Chem Rev Lett 7 (2024) 926-941

Research Article

Chemical Review and Letters
journal homepage: www.chemrevlett.com
ISSN (online): 2645-4947 (print) 2676-7279

3D-QSAR study on the DYRK1A inhibitors and design of new compounds by CoMFA and CoMSIA methods

Mohammad Sarafraz[1], Eslam Pourbasheer[1,*], Reza Mahmoudzadeh Laki[1], Mohammad Fadaeian[2]

[1]*Department of Chemistry, Faculty of Science, University of Mohaghegh Ardabili, P.O. Box 179, Ardabil, Iran,*
[2]*Department of Geology, Payame Noor University, Tehran, Iran*

ARTICLE INFO

Article history:
Received 21 August 2024
Received in revised form 19 September 2024
Accepted 03 October 2024
Available online 11 October 2024

Keywords:
3D-QSAR
Anti-diabetes
CoMFA
CoMSIA
DYRK1A

ABSTRACT

In recent years, drug design for specific diseases has been of great importance for researchers. In this study, 3D-QSAR modeling was used on a series of 1,5-naphthyridine derivatives in order to design new DYRK1A inhibitors as anti-diabetes. After dividing the data set to the training and test sets, the training set was used to generate statistically significant CoMFA (r^2_{cv} = 0.376, r^2_{ncv} = 0.980) and CoMSIA (r^2_{cv} = 0.365, r^2_{ncv} = 0.783) models based on the common substructure-based alignment. Furthermore, a set of 9 compounds was created for testing the ability of the CoMFA and CoMSIA models to accurately predict compound activity. Also, the application of the CoMFA focus model provided better results (r^2_{cv} = 0.566, r^2_{ncv} = 0.988). The design of new analogues based on naphthyridines as DYRK1A inhibitors was carried out using the knowledge obtained from the contours of the CoMFA focus model. Contours were used to identify structural features of this series of analogs that are related to biological activity. Six new designed compounds, in this group of substances, showed stronger DYRK1A inhibitory activity.

1. Introduction

Diabetes is a condition in which there is a disruption in the balance of insulin supply and demand within the body [1, 2]. In a healthy physiological state, pancreatic beta-cells secrete insulin proportionally to increases in blood glucose levels. However, diabetes arises when there is an insufficient insulin secretion to lower blood glucose, often accompanied by a decline in beta-cell mass [3, 4]. Several different types of cells, such as partially recognized facultative progenitor cells, as well as other cells in the islets, and particularly pre-existing beta-cells, have the capability to produce new beta-cells in adult organisms to boost insulin production capacity [5-9]. Expansion of endogenous β -cells is still the most important method for replacing a person's β -cell bulk [10]. Despite the fact that the mass of adult human β-cells is flexible, these cells rarely show severe expansion [7, 11, 12]. On the other hand, protein kinases that regulate and modulate diabetes-related protein activities make up the diabetic kinome. Numerous experimental findings suggest that alterations in metabolic homeostasis and pharmacological modulations of the diabetic kinome are intimately related [13-15]. Insulin modulates target tissue protein kinase activity to control glucose homeostasis. Insulin resistance is caused by a dysfunction in the kinome response to insulin [16]. More recently, the function of β-cells was linked with the dual-specific tyrosine phosphorylation-adjusted kinase A (DYRK1A). This class of enzymes has evolved over the past 20 years into one of the most significant pharmacological targets due to the vast quantity of data relating to mutations, overexpression, and dysregulation of protein kinases in the pathogenesis of many illnesses[17]. However, despite a wealth of data, finding inhibitors of kinases important to diabetes with specific binding mechanisms and structural characteristics is still

* Corresponding author; e-mail: e.pourbasheer@uma.ac.ir
https://doi.org/10.22034/crl.2024.474579.1410

necessary. Particularly, there are still unanswered questions regarding the impact of protein kinases from the DYRKs family on insulin resistance, cell proliferation, cell cycle control, and apoptosis in diabetes. It has been established that DYRK1A controls regeneration pathways crucial for human pancreatic beta cells to operate more effectively. It has been extensively researched how to use these kinase inhibitors to treat various types of diabetes[18, 19].

DYRK1A is widely expressed and participates in the biochemical processes underlying cancer, apoptosis, neurodegenerative disorders, and brain growth and function [20-24]. Additionally, DYRK1A is involved in several pathways crucial for pancreatic β-cell proliferation [25, 26]. Due to their lower β-cell mass and proliferation, as well as high glucose intolerance, DYRK1A haploinsufficient mice eventually develop diabetes [27, 28]. However, upregulating DYRK1A in their β-cells improves the phenotype [18, 29]. Therefore, DYRK1A is seen as one of the most hopeful targets for medications that treat diabetes [18, 25, 30].

Another point is that the production of drugs for human use is inherently costly and time-consuming, requiring multiple rigorous steps to meet stringent standards. Often, drugs fail to meet these standards, leading to significant wastage of time and resources. To mitigate this issue, it is advisable to conduct standard simulations prior to initiating the production process. If these simulations yield positive results, the production can proceed; however, if the drug design theory proves inadequate, the production process should be halted. Based on recent studies, it can be said that computer calculations and modeling are very good methods for drug design.

One of the main computational molecular modeling tools, are the quantitative structure-activity relationship (QSAR) approaches [31-34], which feather a powerful way for predicting and estimating the activities or characteristics of compounds based on their structures [35]. All of the characteristics of organic compounds, containing their biological, chemical, physical and technical characteristics, are influenced by their chemical structure and change in a predictable way as a result. For society to evaluate and advance the technical, medical, and environmental facets of life, a quantifiable relationship between various molecular attributes and chemical constitution is of the utmost relevance. These relationships between the molecular structures of compounds and their physicochemical, chemical and physical properties are known as quantitative structure activity/property relationships (QSAR/QSPR)[36-40]. The main stages in developing QSAR models involve obtaining descriptors from molecular structures, picking relevant descriptors (feature selection) that are deemed essential for interpreting the desired property or activity, and building, validating, and refining QSAR models [41-46]. 3D-QSAR is a method for applying force field calculations that need 3D structures. Prior to descriptor calculation, uniform molecular alignment in space is necessary 3D-QSAR descriptors. Alignment-based 3D-QSAR methods that are widely used include comparative molecular field analysis (CoMFA) [47, 48] and comparative molecular similarity indices analysis (CoMSIA) [47, 49]. These techniques have two limitations. In order to compare molecules that are structurally different, it is important to use the correct form of the molecule, which may not necessarily be the conformation with the lowest energy. Furthermore, proper alignment of the compounds is necessary, leading to an increased analysis time [50]. However, researchers continue to consider them as effective methods.

The aim of this study is the three-dimensional modeling to predict the activity of 1,5-naphthyridine derivatives as DYRK1A inhibitors using 3D-QSAR method and also, design of new compounds with high performance in the treatment of diabetes.

2. Materials and methods

2.1. Data set

In this study, 47 numbers of 1,5-naphthyridine derivatives from the literature were selected as a data set for 3D-QSAR analysis [10]. In order to obtain precise results, the compounds were split into two groups: 20 percent (9 compounds) were in the test set, and 80 percent (38 compounds) were in the training set. In 3D-QSAR studies, the selection of training and test sets is a crucial step to ensure the validity and generalizability of the model. Dividing the dataset into training and testing sets was done randomly so that the two compounds with the lowest and highest pEC_{50} were kept in the training set to ensure that both sets have a balanced representation of activity values and structural diversity. The inhibitory activities of 1,5-naphthyridine were extracted as EC_{50} values for 3D-QSAR analysis and converted to logarithmic values [pEC_{50}]. Converting activity values to a logarithmic scale normalizes data, reduces skewness, stabilizes variance, and makes data analysis easier. Table 1, shows the chemical structure and the experimental and predicted values of the biological activity of the compounds.

