Comprehensive Modeling and Molecular Docking of Phosphodiesterase Inhibitors

zahra khosravi

Golestan University

elham baher (✉ e.baher@gu.ac.ir)

Golestan University <https://orcid.org/0000-0002-7800-0612>

sajad gharaghani

University of Tehran

salma ehsani

Farhangian University

Research Article

Keywords: half maximal inhibitory concentration (IC₅₀), phosphodiesterase 10A, PDE10A, QSAR, molecular docking, multiple linear regression

Posted Date: November 9th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-957884/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Comprehensive modeling and molecular docking of phosphodiesterase inhibitors

z. khosravi¹, e. baher^{1*}, s. gharaghani², s. ehsani³

Abstract

In the present study, the biological activity of some pharmaceutical molecules was investigated and predicted by using quantitative structure-activity relationship (QSAR) studies and molecular docking. The aim of this study is to apply QSAR and molecular docking methods to predict and calculate the half maximal inhibitory concentration (IC₅₀) of phosphodiesterase (PDE) 10A. To apply QSAR method at first multiple linear regression (MLR), genetic algorithm (GA), and successive projection algorithm (SPA) were used to select the best descriptors related to PDE 10A inhibitory activity. The selected descriptors were then applied as inputs to construct MLR and a non-linear support vector machine (SVM). Also we used molecular docking method to extract the descriptors and then by using MLR and SVM methods a linear and non-linear models developed respectively. Consequently, a comparison of the results of QSAR and docking study indicated that the non-linear SVM-SPA2 in QSAR study and SVM-MLR in molecular docking study had a much better prediction power than the other models. Finally, the Y-scrambling and cross-validation tests were used to evaluate the validity of the obtained model and the results of these tests indicates that the model is appropriate for using in prediction.

Keywords: half maximal inhibitory concentration (IC₅₀), phosphodiesterase 10A, PDE10A, QSAR, molecular docking, multiple linear regression

1. Introduction

Schizophrenia is a mental disorder that requires major pharmacological therapies; therefore, a phosphodiesterase (PDE) inhibitor is suggested as a treatment strategy for this disease. Cyclic nucleotides PDE plays an important role in regulating cellular signal transduction by catalyzing the hydrolysis of cyclic adenosine monophosphate (cAMP) and cyclic guanosine monophosphate (cGMP), which are secondary intracellular messengers [1].

The function of many immune cells is also controlled by cAMP and cGMP nucleotides; thus, increasing PDE enzyme inactivates the two messengers by disrupting their function. As a result, the PDE inhibitors improve their function by elevating the levels of these two intracellular messengers. The PDE enzymes are classified into 11 families that have different properties, some of them hydrolyze cAMP and others cGMP, whereas PDE 10A hydrolyzes both cAMP and cGMP under physiological conditions [2]. Hence, this study used molecular docking studies and quantitative structure-activity relationship (QSAR) modeling to predict the activity and identify effective structural factors to design a more active drug in the treatment of schizophrenia.

So far, there are several articles relating to the design of drugs that affect schizophrenia. Prasad et al. [3] evaluated the inhibitory effectiveness of chloroacridone against calmodulin-dependent cAMP PDE1c. In their study, molecular docking was done with PDE1c for acridones to identify protein-ligand interactions.

In the study of Bhansali et al. [4], the 3D-QSAR, comparative molecular field analysis (CoMFA), comparative molecular similarity indices analysis (CoMFA), Topomer CoMFA, 2D-HologramQSAR, and molecular docking methods were evaluated for generating quantitative models to predict the behavior of pyrazolodiazepinone and identify some zones in space that improved their function. Rosendal et al. [5] developed models of the tendency to bind PDE 10A inhibitors using linear interaction energy (LIE) (also called linear responses). Abdul Amin et al [2]. studied pyrazolo (3, 4-d) pyrimidine that has PDE1 binding

