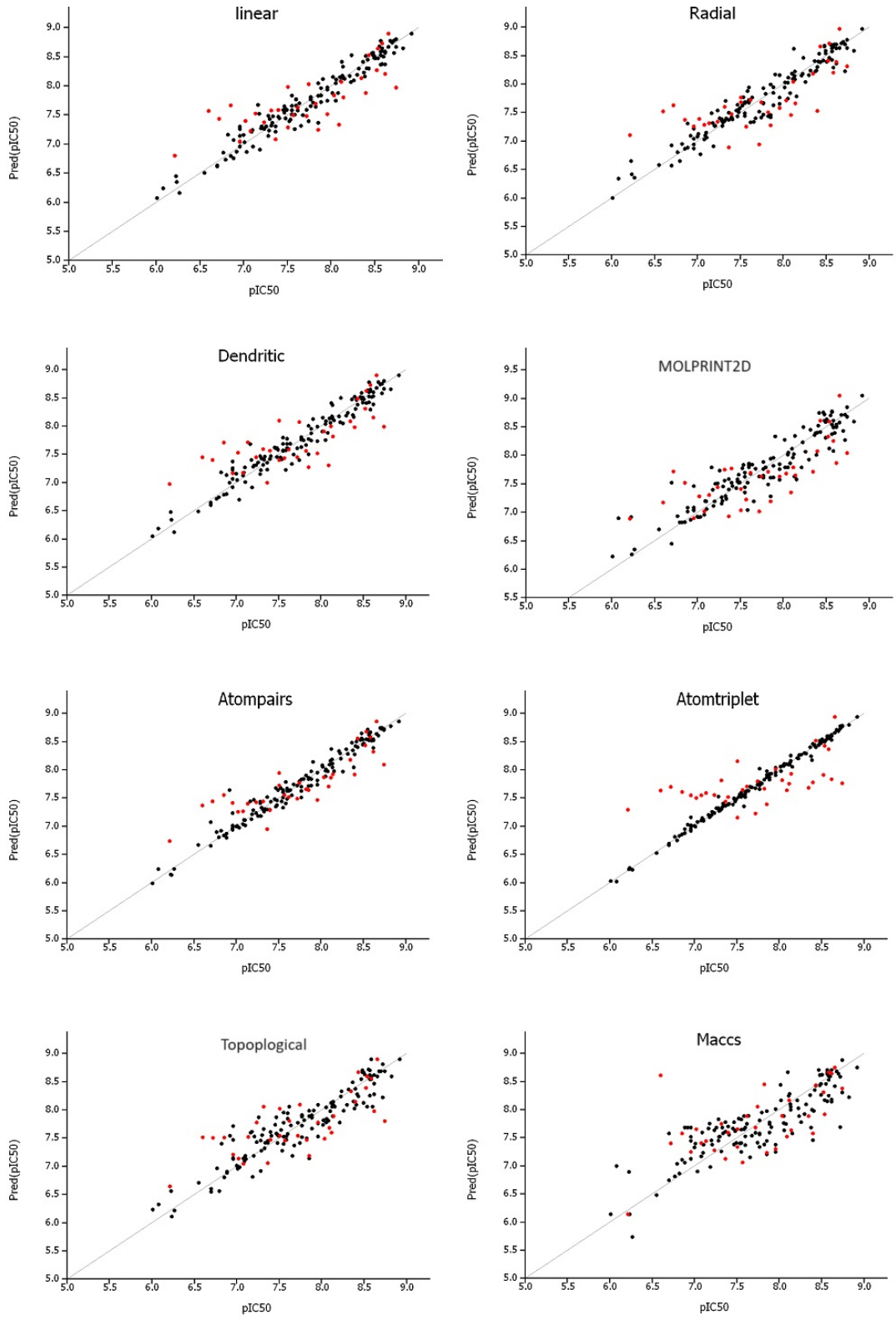

Figure 2. Actual versus predicted activities of FLAP inhibitors from various QSAR models

The analysis of atomic contributions in the KPLS model provides valuable insights into the structural features of molecules that contribute positively or negatively to their biological activity. In the presented study, the relative contributions of individual atoms were visualized, allowing for a detailed assessment of favourable and unfavourable characteristics within the molecular structures.