2.2. Molecular optimization and alignment

The molecules being studied were fine-tuned for ideal shape with a mix of the Powell energy minimization algorithm, Gasteiger-Huckel charges, and an energy gradient convergence criteria of 0.01 kcal/mol. The procedure was executed in collaboration with the standard Tripos molecular mechanics force field in the SYBYL molecular modeling package [51, 52]. It is well known that 3D-QSAR models are often sensitive to the specific alignment strategy employed. Structural alignment, considered the most subjective yet crucial step in CoMFA analysis [53], typically involves selecting a template molecule based on one of the following criteria: (1) the molecule with the most functional groups, (2) the molecule with the highest level of activity, or (3) the lead and/or commercial molecule [54, 55]. In this case, the most active compound, referred to as compound 8, was chosen as the template for aligning all other molecules. Figure 1 displays the chemical structure of molecule 8 (target compound), with its shared portion highlighted in bold. Figure 2, also shows the aligned molecules on the target compound.

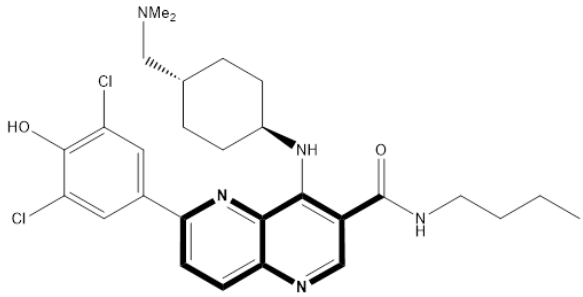


Fig. 1. Chemical structure of template compound (compound 8), Common substure is in bold

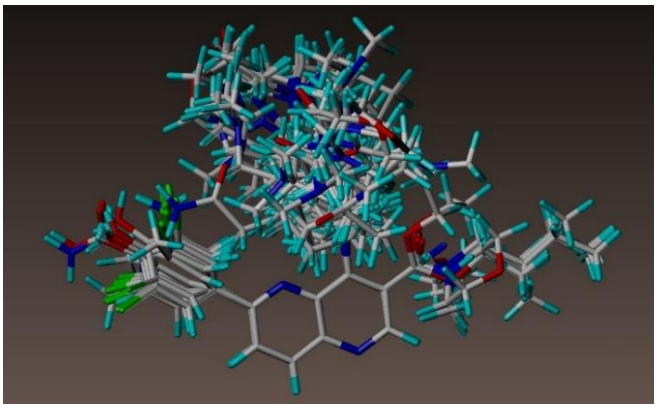


Fig. 2. Alignment of dataset compounds on molecule 8

2.3. CoMFA procedure

A sp^3 carbon atom probe with a van der Waals radius of 1.52 Å and a +1.0 charge was employed to determine the steric and electrostatic fields. The steric and electrostatic impacts were limited to 30 kcal/mol, and the electrostatic impacts were disregarded at lattice intersections with the most intense steric interactions. One can adjust the contribution of lattice points to subsequent analysis by assigning weights to them within a CoMFA region. In this case, 'StDev * Coefficients' values were used in addition to other weighting factors and grid spacing to produce better models [56].

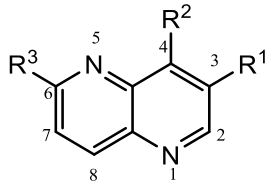
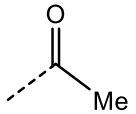
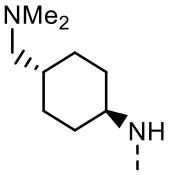
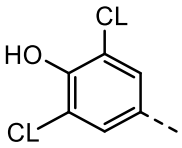
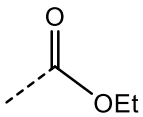
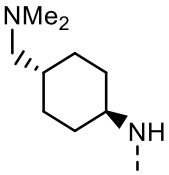
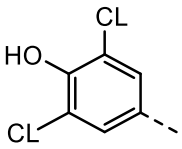
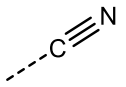
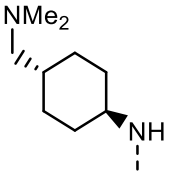
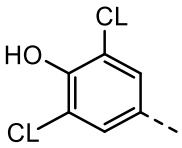
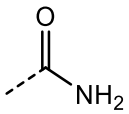
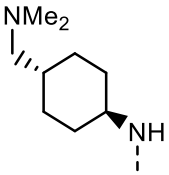
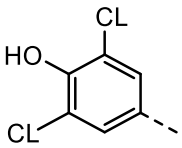
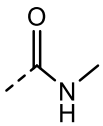
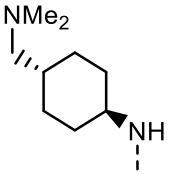
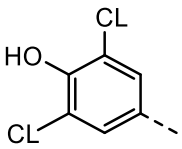
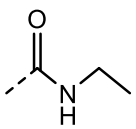
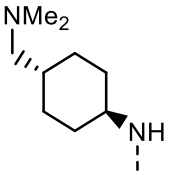
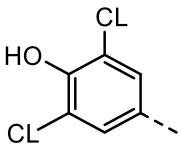
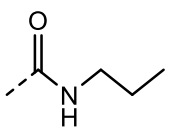
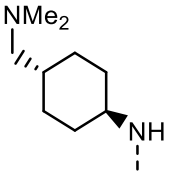
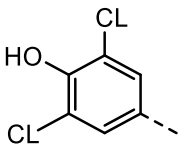
2.4. CoMSIA procedure

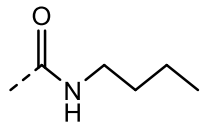
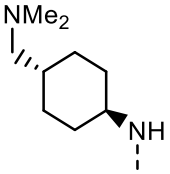
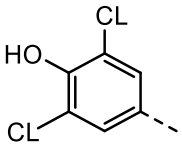
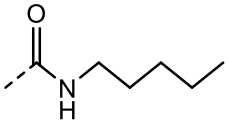
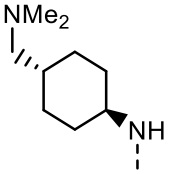
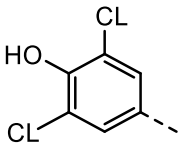
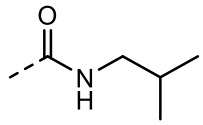
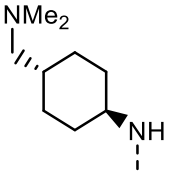
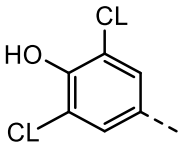
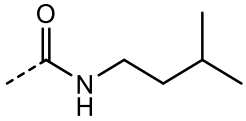
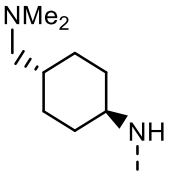
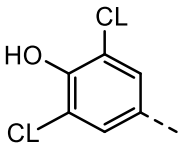
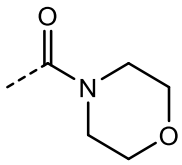
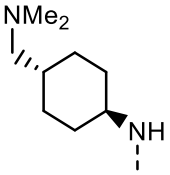
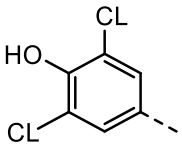
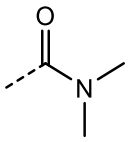
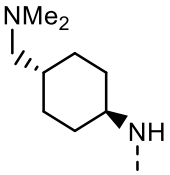
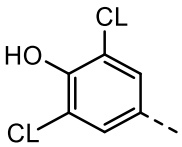
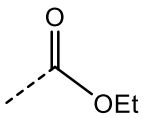
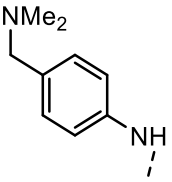
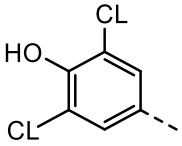
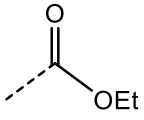
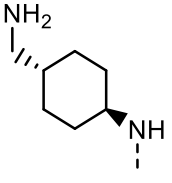
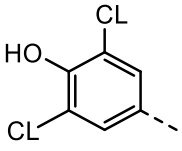
Five physicochemical properties of CoMSIA model, including steric, electrostatic, hydrophobicity, hydrogen bond donor and hydrogen bond acceptor fields, as in the CoMFA model, were determined by a sp3 carbon atom probe with a van der Waals radius of 1.52 Å and charge of +1 [57]. In CoMSIA, steric indices are calculated based on the cube of the atomic radius; the indices for hydrogen bond donors and acceptors are determined based on empirical data [58]. Atom-based parameters developed by Viswanadhan and colleagues are employed in the determination of partial atomic charges and hydrophobic fields [59].