Zahra khosravi
Elham Baher*(Corresponding Author)
Sajjad Gharaghani
Salma Ehsani

Email: z.khosravi972@gmail.com
Email: e.baher@gu.ac.ir
Email: s.gharaghani@aut.ac.ir
Email: salmaehsani@yahoo.com

¹ department of chemistry, faculty of science, golesan university, Gorgan, Iran

² Department of Bioinformatics, Laboratory of Chemoinformatics, Institute of Biochemistry and Biophysics, University of Tehran, Tehran, Iran

³ Department of basic science, Farhangian University, Tehran, Iran

inhibitory functions through QSAR studies. They applied the regression method including MLR, SVM, and artificial neural network (ANN) and classification methods such as bayesian modeling and recursive partitioning via Monte Carlo simulation. Halder et al. [6] investigated the structural requirements of pyrimidine-based PDE 10A inhibitors with several valid 3D-QSAR approaches. They used five 3D-QSAR tools, including CoMFA, CoMSIA, k-nearest neighbor molecular field analysis (kNN-MFA), Open3DQSAR, and Topomer CoMFA for a pyrimidine-based PDE1A inhibitor dataset. They used a non-commercial Open3DQSAR product for statistical analysis. The structural interpretations of methods that were analyzed based on the information obtained from the crystal structure and concluded, provided more detailed information on the structural requirements of this type of compound. Based on the observations of their study, 12 novel compounds were designed for better PDE 10A inhibitory activity. Goodarzi et al. [7] studied imidazo [1,5-a] pyrido [3,2-e] pyrazines of PDE10A inhibitors in QSAR modeling using a genetic algorithm (GA) and particle swarm optimization (PSO) as techniques to select the descriptors. In their study, modeling of the relationship between the selected descriptors and pIC50 activity data was performed using linear and non-linear weighted regression methods based on Euclidean and Mahalanobis distances. A stepwise MLR model with a limited number of quantum descriptors was also used. The results indicated that Euclidean non-linear weighted regression method on the descriptors selected by the PSO algorithm had the best ability to predict.

Dym et al. [8] proposed a structural model for the PDE4 substrate (cAMP) and a specific PDE4 inhibitor (rolipram) within the active site of the enzyme. In addition, these docked models were applied to test the structural results of mutations that guarantee resistance to rolipram and other specific PDE4 inhibitors. Fossa et al [9] performed a study on the structural basis of PDE3 selective inhibition using molecular docking. In this study, based on the available experimental data, a theoretical model with respect to the catalytic site of PDE3A was developed using homology-modeling techniques. Chidochangu et al. [10] conducted a docking study on PDE4B ligands and analyzed the behavior of selected scoring functions. The aim of their study was to identify a scoring method capable of providing a 4-fold enrichment in a screening sample in a random selection process. To do so, Ligand Fit, a fast-form docking algorithm, was used to dock various inhibitors of PDE4B into its binding site. Fernanda et al. [11] conducted a study on molecular docking and developed an empirical binding-free energy model for PDE 4 inhibitors. They applied many computational methods to develop a model to predict the PDE4B inhibitory activity. The suitability of using the ligand docking approach and maintaining enzyme rigidity in the study of a number of PDE4 inhibitors was confirmed by previous analyses of the molecular dynamics of the complete enzyme. Moreover, a comprehensive docking process was carried out to identify the most movable site of ligand binding to the enzyme, which includes the active regions of metal ions and structural molecules of the surrounding water. Kang et al. [12] performed docking and 3-D QSAR studies for PDE4 and PDE7 inhibitors. They applied CoMFA to fulfill target-based docking and a 3-D QSAR study in which 3 compounds were synthesized, and their inhibitory activity was predicted using the 3-D QSAR model. The effect of a new derivative of pyrazole (LQFM 021) on PDE was evaluated by Ramos et al. [13] in their study, the synthesis, docking, pharmacological and toxic activity of this novel compound was evaluated to investigate its possible effect on PDE. Cicher et al. [14] conducted a study on probing the PDE11 catalytic site. Docking studies, homology modeling, and molecular dynamic simulations were used through graphical processing unit architecture.