That is the relative contributions of individual atoms can be displayed in the molecules, allowing the visualization favourable and unfavourable characteristics of individual atoms in the structures

that helps to assess which atoms contributed positively or negatively towards the compound's activity as depicted in Figure 3. For the assessment of atomic contribution to the model, three molecules were taken from each the dataset. The use of colored circles to denote the contributing roles of atoms provides a clear and intuitive visualization of their impact on the compound's activity. Atoms with positive contributions are colored green, indicating their favorable influence on biological activity, while atoms with negative contributions are colored red, suggesting their detrimental effect on activity. The intensity of the color reflects the magnitude of the contribution, enabling the identification of atoms with the most significant impact.

Most of the atoms in 5-(5-cyclobutylpyridin-2-yl)pyrimidin-2-amine group are green in color. The observation that most of the atoms in this group are green suggests that the retention of this structural motif is crucial for predicting FLAP inhibitory activity. This indicates that the presence of the 5-(5-cyclobutylpyridin-2-yl)pyrimidin-2-amine group is associated with increased biological activity, aligning with the goals of drug design to enhance potency. Furthermore, the analysis suggests that substituting the cyclobutyl ring with any other ring structure may decrease the activity. This observation underscores the importance of the specific arrangement and substitution pattern within the molecule for maintaining or enhancing biological activity. It implies that the cyclobutyl ring provides a favorable interaction with the target receptor or enzyme, and altering this moiety may disrupt the binding interactions critical for activity.

This study significantly advances the field by demonstrating the efficacy of KPLS modeling with molecular binary fingerprints in capturing the complex, non-linear relationships in the QSAR of FLAP inhibitors. Previous research often combined multiple techniques, such as pharmacophore modeling, docking, QSAR, and ADMET analyses, or utilized machine learning with GRID-Independent Molecular Descriptors (GRIND) [21], [22], [23]. These methods provided robust predictions but required significant computational resources. Traditional QSAR and 3D-QSAR studies, which relied on simpler linear models and spatial descriptors, offered valuable insights but struggled with complex interactions. Our findings highlight the effectiveness of KPLS in managing non-linearities and leveraging detailed structural data, emphasizing the importance of advanced modeling techniques in QSAR studies.

Overall, the detailed assessment of atomic contributions provided by the KPLS analysis offers valuable guidance for structure-activity relationship (SAR) studies and rational drug design efforts. By pinpointing key structural features that positively influence activity, researchers can focus on optimizing these elements to develop more potent and selective compounds with improved therapeutic properties.

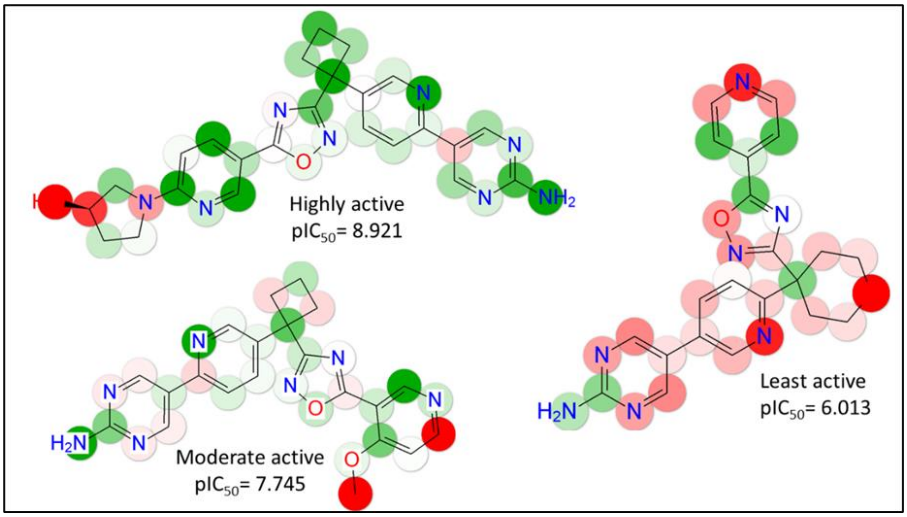

Figure 3. Colour map of atomic effects in KPLS model built from atom triplets fingerprints

4. Conclusion

In conclusion, our study has provided valuable insights into the QSAR of FLAP inhibitors using KPLS modeling. Leveraging the Schrödinger Canvas environment, we analyzed a dataset of FLAP inhibitors to correlate molecular descriptors with biological activity. By employing KPLS, we were able to uncover intricate relationships between molecular structure and FLAP inhibitor potency. Out of the eight models developed, the "atom pairs" fingerprint yielded a statistically significant 2D QSAR model with outstanding regression coefficient values (R^2=0.9624). The model demonstrated high predictive ability for external test set data (R^2pred = 0.7105), underscoring its robustness and reliability in accurately predicting biological activity based on molecular structure.