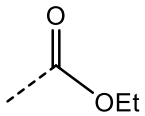
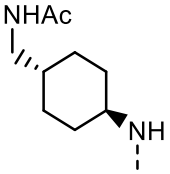
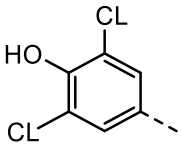
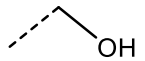
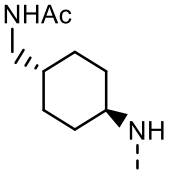
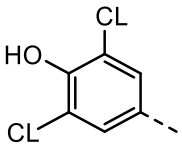
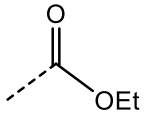
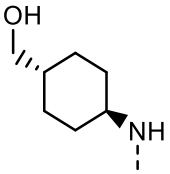
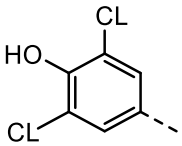
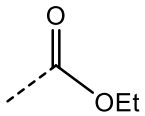
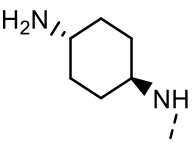
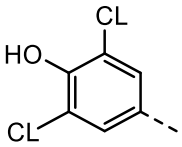
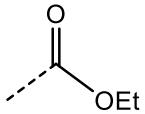
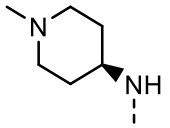
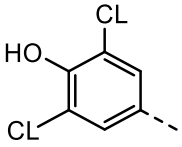
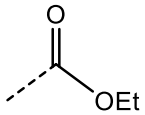
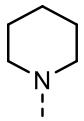
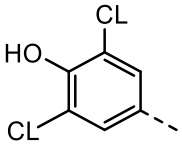
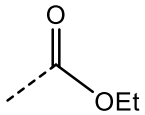
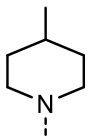
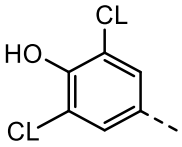
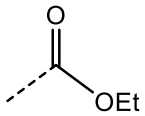
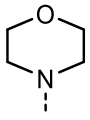
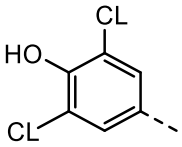
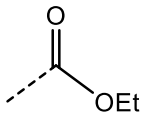
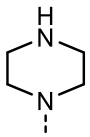
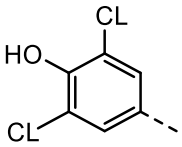
2.5. Model construction

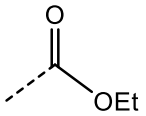
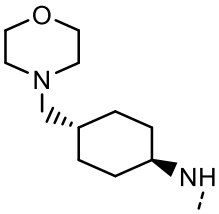
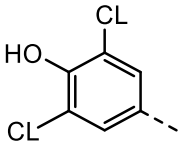
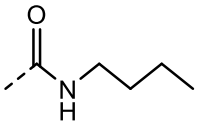
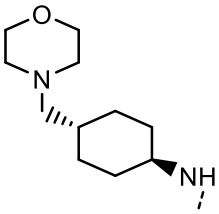
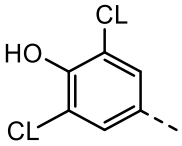
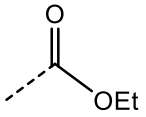
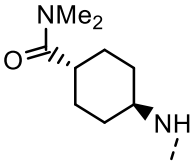
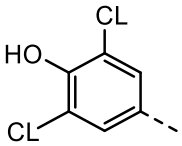
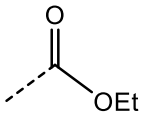
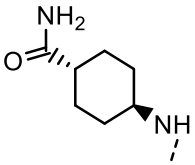
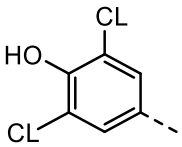
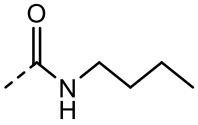
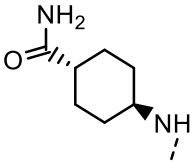
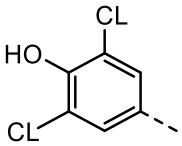
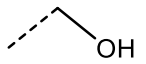
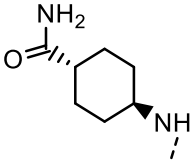
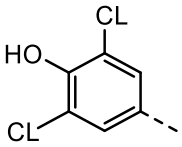
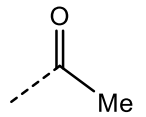
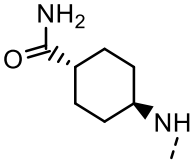
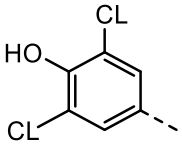
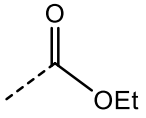
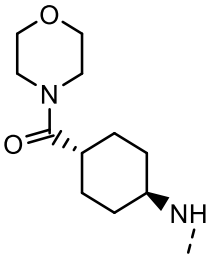
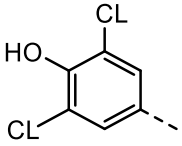
The partial least-squares (PLS) approach was applied to all 3D-QSAR modelling. Using the usual implementation in the SYBYL program, 3D-QSAR models were constructed by means of PLS regression analysis, with the CoMFA and CoMSIA descriptors serving as independent variables and the pEC_{50} values as dependent factors [60, 61]. To decrease noise and speed up the study, a column filtering of 1.0 kcal/mol was implemented. The top model for each instance was chosen through cross-validation employing the leave one out (LOO) method, resulting in the calculation of the cross-validation coefficient squared (q^2) and determining the optimal number of components [62]. A PLS analysis was conducted in non-validation mode with the optimal number of components to obtain the correlation coefficient (r^2_{ncv}), standard error of estimate (SEE), and F values. Moreover, in order to evaluate the statistical significance of the created models, a bootstrapping analysis (r^2_{boot}) and leave-group-out cross-validation (r^2_{Group}) (10 groups) were carried out [63]. The evaluation of the model's performance as a prediction tool was done using the test set, which consisted of chemicals with known activity not used in model development.

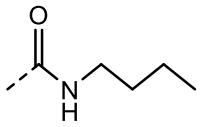
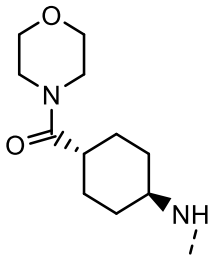
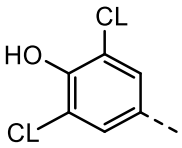
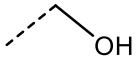
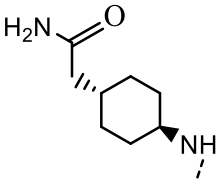
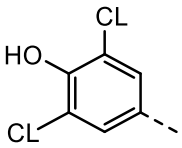
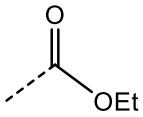
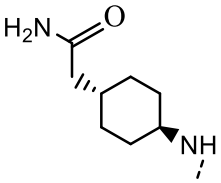
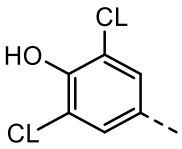
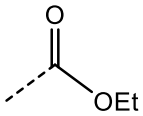
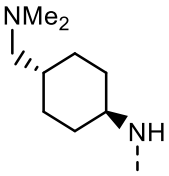
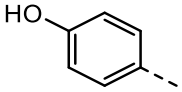
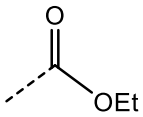
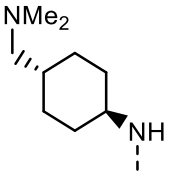
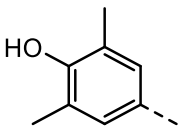
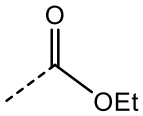
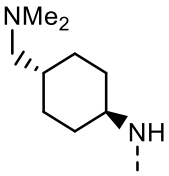
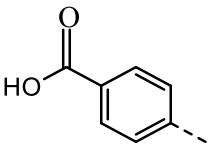
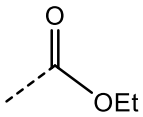
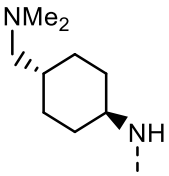
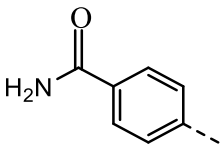
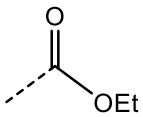
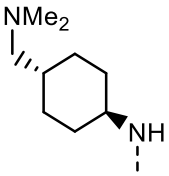
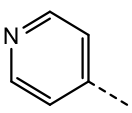
Table 1. Chemical structures of the 1,5-naphthyridines derivatives and their experimental and predicted DYRK1A inhibitory activity.