Given the success of the QSAR method, the current researchers decided to create a comprehensive model to predict and explain the factors affecting the IC50 of PDE-10A inhibitors. For this purpose, the linear and non-linear modeling methods were developed in two parts based on linear (MLR, SPA) and non-linear (GA) variable selection methods. The first part involved QSAR modeling of the inhibitory activity of PDE 10A inhibitors based on the ligand method. The second part was based on molecular docking studies and the development of the QSAR model on their obtained descriptors.

In QSAR modeling, direct calculation of the molecular activity is not possible, because the molecular descriptors would be computed, and finally, a predictive model can be developed by measuring their information [15, 16]. Generally, drug design methods are divided into two categories based on QSAR techniques, Ligand-based methods and Protein-based methods.

Up to now, most of the QSAR studies have been conducted using ligand-based methods. In this method, a large set of molecules with similar structures that have the same functional mechanism and distinct biological activity are studied. In protein-based QSAR methods, knowledge of the structure of the target protein, and in particular, the binding site of the protein for drug design are important because accurate information on drug-protein interaction at the protein binding site can be very effective in drug design. Some drugs that effectively interact with the target protein binding site can be designed by having information on

drug structure, protein structure and protein binding site. The study of drug-protein interactions using this information is called molecular docking.

For developing the comprehensive QSAR model, we used both ligand-based and protein-based methods.

Materials and Methods

Data set

The data set was chosen from Malamas et al. [1]. The data of the current study included 42 compounds of PDE 10A whose inhibitory activity was exhibited by IC_{50} and its range was from 0.28 to 2430. Fig. 9 and fig. 10 represent the basic structure of derivatives of PDE 10A inhibitors and the corresponding structural specification and their activities represent in Table 7 and Table 8, respectively.

In this study, IC_{50} was converted to a logarithmic scale and used as $\log IC_{50}$. The Y-ranking method was used to divide the data series. To do this, the data were first ranked ascending or descending according to $\log IC_{50}$ values so that 80% of the data series were in the training group and the other 20% in the experimental group. At first, the structure of the compounds was plotted using *Hyperchem* software and their three-dimensional (3D) structure was optimized using semi-empirical AM1 method [17].

Fig. 9 Basic structure of derivatives of PDE 10A inhibitors

Fig. 10 Basic structure of derivatives of PDE 10A inhibitors

Table 7 Complete structural characteristics of PDE 10A inhibitors and their empirical activity

Compound	R^2	R^4	R^9	IC_{50} (nM)
1	OMe	H	4-Me-3-pyridyl	388
2	OMe	H	5-Me-3-pyridyl	51.5
3	OMe	H	5-OMe-3-pyridyl	91.3
4	OMe	H	2,4-Me, Me-thiazole	64.5
5	OMe	H	2-Me-Ph	47.6
6	OMe	H	4-F, 2-Me-Ph	45.9
7	OMe	H	2-Cl-Ph	13.1
8	OMe	H	2,4-Cl, Cl-Ph	15
9	OMe	H	2-Cl, 4-F-Ph	13.1
10	OMe	H	2,5-Cl, Cl-Ph	53.1
11	OMe	H	2-F, 5-OMe-Ph	9.83
12	OMe	H	2-F, 5-OEt-Ph	114
13	OMe	H	2-F, 5-O(CHMe ₂)-Ph	2430
14	H	OMe	4-Me-3-pyridyl	42.4
15	H	OMe	3-Me-4-pyridyl	10.1

16	H	OMe	2-Me-3-pyridyl	12.4
----	---	-----	----------------	------

Table 8 Complete structural characteristics of PDE 10A inhibitors and their empirical activity