A novel aspect of our study was the visualization of atomic contributions within FLAP inhibitors, allowing for the identification of key structural motifs influencing biological activity. Through this analysis, we identified specific atoms and structural features that contribute positively or negatively to FLAP inhibitor efficacy. Notably, the retention of the 5-(5-cyclobutylpyridin-2-yl) pyrimidin-2-amine group emerged as crucial for predicting FLAP inhibitory activity, highlighting its importance in drug design efforts targeting this pathway.

Our findings underscore the utility of KPLS modeling and atomic contribution analysis in elucidating QSAR in drug discovery. By providing insights into the structural determinants of FLAP inhibitor activity, our study contributes to the development of more effective anti-inflammatory therapeutics. Future research in this area may further refine our understanding of FLAP inhibitor SAR and facilitate the design of novel compounds with enhanced potency and selectivity.

Author contribution

Mandeth Kodiyil Geetha Nambiar: Conceptualization, Supervision, Methodology, Editing. Thaikadan Shameera Ahamed: Software, Data curation, Visualization, Investigation, Writing-Original draft preparation.

Funding statement

No funders to declare.

Acknowledgements

University of Calicut for Providing Software Canvas.

Competing interest

The authors declare that they have no known competing financial interests or personal relationships that could appear to have influenced the work reported in this paper.

References

[1] K. A. Heidenreich and C. E. Corser-Jensen, “Chapter 12 - 5-Lipoxygenase-Activating Protein Inhibitors: Promising Drugs for Treating Acute and Chronic Neuroinflammation Following Brain Injury,” in *New Therapeutics for Traumatic Brain Injury*, K. A. Heidenreich, Ed., San Diego: Academic Press, 2017, pp. 199–210. https://doi.org/https://doi.org/10.1016/B978-0-12-802686-1.00012-2

[2] D. Pettersen *et al.*, “Discovery and Early Clinical Development of an Inhibitor of 5-Lipoxygenase Activating Protein (AZD5718) for Treatment of Coronary Artery Disease,” *J*

Med Chem, vol. 62, no. 9, pp. 4312–4324, May 2019, https://doi.org/10.1021/acs.jmedchem.8b02004

[3] Z. T. Gür, B. Çalışkan, and E. Banoglu, "Drug discovery approaches targeting 5-lipoxygenase-activating protein (FLAP) for inhibition of cellular leukotriene biosynthesis," *Eur J Med Chem*, vol. 153, pp. 34–48, 2018, https://doi.org/https://doi.org/10.1016/j.ejmech.2017.07.019

[4] A. Cherkasov *et al.*, "QSAR Modeling: Where have you been? Where are you going to?," vol. 57, no. 12, pp. 4977–5010, 2015, https://doi.org/10.1021/jm4004285.QSAR

[5] A. Golbraikh, X. S. Wang, H. Zhu, and A. Tropsha, "Predictive QSAR Modeling: Methods and Applications in Drug Discovery and Chemical Risk Assessment BT - Handbook of Computational Chemistry," J. Leszczynski, Ed., Springer Netherlands, 2016, pp. 1–38. https://doi.org/10.1007/978-94-007-6169-8_37-2

[6] E. N. Muratov *et al.*, "QSAR without borders," *Chem Soc Rev*, vol. 49, no. 11, pp. 3525–3564, 2020, https://doi.org/10.1039/D0CS00098A

[7] J. Yang, Y. Cai, K. Zhao, H. Xie, and X. Chen, "Concepts and applications of chemical fingerprint for hit and lead screening," *Drug Discov Today*, vol. 27, no. 11, p. 103356, 2022, https://doi.org/https://doi.org/10.1016/j.drudis.2022.103356

[8] B. Zagidullin, Z. Wang, Y. Guan, E. Pitkänen, and J. Tang, "Comparative analysis of molecular fingerprints in prediction of drug combination effects," *Brief Bioinform*, vol. 22, no. 6, p. bbab291, Nov. 2021, https://doi.org/10.1093/bib/bbab291

[9] Á. Orosz, K. Héberger, and A. Rácz, "Comparison of Descriptor- and Fingerprint Sets in Machine Learning Models for ADME-Tox Targets," *Front Chem*, vol. 10, 2022, https://doi.org/10.3389/fchem.2022.852893