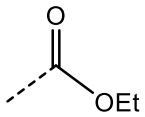
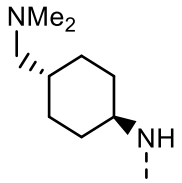
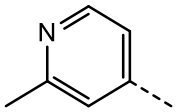
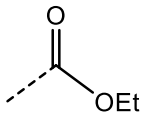
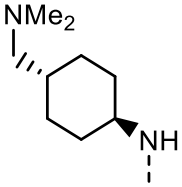
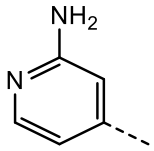
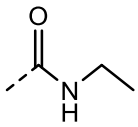
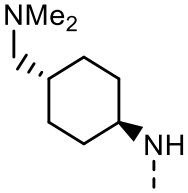
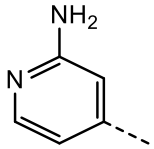
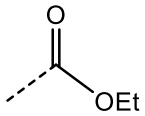
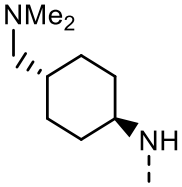
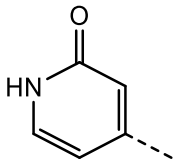
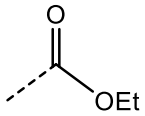
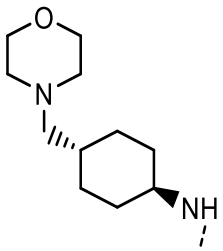
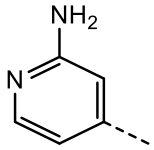
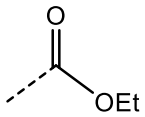
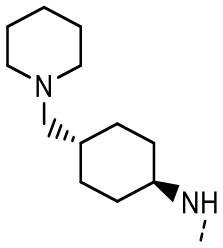
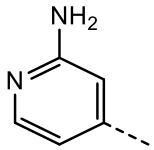
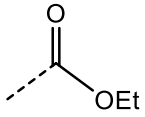
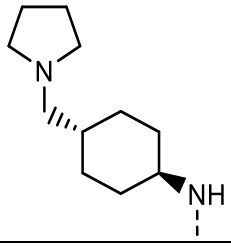
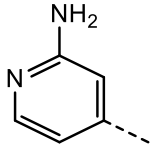
No.	R_1	R_2	R_3	Exp. (pEC_{50})	Predicted		
					CoMFA-1	CoMFA-2	CoMSIA
							
1				8.3	8.38	8.41	8.05
2				8.15	8.02	8.08	7.98
3				7.35	7.33	7.30	7.48
4[a]				7.82	7.95	7.75	8.19
5				8.22	8.19	8.15	7.92
6				8.22	8.05	8.15	7.70
7				8.22	8.17	8.23	8.18

No.	R_1	R_2	R_3	Exp. (pEC_{50})	Predicted		
					CoMFA-1	CoMFA-2	CoMSIA
8				8.52	8.60	8.62	8.38
9[a]				8.3	8.19	8.26	8.28
10				8.22	8.20	8.21	8.18
11				8.52	8.42	8.47	8.34
12				6.19	6.14	6.05	6.91
13				6.45	6.73	6.59	7.32
14[a]				7.59	7.70	7.68	7.73
15				8.1	7.94	7.88	7.42

No.	R_1	R_2	R_3	Exp. (pEC_{50})	Predicted		
					CoMFA-1	CoMFA-2	CoMSIA
16				7.82	7.85	7.84	7.79
17				7.8	7.92	7.83	7.90
18				7	7.25	7.12	7.09
19[a]				8.05	6.11	5.73	6.82
20				8.22	8.03	8.06	7.65
21				6.69	6.56	6.51	7.15
22				6.91	7.05	6.93	7.17
23				6.98	7.10	6.99	7.34
24[a]				7.14	7.19	7.08	7.33

No.	R_1	R_2	R_3	Exp. (pEC_{50})	Predicted		
					CoMFA-1	CoMFA-2	CoMSIA
25				8.15	8.33	8.26	7.94
26				8.4	8.34	8.34	8.41
27				5.9	5.93	6.05	6.27
28				7.3	7.54	7.45	7.66
29				6.76	6.81	6.82	7.06
30[a]				6.17	5.83	5.67	6.89
31				7.96	7.78	7.92	8.43
32				7.01	7.04	7.11	7.06

No.	R_1	R_2	R_3	Exp. (pEC_{50})	Predicted		
					CoMFA-1	CoMFA-2	CoMSIA
33				7.25	7.10	7.31	7.58
34				7.13	7.13	7.17	7.51
35[a]				7.37	6.94	7.05	7.09
36				6.52	6.51	6.44	6.47
37				7.92	7.98	7.95	6.81
38				6	5.93	6.00	5.74
39				5.33	5.29	5.36	5.61
40[a]				6.65	6.89	6.90	7.02

No.	R_1	R_2	R_3	Exp. (pEC_{50})	Predicted		
					CoMFA-1	CoMFA-2	CoMSIA
41				5.82	5.76	5.75	5.95
42				7.23	7.20	7.25	6.84
43				7.34	7.48	7.45	7.66
44				6.42	6.35	6.34	6.00
45				7.43	7.33	7.36	7.19
46[a]				6.94	6.82	6.68	6.25
47				6.97	6.97	6.99	6.61

[a] Compounds used as test set

3. Results and discussion

3. 1. CoMFA analysis

Table 2 displays the statistical parameters of the CoMFA model (CoMFA-1). The CoMFA-1, PLS analysis showed five components and a q^2 value of 0.376. The PLS analysis non-cross-validation analysis showed traditional r^2, F, and SEE values of 0.980, 317.353, and 0.129, respectively. In all the built models, the maximum number of components tested by PLS for modeling was 10. Steric and electrostatic fields contributed 42.5% and 57.5%, respectively. The association between the experimental and projected pEC_{50} values for the CoMFA-1 model in non-cross-validated analysis is depicted in Figure 3. The best outcomes were achieved with a column filter set at 1 kcal/mol for both electrostatic and steric fields. The analytical parameters improved when these fields were focused. The CoMFA focus model (CoMFA-2) shows a substantial cross-validated correlation value (q^2) of 0.566 whit component number of 6. Also, this model obtained values of r^2=0.988, SEE=0.102, and F=430.340 in non-cross-validation mode. The steric and electrostatic fields contributed 43.6% and 56.4%, respectively. Figure 4 shows the association between the experimental and pEC_{50} activities using the CoMFA-2 model. Statistically, data points are considered outliers if their residual value exceeds three times than the standard deviation of the residuals [64]. As the pEC_{50} value forecast for compound 19 was more than three times the standard deviation of the other values in both CoMFA-1 and CoMFA-2 models, it was excluded as an outlier from the test set.