Compound	R^1	R^6	R^8	IC50 (nM)
17	2-Me-Ph	OMe	H	3.3
18	2-Cl-Ph	OMe	H	2.41
19	2-Cl-Ph	OMe	OMe	0.28
20	2-Me-Ph	OMe	OMe	5.13
21	2-CF ₃ -Ph	OMe	OMe	3.35
22	2-Me-3-pyridyl	OMe	OMe	1.48
23	4-Me-3-pyridyl	OMe	OMe	4.16
24	3-Me-4-pyridyl	OMe	OMe	2.06
25	2-Me-Ph	OMe	morpholine	1.53
26	2-Cl-Ph	OMe	morpholine	0.85
27	3-Me-4-pyridyl	OMe	morpholine	3.99
28	2-OMe-Ph	OMe	F	6.53
29	3-OMe-Ph	OMe	F	9.11
30	3-F, 2-Me-Ph	OMe	F	3.55
31	4-Me-3-pyridyl	OMe	F	25.4
32	3-Me-4-pyridyl	OMe	F	7.2
33	2-Me-3-pyridyl	OMe	F	7.84
34	3,5-Me, Me-N-Me-pyrazole	OMe	F	13.1
35	3-Me-2-thienyl	OMe	F	1.65
36	2,4-Me, Me-thiazole	OMe	F	11.7
37	3-Me-4-pyridyl	OMe	CF ₃	1.48
38	4-Me-3-pyridyl	OMe	CF ₃	5.11
39	3-F, 2-Me-Ph	OMe	CF ₃	1.85
40	3-Me-4-pyridyl	OMe	OCF ₂	7.0
41	2-OMe-Ph	F	OMe	8.33
42	3-Me-2-thienyl	F	OMe	9.43

Descriptors Calculation

After drawing and optimizing the structure of the compounds using *HyperChem* software, the output of this software was stored with the HIN suffix to calculate suitable descriptors, and then was used as inputs in *PaDEL*-Descriptor [18] and *AutoDock* [19] softwares, separately. The obtained output was also stored once again with the Zmt suffix used as input in *WinMOPAC* software [1].

The first part, ligand-based model, contained a large number of descriptors obtained by the *PaDEL* and *WinMOPAC* that, the most important descriptors depended to the inhibitory activity of PDE 10A inhibitors were selected in the next step.

At first, the descriptors whose all values related to the compounds were zero or constant number were deleted. The pair of descriptors with a correlation coefficient above 0.8 was classified as internal correlation and one of them was omitted. Finally, variable selection methods including MLR, GA and SPA were used to select the best descriptors among the remaining 130 descriptors.

The stepwise MLR method was applied to obtain suitable descriptors for the MLR model using SPSS software. MATLAB software was then used to perform the variable selection process with GA and SPA. The selected descriptors are shown in Tables 9 through 12.

Table 9 Descriptors calculated by the MLR method		
Variable No.	Descriptors	Definition
V1	ATSC1c	Centered Broto-Moreau autocorrelation-lag1/weighted by charges
V2	ATSC4v	Centered Broto-Moreau autocorrelation-lag4/weighted by van der Waals volumes
V3	AATSC6s	Average centered Broto-Moreau autocorrelation-lag6/weighted by I-state
V4	GATS3c	Geary autocorrelation-lag3/weighted by charges
V5	GATS6v	Geary autocorrelation-lag6/weighted by van der Waals volumes
V6	SPMAX5_Bhv	Largest absolute eigenvalue of Burden modified matrix-n5/weighted by relative van der Waals volumes
V7	TDB6u	3D topological distance-based autocorrelation-lag6/without weight

Table 10 Descriptors calculated by the GA method		
Variable No.	Descriptors	Definition
V1	ALOGP	Logarithm K of water and alcohol Ghose-Crippen
V2	Nn	Number of nitrogen atoms
V3	ATS8s	Broto-Moreau autocorrelation-lag4/weighted by I-state
V4	AATS8i	Average centered Broto-Moreau autocorrelation-lag8/weighted by first ionization potential
V5	AATSOs	Average centered Broto-Moreau autocorrelation-lag0/weighted by I-state
V6	AATS1s	Average centered Broto-Moreau autocorrelation-lag1/weighted by I-state