[10] V. Consonni, D. Ballabio, and R. Todeschini, "Chapter 12 - Chemical space and molecular descriptors for QSAR studies," in *Cheminformatics, QSAR and Machine Learning Applications for Novel Drug Development*, K. Roy, Ed., Academic Press, 2023, pp. 303–327. https://doi.org/https://doi.org/10.1016/B978-0-443-18638-7.00022-0

[11] Y. An, W. Sherman, and S. L. Dixon, "Kernel-Based Partial Least Squares: Application to Fingerprint-Based QSAR with Model Visualization," *J Chem Inf Model*, vol. 53, no. 9, pp. 2312–2321, Sep. 2013, https://doi.org/10.1021/ci400250c

[12] S. Kim *et al.*, "PubChem 2023 update," *Nucleic Acids Res*, vol. 51, no. D1, pp. D1373–D1380, Jan. 2023, https://doi.org/10.1093/nar/gkac956

[13] S. Charleson *et al.*, "Characterization of a 5-lipoxygenase-activating protein binding assay: correlation of affinity for 5-lipoxygenase-activating protein with leukotriene synthesis inhibition.," *Mol Pharmacol*, vol. 41, no. 5, p. 873, 1992, [Online]. Available: http://molpharm.aspetjournals.org/content/41/5/873.abstract

[14] M. J. ; T. Frisch G. W.; Schlegel H. B.; Scuseria G. E.; Robb M. A.; Cheeseman J. R.; Scalmani G.; Barone V.; Mennucci B.; Petersson G. A.; Nakatsuji H.; Caricato M.; Li X.; Hratchian H. P.; Izmaylov A. F.; Bloino J.; Zheng G.; Sonnenb Wallingford CT, "Gaussian 09 (Revision A.2)." 2009.

[15] P. Willett, "Similarity Searching Using 2D Structural Fingerprints," in *Chemoinformatics and Computational Chemical Biology*, J. Bajorath, Ed., Totowa, NJ: Humana Press, 2011, pp. 133–158. https://doi.org/10.1007/978-1-60761-839-3_5

[16] K. Gao, D. D. Nguyen, V. Sresht, A. M. Mathiowetz, M. Tu, and G.-W. Wei, "Are 2D fingerprints still valuable for drug discovery?," *Physical Chemistry Chemical Physics*, vol. 22, no. 16, pp. 8373–8390, 2020, https://doi.org/10.1039/D0CP00305K

[17] J. Duan, S. L. Dixon, J. F. Lowrie, and W. Sherman, "Analysis and comparison of 2D fingerprints: Insights into database screening performance using eight fingerprint methods," *J Mol Graph Model*, vol. 29, no. 2, pp. 157–170, 2010, https://doi.org/https://doi.org/10.1016/j.jmgm.2010.05.008

[18] Y. An, W. Sherman, and S. L. Dixon, "Kernel-Based Partial Least Squares: Application to Fingerprint-Based QSAR with Model Visualization," *J Chem Inf Model*, vol. 53, no. 9, pp. 2312–2321, Sep. 2013, https://doi.org/10.1021/ci400250c

[19] "Canvas, version 2.1, Schrödinger, LLC, New York, NY,." 2020.

[20] P. De, S. Kar, P. Ambure, and K. Roy, "Prediction reliability of QSAR models: an overview of various validation tools," *Arch Toxicol*, vol. 96, no. 5, pp. 1279–1295, 2022, https://doi.org/10.1007/s00204-022-03252-y

[21] K. Rullah *et al.*, "Identification of Novel 5-Lipoxygenase-Activating Protein (FLAP) Inhibitors by an Integrated Method of Pharmacophore Virtual Screening, Docking, QSAR and ADMET Analyses," *Journal of Computational Biophysics and Chemistry*, vol. 22, no. 01, pp. 77–97, 2023, https://doi.org/10.1142/S2737416523500059

[22] H. A. Khan and I. Jabeen, "Combined Machine Learning and GRID-Independent Molecular Descriptor (GRIND) Models to Probe the Activity Profiles of 5-Lipoxygenase Activating Protein Inhibitors," *Front Pharmacol*, vol. 13, 2022, https://doi.org/10.3389/fphar.2022.825741

[23] X. Ma, L. Zhou, Z.-L. Zuo, J. Liu, M. Yang, and R.-W. Wang, "Molecular Docking and 3-D QSAR Studies of Substituted 2,2-Bisaryl-Bicycloheptanes as Human 5-Lipoxygenase-Activating Protein (FLAP) Inhibitors," *QSAR Comb Sci*, vol. 27, no. 9, pp. 1083–1091, 2008, https://doi.org/https://doi.org/10.1002/qsar.200810053