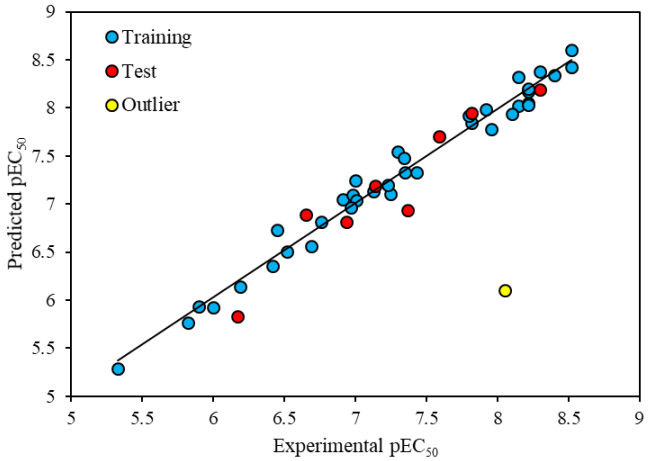


Fig. 3. Plots obtained by the CoMFA-1 model for the relationship between experimental values and predicted pEC_{50} values

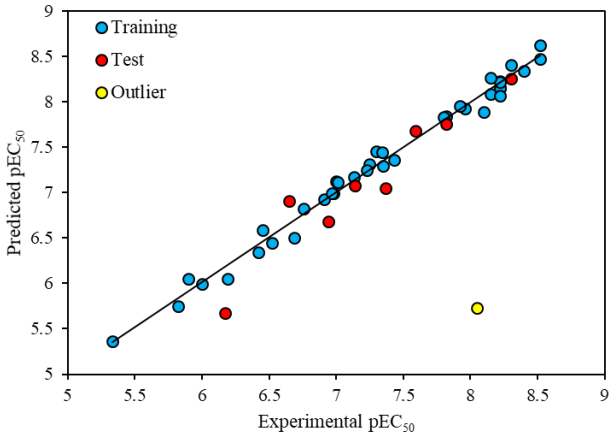


Fig. 4. Plots obtained by the CoMFA-2 model for the relationship between experimental values and predicted pEC_{50} values

3.2. CoMSIA analysis

The CoMSIA method incorporates three additional physicochemical properties, hydrophobicity, hydrogen bond donor, and acceptor, that are not considered in CoMFA. By combining these three characteristics with steric and electrostatic aspects, numerous models can be created. In this research, altering these five physicochemical properties resulted in the development of 31 unique models, as indicated in Table 3. The optimal CoMSIA model included the incorporation of electrostatics, hydrogen bond donor, and acceptor properties (model 23). This particular model obtained an r^2 of 0.365 with leave-one-out cross-validation for two components, an r^2 of 0.783 non-cross-validation, an F value of 63.006, and a SEE of 0.410. The electrostatic, hydrogen bond donor, and acceptor properties contributed 37%, 44.7%, and 18.3%, respectively. Moreover, the CoMSIA model was used to forecast the activity of the molecules in the test set, as demonstrated in Table 1. Figure 5 illustrates the correlation between the predicted pEC_{50} values and the actual activities observed in experiments. The research shows that the CoMSIA model is able to accurately forecast the performance of all compounds in the examination. Since, the CoMFA based model was discovered to excel in cross-validation, hence, the CoMFA-2 model was chosen for the subsequent stages of the study.

Table 2. Statistical results obtained from CoMFA and CoMSIA models

Parameters	CoMFA-1	CoMFA-2(after region focusing)	CoMSIA
PLS statistics			
LOO cross q^2 /SEP	0.376 / 0.727	0.566 / 0.616	0.365 / 0.701
Group cross q^2 /SEP	0.377 / 0.724	0.533 / 0.629	0.394 / 0.685
Non-validated r^2 /SEE	0.980 / 0.129	0.988 / 0.102	0.783 / 0.410
F	317.353	430.340	63.006
$r^2_{bootstrap}$	0.989 ± 0.005	0.994 ± 0.003	0.861 ± 0.049
$S_{bootstrap}$	0.094 ± 0.062	0.071 ± 0.053	0.320 ± 0.173
Optimal compounds	5	6	2
r^2_{test}	0.906	0.919	0.367
Field distribution%			
Steric	0.425	0.436	-
Electrostatic	0.575	0.564	0.37
Hydrophobic	-	-	-
H-bond donor	-	-	0.447
H-bond acceptor	-	-	0.183

SEP: standard error of prediction, SEE: standard errors of estimate.

Table 3. Findings from CoMSIA models utilizing various combinations of the five descriptors

Model	Descriptors	q^2_{LOO} / SEP	Number of Components
1	S	0.078 / 0.833	1
2	E	0.290 / 0.731	1
3	D	0.281 / 0.747	2
4	A	0.196 / 0.789	2
5	H	0.187 / 0.794	2
6	S, E	0.214 / 0.769	1
7	S, D	0.247 / 0.764	2
8	S, A	0.157 / 0.845	5
9	S, H	0.160 / 0.795	1
10	E, D	0.338 / 0.716	2
11	E, A	0.293 / 0.730	1
12	E, H	0.242 / 0.756	1
13	D, A	0.355 / 0.707	2
14	D, H	0.344 / 0.723	3
15	A, H	0.197 / 0.778	1
16	S, E, D	0.311 / 0.731	2
17	S, E, A	0.228 / 0.773	2
18	S, E, H	0.229 / 0.762	1
19	D, A, H	0.346 / 0.712	2
20	D, A, S	0.296 / 0.739	2
21	H, D, S	0.321 / 0.747	4
22	H, A, E	0.264 / 0.745	1
23	**D, E, A**	**0.365 / 0.701**	**2**
24	S, A, H	0.174 / 0.800	2
25	D, E, H	0.334 / 0.718	2
26	S, E, D, A	0.326 / 0.723	2
27	S, E, D, H	0.306 / 0.744	3
28	S, E, A, H	0.232 / 0.761	1
29	D, A, H, S	0.318 / 0.738	3
30	D, A, H, E	0.341 / 0.715	2
31	S, E, D, A, H	0.312 / 0.730	2

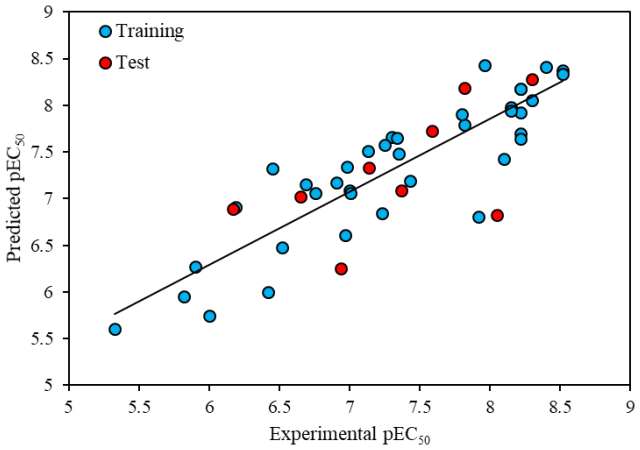


Fig. 5. Plots obtained by the CoMSIA model for the relationship between experimental values and predicted pEC_{50} values

3.3. CoMFA contour maps

The steric and electrostatic contours for the CoMFA-2 model around the target molecule (compound 8) are shown in Figures 6 and 7. The green areas in Figure 6, related to steric contours, show where substituting bulky groups enhances the compound's biological activity. Also, the yellow parts indicate the unfavorable areas for replacing bulky groups to increase biological activity. The contribution of green parts was 80% and the contribution of yellow parts was 20%. Figure 6 shows that the green contours around the target molecule are scattered in three regions R_1, R_2, and R_3. It is understood that the presence of large groups in these areas can enhance the DYRK1A inhibitory activity of the

compounds. Also, the largest green contour is located in the R_2 area. It is known that substituting bulky groups in these regions can increase the DYRK1A inhibitory activity of the compound. To compare the changes in the activity of the compounds in the data set (Table 1) with the change in their structure, we can refer to compounds 23, 24, and 2, respectively. By replacing the bulkier group in the R_2 position, their biological activity also increases (6.98 < 7.14 < 8.15). Yellow contours are mostly spread in the end areas of R_2 position and R_1 position. A comparison of two compounds 25 and 20 shows that substituting smaller groups at the end of the R_2 position increases the biological activity of the compound, as well as replacing the methyl group in compound 20 instead of the bulky morpholine group in compound 25 increases the activity from 8.15 to 8.22. Additionally, when examining compounds 27, 32, and 28, it is evident that the introduction of NMe_2, morpholine, and NH_2 groups in the R_2 position's terminal part correspondingly enhances the compounds' biological activity as a result of smaller agent groups. By comparing compounds 1 and 2, it is confirmed that there are yellow outlines around the location of R_1. Replacing the smaller methyl group instead of the bulky ethyl group in these compounds increases the activity from 8.15 to 8.30. The results of the contours and visual comparisons show that the positions of R_1 and R_2 for spatial fields are of great importance for designing new compounds with higher activity. The corresponding electrostatic contour plots are shown in Figure 7. In the areas marked with red color (20% contribution) replacing negatively charged groups and in blue areas (80% contribution) replacing positively charged groups increases the biological activity of the compound. The optimal contours for replacing negative charges are scattered around the R_1 (high intensity) and R_2 (low intensity) positions. The desired contours for positive charges are also partially specified in the R_3 position of the template molecule. Substituting the morpholine group for the piperidine group in the R_2 position enhances the activity from 6.69 to 6.98, as demonstrated by the comparison of compounds 23 and 21. Also, the comparison of compounds 20 and 21 shows that replacing the more negative N,1-dimethylpiperidin-4-amine group instead of less negatively charged group in the same position has caused a significant increase in biological activity. In the comparison of compounds 24 and 23, replacing the group with a more negative charge instead of the group with a less negative charge in the R2 position increases the biological activity.