Table 11 Descriptors calculated by the SPA1 method		
Variable No.	Descriptors	Definition
V1	ATSOm	Broto-Moreau autocorrelation-lag0/weighted by mass
V2	ATSC7m	Centered Broto-Moreau autocorrelation-lag7/weighted by mass
V3	ATSC8m	Centered Broto-Moreau autocorrelation-lag8/weighted by mass
V4	ATSC7v	Centered Broto-Moreau autocorrelation-lag7/weighted by van der Waals volumes
V5	VR1_Dzv	Randic-like eigenvector-based index from Barysz matrix weighted by van der Waals volume
V6	SIC3	Structural information content index
V7	VR1_D	Randic-like eigenvector-based index from topological distance matrix

Table 12 Descriptors calculated by the SPA2 Method		
Variable No.	Descriptors	Definition
V1	ATSC4m	Centered Broto-Moreau autocorrelation-lag4/weighted by mass
V2	ATSC2s	Centered Broto-Moreau autocorrelation-lag2/weighted by I-state
V3	GATS7c	Geary autocorrelation-lag7/weighted by charges
V4	GATS7i	Geary autocorrelation-lag7/weighted by first ionization potential
V5	GATS1s	Geary autocorrelation-lag1/weighted by I-state
V6	VR1_Dzv	Randic-like eigenvector-based index from Barysz matrix weighted by van der Waals volume
V7	RNCS	Area of relative negative charged surface

In the second part, after docking calculations, the best configuration obtained by docking for each molecule with the most favorable energy (the most negative bonding energy) was used to compute the

molecular descriptors. After each docking process, *AutoDock* was able to calculate eight types of energy for per molecule. These energies are as follows.

1. Estimated free energy of binding (E_{FreeBind})
2. Estimated inhibition constant (K_i)
3. Van der Waals energy (E_{vdw})+hydrogen bond (Hbond)+ desolvation (desolv): (E_{VHD})
4. Electrostatic energy (E_{Elec})
5. Final intermolecular energy (E_{InterMol}) which includes E_{vdw} , Hbond, desolv and E_{Elec} .
6. Final total internal energy (E_{FTot})
7. Torsional free energy (E_{Tors})
8. Unbounded system energy (E_{Unb})

Next, the most important descriptors were selected using MLR. These descriptors included E_{FreeBind} , E_{Elec} , E_{FTot} and E_{Tors} .

QSAR Model development

QSAR models indicate a relationship between the structural descriptors of compounds and their biological activity and were applied as one of the main computational tools in pharmaceutical chemistry.

These models consist of a single dependent variable (Y) and independent variables (X_{ij}) that are descriptors. In the first section of the present study, modeling was performed using QSAR statistical methods. To do so, MLR, GA and SPA were used as variable selection methods as well as MLR and SVM were applied as linear and non-linear modeling methods.

Multiple linear regression (MLR)

A regression relation called MLR was used to investigate the effect of one or more variables (independent variables) on one variable (dependent variable).

In the MLR method, a mathematical equation was developed which equation (3) shows the general form of these linear models [21].

$$y = a_0 + a_1x_1 + a_2x_2 + \dots + a_nx_n \quad (3)$$

In this equation, a_0 , a_1 , and a_n are constant and coefficient of variables, respectively. MLR model was used to predict the inhibitory activity of PDE inhibitors using SPSS software. The best model of all obtained MLR models to predict LogIC₅₀, is expressed in equation (4).

$$\log\text{IC}_{50} = 37.005 - 6.185(V_1) + 0.002(V_2) + 1.475(V_3) + 4.139(V_4) + 1.059(V_5) - 3.532(V_6) - 5.790(V_7) \quad (4)$$

Genetic algorithm (GA) in chemistry

For the first time, in the 1970s, a scientist from the University of Michigan, John Holland, proposed the use of a GA. The application of GA in QSAR modeling is to search among descriptors to find the most effective ones for modeling based on Darwin's law (LAW??). The GA is often a good option for regression-based prediction techniques. GA as a computational optimization algorithm effectively searches for different areas of the answer space considering a set of answer space points within each computational iteration [22].