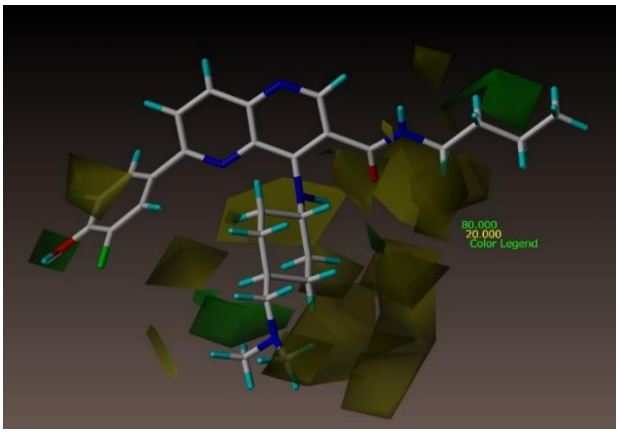


Fig. 6. Contour maps obtained from COMFA-2 model for steric fields

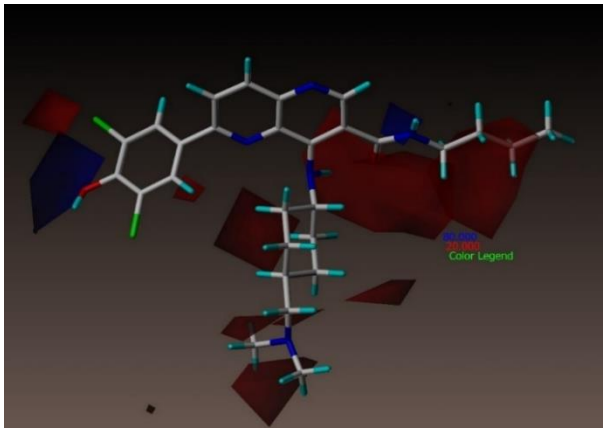
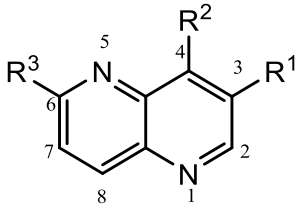
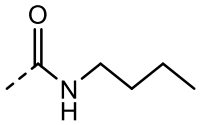
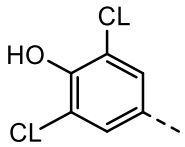
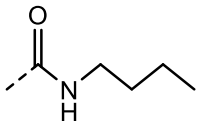
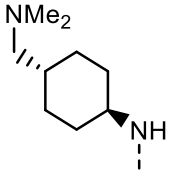
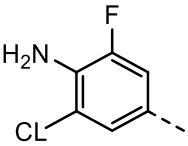
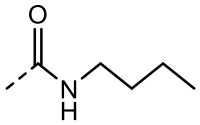
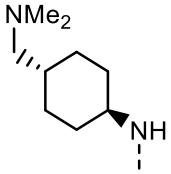
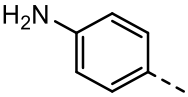
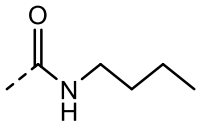
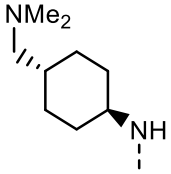
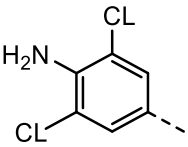
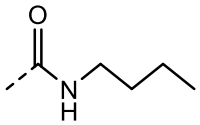
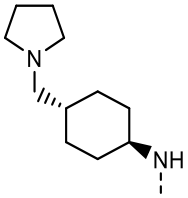
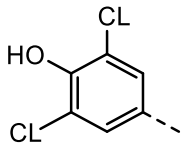
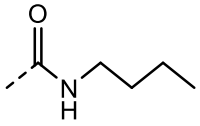
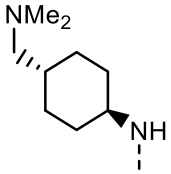
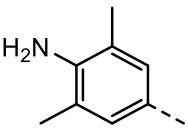
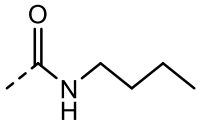
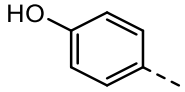


Fig. 7. Contour maps obtained from COMFA-2 model for electrostatic fields

3.4. Design of new DYRK1A inhibitors

The analysis of the contours corresponding to the CoMFA-2 model, discussed above, provided a wealth of knowledge about the structural prerequisites for the design of compounds with high biological activity. Compound 8 (the most potent molecule) was chosen as the template structure for making new compounds. Substitution of specific groups in the positions of R_1, R_2, and R_3 based on the results of the contours in Figures 6 and 7 led to the creation of new compounds. By adding bulky groups like 2,6-dimethylaniline to the R_3 location and an electrostatic group like $C_{11}H_{21}N_2$ to the R_2 location, multiple novel compounds were developed. Table 4 displays the chemical structure of 6 newly designed compounds, as well as their inhibitory activity predicted by the CoMFA-2 model. The results of this table show that all designed compounds have high DYRK1A inhibitory activity (pEC_{50} > 8.52). These results also demonstrate the strong predictive power of the CoMFA-2 model, which can be applied to drug design or structural modification in the future.

Table 4. Chemical structures of newly DYRK1A inhibitors with their predicted pEC_{50} values

No.	R_1	R_2	R_3	Predicted (CoMFA-2)
				
Template Molecule				8.52
M1				8.85
M2				8.84
M3				8.70
M4				8.68
M5				8.61
M6				8.59

4. Conclusion

This work is focused on 3D-QSAR of 1,5-naphthyridine derivatives as effective DYRK1A inhibitors and antidiabetic agents. PLS analysis was used to develop robust CoMSIA and CoMFA models with high predictive power. Using the CoMFA focus model, it was found that there is a significant relationship between structural changes and DYRK1A inhibitory activity based on electrostatic and spatial fields. It was also found that CoMFA focusing model, produces a statistically superior model than the CoMFA and CoMSIA models in this dataset. Examination of the model features and contour maps revealed the details of the structure activity relationship, revealing conformational changes that were used to generate more active analogs. Additionally, these models were employed to forecast if newly created medications exhibit anti-diabetic properties.

References

[1] L. A. DiMeglio, C. Evans-Molina, and R. A. Oram, Type 1 diabetes. Lancet. 391 (2018) 2449-2462.

[2] S. Chatterjee, K. Khunti, and M. J. Davies, Type 2 diabetes. Lancet. 389 (2017) 2239-2251.

[3] N. Damond, S. Engler, V. R. Zanotelli, D. Schapiro, C. H. Wasserfall, I. Kusmartseva, H. S. Nick, F. Thorel, P. L. Herrera, and M. A. Atkinson, A map of human type 1 diabetes progression by imaging mass cytometry. Cell Metab. 29 (2019) 755-768.e5.

[4] E. U. Alejandro, B. Gregg, M. Blandino-Rosano, C. Cras-Méneur, and E. Bernal-Mizrachi, Natural history of β-cell adaptation and failure in type 2 diabetes. Mol. Aspects Med. 42 (2015) 19-41.