In the current study, GA was used as a non-linear variable selection method to find the most effective descriptors. Then, the relationship between LogIC₅₀ and descriptors selected by GA was evaluated using MLR modeling. The best model obtained by this method is given in equation (5).

$$\log\text{IC}_{50} = 8.817 + 0.181(V_1) + 0.764(V_2) + 0.005(V_5) - 0.072(V_6) + 0.384(V_7) + 0.056(V_8) \quad (5)$$

Successive projection algorithm (SPA)

The SPA is a technique specifically designed to select subgroups of variables with low co-linearity to improve the suitability of MLR models. The SPA method has three steps as follows:

Step 1: The algorithm at least constructs a set of candidates from a set of variables based on the co-linearity criterion.

Step 2: At this stage, the best candidates are selected based on the ability criterion to predict the MLR model.

Step 3: From the best selection set, the variables which have no significant effect on decreasing predictive ability are eliminated [23].

Then, according to the previous steps, the MLR modeling was done using descriptors obtained by SPA method. Here, there were two sets of descriptors (SPA1 and SPA2). The best obtained model using SPA1

descriptors represents by equation (6) and Finally, the best obtained model using SPA2 descriptors represents by equation (7).

$$\log IC_{50} = -23.516 + 0.000(V_1) - 0.001(V_2) - 0.001(V_3) + 2.46 \times 10^{-5}(V_4) + 2.35 \times 10^{-5}(V_5) + 27.680(V_6) + 0.000(V_7) \quad (6)$$

$$\log IC_{50} = 2.570 - 0.001(V_1) - 0.009(V_2) + 2.195(V_3) - 0.924(V_4) + 0.863(V_5) - 4.5 \times 10^{-5}(V_6) - 0.485(V_7) \quad (7)$$

Support vector machine (SVM)

SVM is one of the supervised learning methods used for classification and regression. The foundation of the SVM was invented by Vladimir Vapnik and the SVM was highly regarded for its remarkable features and experimental performance [24, 25]. Furthermore, SVM has found wide application in solving non-linear regression calculations.

In the SVM method, the input data are first mapped to a higher-dimensional feature space via the kernel function, and next, the linear regression is performed in this new feature space.

Mapping the variables allows for the solving of non-linear problems in a linear space. A radial basis function (RBF) used in the present study is the most common function in SVM. This function is estimated by minimizing the regularized risk function $R(c)$ which is shown by equation (8).

$$R(C) = C \frac{1}{N} \sum_{i=1}^N L_{\varepsilon}(d_i, y_i) + \frac{1}{2} \|W\|^2 \quad (8)$$

R_{SVM} represents the adjusted error function In Equation (8) and $C \frac{1}{n} \sum_{i=1}^n L_{\varepsilon}(d_i, y_i)$ stands for the experimental error measured by $L_{\varepsilon}(d_i, y_i)$ as a loss function and shows that there will be no error in values less than ε . $\frac{1}{2} \|w\|^2$ indicates the complexity of model and C is an adjusted constant which compromise between experimental error and model complexity.

d_i is tube length and y_i represents the distance of $-\varepsilon$ to $+\varepsilon$ around regression line and determines the accuracy of the training data set, and the data inside of this hypothetical tube are not considered as errors, and the data outside of this range are considered as error.

Molecular docking

Molecular docking studying is a powerful computational tool for predicting the ligand binding state on the active site of a protein. Therefore, after the docking process, a suitable configuration of the ligand is obtained on the active site of the protein and is used to calculate molecular descriptors.

The difference between this type of protein-based modeling and ligand-based modeling is that with the ligand-based, the target molecule structure is considered as a single molecule while in the first one, the ligand structure is considered in association with be the protein, and the results of the calculations are affected by the interactions between the ligand and protein.