[5] Y. Dor, J. Brown, O. I. Martinez, and D. A. Melton, Adult pancreatic β-cells are formed by self-duplication rather than stem-cell differentiation. Nature. 429 (2004) 41-46.

[6] J. J. Meier, A. E. Butler, Y. Saisho, T. Monchamp, R. Galasso, A. Bhushan, R. A. Rizza, and P. C. Butler, β-Cell replication is the primary mechanism subserving the postnatal expansion of β-cell mass in humans. Diabetes. 57 (2008) 1584-1594.

[7] Y. Saisho, A. E. Butler, E. Manesso, D. Elashoff, R. A. Rizza, and P. C. Butler, β-cell mass and turnover in humans: effects of obesity and aging. Diabetes Care. 36 (2013) 111-117.

[8] S. Chera, D. Baronnier, L. Ghila, V. Cigliola, J. N. Jensen, G. Gu, K. Furuyama, F. Thorel, F. M. Gribble, and F. Reimann, Diabetes recovery by age-dependent conversion of pancreatic δ-cells into insulin producers. Nature. 514 (2014) 503-507.

[9] C. J. Lam, A. R. Cox, D. R. Jacobson, M. M. Rankin, and J. A. Kushner, Highly proliferative α-cell–related islet endocrine cells in human pancreata. Diabetes. 67 (2018) 674-686.

[10] P. A. Allegretti, T. M. Horton, Y. Abdolazimi, H. P. Moeller, B. Yeh, M. Caffet, G. Michel, M. Smith, and J. P. Annes, Generation of highly potent DYRK1A-dependent inducers of human β-Cell replication via multi-dimensional compound optimization. Bioorg. Med. Chem. 28 (2020) 115193.

[11] A. E. Butler, S. Dhawan, J. Hoang, M. Cory, K. Zeng, H. Fritsch, J. J. Meier, R. A. Rizza, and P. C. Butler, β-cell deficit in obese type 2 diabetes, a minor role of β-cell dedifferentiation and degranulation. J. Clin. Endocrinol. Metab. 101 (2016) 523-532.

[12] A. Butler, L. Cao-Minh, R. Galasso, R. Rizza, A. Corradin, C. Cobelli, and P. Butler, Adaptive changes in pancreatic beta cell fractional area and beta cell turnover in human pregnancy. Diabetologia. 53 (2010) 2167-2176.

[13] T. M. Frayling, J. C. Evans, M. P. Bulman, E. Pearson, L. Allen, K. Owen, C. Bingham, M. Hannemann, M. Shepherd, and S. Ellard, beta-cell genes and diabetes: molecular and clinical characterization of mutations in transcription factors. Diabetes. 50 (2001) S94.

[14] G. S. Eisenbarth, Type 1 diabetes: molecular, cellular and clinical immunology. (2004).

[15] H. Inadera, Developmental origins of obesity and type 2 diabetes: molecular aspects and role of chemicals. Environ. Health Prev. Med. 18 (2013) 185-197.

[16] C. M. Rondinone, Serine kinases as new drug targets for the treatment of type 2 diabetes. Curr. Med. Chem.-Immunol. Endocr. Metab. Agents. 5 (2005) 529-536.

[17] P. Wu, T. E. Nielsen, and M. H. Clausen, FDA-approved small-molecule kinase inhibitors. Trends Pharmacol. Sci. 36 (2015) 422-439.

[18] E. Dirice, D. Walpita, A. Vetere, B. C. Meier, S. Kahraman, J. Hu, V. Dančík, S. M. Burns, T. J. Gilbert, and D. E. Olson, Inhibition of DYRK1A stimulates human β-cell proliferation. Diabetes. 65 (2016) 1660-1671.

[19] T. L. Nguyen, C. Fruit, Y. Hérault, L. Meijer, and T. Besson, Dual-specificity tyrosine phosphorylation-regulated kinase 1A (DYRK1A) inhibitors: A survey of recent patent literature. Expert Opin. Ther. Pat. 27 (2017) 1183-1199.

[20] W. Becker, and W. Sippl, Activation, regulation, and inhibition of DYRK1A. FEBS J. 278 (2011) 246-256.

[21] S. Wiechmann, H. Czajkowska, K. de Graaf, J. Grötzinger, H.-G. Joost, and W. Becker, Unusual function of the activation loop in the protein kinase DYRK1A. Biochem. Biophys. Res. Commun. 302 (2003) 403-408.

[22] S. Stotani, F. Giordanetto, and F. Medda, DYRK1A inhibition as potential treatment for Alzheimer's disease. Future Med. Chem. 8 (2016) 681-696.

[23] N. Janel, M. Sarazin, F. Corlier, H. Corne, L. De Souza, L. Hamelin, A. Aka, J. Lagarde, H. Blehaut, and V. Hindié, Plasma DYRK1A as a novel risk factor for

Alzheimer's disease. Transl. Psychiatry. 4 (2014) 425-437.

[24] R. B. Kargbo, Selective DYRK1A inhibitor for the treatment of neurodegenerative diseases: Alzheimer, Parkinson, Huntington, and Down syndrome. ACS Publ. 11 (2020) 1795-1796.

[25] K. Kumar, C. Suebsuwong, P. Wang, A. Garcia-Ocana, A. F. Stewart, and R. J. DeVita, DYRK1A inhibitors as potential therapeutics for β-cell regeneration for diabetes. J. Med. Chem. 64 (2021) 2901-2922.

[26] L. Rachdi, D. Kariyawasam, V. Aïello, Y. Herault, N. Janel, J.-M. Delabar, M. Polak, and R. Scharfmann, Dyrk1A induces pancreatic β-cell mass expansion and improves glucose tolerance. Cell Cycle. 13 (2014) 2221-2229.

[27] J. Ji, H. Lee, B. Argiropoulos, N. Dorrani, J. Mann, J. A. Martinez-Agosto, N. Gomez-Ospina, N. Gallant, J. A. Bernstein, and L. Hudgins, DYRK1A haploinsufficiency causes a new recognizable syndrome with microcephaly, intellectual disability, speech impairment, and distinct facies. Eur. J. Hum. Genet. 23 (2015) 1473-1481.

[28] L. Rachdi, D. Kariyawasam, F. Guez, V. Aïello, M. L. Arbonés, N. Janel, J.-M. Delabar, M. Polak, and R. Scharfmann, Dyrk1a haploinsufficiency induces diabetes in mice through decreased pancreatic beta-cell mass. Diabetologia. 57 (2014) 960-969.

[29] P. Gehlot, R. Pathak, S. Kumar, NK. Choudhary, and VK. Vyas, A review on synthetic inhibitors of dual-specific tyrosine phosphorylation-regulated kinase 1A (DYRK1A) for the treatment of Alzheimer's disease (AD). Bioorg. Med. Chem. 14 (2024) 117925.

[30] K. Hu, L. Liu, S. Tang, X. Zhang, H. Chang, W. Chen, T. Fan, L. Zhang, B. Shen, and Q. Zhang. MicroRNA-221-3p inhibits the inflammatory response of keratinocytes by regulating the DYRK1A/STAT3 signaling pathway to promote wound healing in diabetes. Commun. Biol. 1 (2024) 300.

[31] A. Tropsha, Best practices for QSAR model development, validation, and exploitation. Mol. Inform. 29 (2010) 476-488.

[32] Z. Wan, and R. J. Sun, QSAR and molecular docking study on the biological activity of levofloxacin and thiodiazole histone deacetylase inhibitors. Chem. Rev. Lett. 3 (2020) 12-18.

[33] E. Pourbasheer, R. Aalizadeh, M. R. Ganjali, and P. Norouzi, QSAR study of alpha 1 beta 4 integrin inhibitors by GA-MLR and GA-SVM methods. Struct. Chem. 25 (2014) 355-370.

[34] R. Mahmoudzadeh Laki, and E. Pourbasheer, 3D-QSAR Modeling on 2-Pyrimidine Carbohydrazides as Utrophin Modulators for the Treatment of Duchenne Muscular Dystrophy by Combining CoMFA, CoMSIA, and Molecular Docking Studies. ACS Omega. (2024).

[35] J. Ghasemi, M. Salahinejad, and M. Rofouei, Review of the quantitative structure–activity relationship modelling methods on estimation of formation constants of macrocyclic compounds with different guest molecules. Supramol. Chem. 23 (2011) 614-629.

[36] A. Beheshti, E. Pourbasheer, M. Nekoei, and S. Vahdani, QSAR modeling of antimalarial activity of urea derivatives using genetic algorithm-multiple linear regressions. J. Saudi Chem. Soc. 20 (2016) 282-290.

[37] E. Pourbasheer, S. Vahdani, R. Aalizadeh, A. Banaei, and M. R. Ganjali, QSAR study of prolylcarboxypeptidase inhibitors by genetic algorithm: Multiple linear regressions. J. Chem. Sci. 127 (2015) 1243-1251.

[38] Z. M. Avval, E. Pourbasheer, M. R. Ganjali, and P. Norouzi, Application of genetic algorithm - multiple linear regressions to predict the activity of RSK inhibitors. J. Serb. Chem. Soc. 80 (2015) 187-196.

[39] S. Al-Sultan, M. Mohammed, and W. Talib, Newly designed 2-(aminomethyl)benzimidazole derivatives as possible tyrosine kinase inhibitors: synthesis, characterization, preliminary cytotoxic evaluation and in Silico studies. Chem. Rev. Lett., 3(2024) 545-559.

[40] A. Tropsha, O. Isayev, A. Varnek, G. Schneider, and A. Cherkasov, Integrating QSAR modelling and deep learning in drug discovery: the emergence of deep QSAR. Nat. Rev. Drug Discovery. 2 (2024) 141-55.

[41] S. Ejeh, A. Uzairu, G. A. Shallangwa, and S. E. Abechi, Computational techniques in designing a series of 1,3,4-trisubstituted pyrazoles as unique hepatitis C virus entry inhibitors. Chem. Rev. Lett. 4 (2021) 108-119.

[42] S. Alimohammadi, A. Hamidi, P. Pargolghasemi, N. Nourani, and M. S. Hoseininezhad-Namin, QSAR study of antiproliferative drug against A549 by GA-MLR and SW-MLR methods. Chem. Rev. Lett. 2 (2019) 193-198.

[43] H. Rafiei, M. Khanzadeh, S. Mozaffari, M. H. Bostanifar, Z. M. Avval, R. Aalizadeh, and E. Pourbasheer, QSAR study of HCV NS5B polymerase inhibitors using the genetic algorithm-multiple linear regression (GA-MLR). Excli J. 15 (2016) 38-53.

[44] E. Pourbasheer, A. Banaei, R. Aalizadeh, M. R. Ganjali, P. Norouzi, J. Shadmanesh, and C. Methenitis, Prediction of PCE of fullerene (C-60) derivatives as polymer solar cell acceptors by genetic algorithm-multiple linear regression. J. Ind. Eng. Chem. 21 (2015) 1058-1067.

[45] E. Pourbasheer, R. Aalizadeh, M. R. Ganjali, and P. Norouzi, Prediction of Superoxide Quenching Activity of Fullerene (C-60) Derivatives by Genetic Algorithm-Support Vector Machine. Fuller. Nanotub. Carbon Nanostruct. 23 (2015) 290-299.

[46] E. Pourbasheer, R. Aalizadeh, J. S. Ardabili, and M. R. Ganjali, QSPR study on solubility of some fullerene's derivatives using the genetic algorithms - Multiple linear regression. J. Mol. Liq. 204 (2015) 162-169.

[47] H. Kubinyi, Comparative molecular field analysis (CoMFA). The Encyclopedia of Computational Chemistry. 1 (1998) 448-460.

[48] K. H. Kim, G. Greco, and E. Novellino, A critical review of recent CoMFA applications. 3D QSAR Drug Des. (1998) 257-315.

[49] G. Klebe, U. Abraham, and T. Mietzner, Molecular similarity indices in a comparative analysis (CoMSIA) of

drug molecules to correlate and predict their biological activity. J. Med. Chem. 37 (1994) 4130-4146.
[50] M. Pastor, G. Cruciani, I. McLay, S. Pickett, and S. Clementi, GRid-INdependent descriptors (GRIND): a novel class of alignment-independent three-dimensional molecular descriptors. J. Med. Chem. 43 (2000) 3233-3243.
[51] E. Pourbasheer, R. Aalizadeh, S. S. Tabar, M. R. Ganjali, P. Norouzi, and J. Shadmanesh, 2D and 3D Quantitative Structure Activity Relationship Study of Hepatitis C Virus NS5B Polymerase Inhibitors by Comparative Molecular Field Analysis and Comparative Molecular Similarity Indices Analysis Methods. J. Chem. Inf. Model. 54 (2014) 2902-2914.
[52] E. Pourbasheer, S. S. Tabar, V. H. Masand, R. Aalizadeh, and M. R. Ganjali, 3D-QSAR and docking studies on adenosine A(2A) receptor antagonists by the CoMFA method. SAR QSAR Environ. Res. 26 (2015) 461-477.
[53] X.-J. Zou, L.-H. Lai, G.-Y. Jin, and Z.-X. Zhang, Synthesis, Fungicidal Activity, and 3D-QSAR of Pyridazinone-Substituted 1,3,4-Oxadiazoles and 1,3,4-Thiadiazoles. J. Agric. Food Chem. 50 (2002) 3757-3760.
[54] A. Agarwal, and E. W. Taylor, 3-D QSAR for intrinsic activity of 5-HT1A receptor ligands by the method of comparative molecular field analysis. J. Comput. Chem. 14 (1993) 237-245.
[55] N. Baurin, E. Vangrevelinghe, L. Morin-Allory, J.-Y. Mérour, P. Renard, M. Payard, G. Guillaumet, and C. Marot, 3D-QSAR CoMFA Study on Imidazolinergic I2 Ligands: A Significant Model through a Combined Exploration of Structural Diversity and Methodology. J. Med. Chem. 43 (2000) 1109-1122.
[56] E. Pourbasheer, and R. Aalizadeh, 3D-QSAR and molecular docking study of LRRK2 kinase inhibitors by CoMFA and CoMSIA methods. SAR QSAR Environ. Res. 27 (2016) 385-407.
[57] P. Gupta, N. Roy, and P. Garg, Docking-based 3D-QSAR study of HIV-1 integrase inhibitors. Eur. J. Med. Chem. 44 (2009) 4276-4287.
[58] H. Kubinyi, 3D QSAR in drug design: volume 1: theory methods and applications, (1993).
[59] V. N. Viswanadhan, A. K. Ghose, G. R. Revankar, and R. K. Robins, Atomic physicochemical parameters for three-dimensional structure directed quantitative structure-activity relationships. 4. Additional parameters for hydrophobic and dispersive interactions and their application for an automated superposition of certain naturally occurring nucleoside antibiotics. J. Chem. Inf. Comput. Sci. 29 (1989) 163-172.
[60] S. Wold, Cross-Validatory Estimation of the Number of Components in Factor and Principal Components Models. Technometrics. 20 (1978) 397-405.
[61] R. D. Cramer III, J. D. Bunce, D. E. Patterson, and I. E. Frank, Crossvalidation, bootstrapping, and partial least squares compared with multiple regression in conventional QSAR studies. Quant. Struct. Act. Relat. 7 (1988) 18-25.
[62] J. Romero-Parra, H. Chung, R. A. Tapia, M. Faúndez, C. Morales-Verdejo, M. Lorca, C. F. Lagos, V. Di Marzo, C. D. Pessoa-Mahana, and J. Mella, Combined CoMFA and CoMSIA 3D-QSAR study of benzimidazole and benzothiophene derivatives with selective affinity for the CB2 cannabinoid receptor. Eur. J. Pharm. Sci. 101 (2017) 1-10.
[63] E. Pourbasheer, R. Aalizadeh, and M. R. Ganjali, Molecular docking and three-dimensional quantitative structure-activity relationship studies on 5-HT6 receptor inhibitors and design of new compounds. Eurasian Chem. Commun. 5 (2022) 154-172.
[64] J. P. Stevens, Outliers and influential data points in regression analysis. Psychol. Bull. 95 (1984) 334.