In this section, the structure of the compounds was plotted using *HyperChem* and their 3D structure was optimized using semi-empirical AM1 method. The output was then stored with the HIN suffix and used as input for molecular docking studies.

The X-ray crystal structure of human protein of PDE 10A encoded by 4FCB is shown in fig. 11, taken from the Protein Data Bank in PDB format. The PDB file obtained from 4FCB had a resolution of 1.2 and an R value of 0.24

Fig. 11 Crystal structure of human protein of PDE 10A encoded by 4FCB

Prior to docking, hydrogens which are not commonly seen in the crystallographic structure must be added to the structure. The water molecules associated with the molecule or ions should also be removed. The status of all rings, side chains and amino acids should be examined.

The files prepared for docking should only contain the protein PDB file alone and PDB ligand file alone.

In the present study, *AutoDock* software was used for molecular docking calculations. For docking calculation, the ligand and the protein was considered as flexible and rigid form, respectively. Given this state, the ligand was inserted into the active site of the PDE 10A protein.

To find the active site of PDE 10A protein, the information of Lu et al.'s article in 2010 was used [26]. To further find the active site of the PDE 10A protein, the OT7 complex of ligand fig. 12 with this protein fig. 13 was removed from the X-ray crystallographic structure of the PDE 10A. Docking with this ligand was again performed to confirm the docking process.

The process was conducted to examine whether the OT7 ligand finds its way back to its previous site by docking again. After performing the re-dock process and searching for all possible sites in the PDE 10A protein, in according with fig. 14, the OT7 ligand again returned to its previous site with little displacement that indicated a successful docking.

Thus, by achieving a precise docking, each of the studied molecules can be docked to the PDE 10A protein.

Fig. 12 Structure of ligand complexed with PDE 10 A (PDBID: N7F).

Fig. 13 Complex structure of PDE 10 A and N7F

Fig. 14 Adaptive image of ligand before and after Redock

Knowing the binding site of the PDE 10A protein, a 94×76×116 network with a distance of 0.5519 was defined around the protein binding site for generating receptor network files.

The Lamarkian genetic algorithm (LGA) search method available in *AutoDock* software was used to find the best ligand-protein binding site [28].

Subsequently, after determining the appropriate descriptors by *AutoDock*, and all the obtained MLR models to predict LogIC₅₀ drugs, the best model was calculated by Equation (9). The R, SE and F values in this model are equal to 0.38, 0.72 and 1.56, respectively.

$$\text{LogIC}_{50} = 0.407 - 0.034(\text{EFreeBind}) + 1.232(\text{EElec}) + 0.899(\text{EFTot}) + 1.262(\text{ETors}) \quad (9)$$

In the current study, LigPlot was applied to demonstrate the hydrogen bonding and hydrophobic interactions of the studied pharmaceutical compounds with the PDE 10A protein. In this study, the compounds with the highest activity (compound 13) and with the lowest activity (compound 19) were evaluated. These interactions are illustrated in fig. 15 through fig. 18.

Fig. 15 Two-dimensional LigPlot representation of hydrogen bond interactions between compound 13 and PDE 10A protein amino acids at the binding site

Fig. 16 Two-dimensional LigPlot representation of hydrophobic interactions between compound 13 and PDE 10A protein amino acids at the binding site

Fig. 17 Two-dimensional LigPlot representation of hydrogen bonding interactions between compound 19 and PDE 10A protein amino acids at the binding site

Fig. 18 Two-dimensional LigPlot representation of hydrophobic interactions between compound 19 and PDE10A protein amino acids at the binding site

Results and discussion

The first aim of the current study was to present a QSAR model to predict and calculate IC_{50} of PDE 10A. The test series was used to measure the predictive power of models made by MLR and SVM methods. The statistical parameters of the MLR models created for the training and test series at their best mode are presented in table 1, which shows that the best model is the MLR-MLR method among the linear ones.

Table 1 Statistical parameters of MLR models						
Model	Training Set			Test Set		
	R	SE	F	R	SE	F
MLR-MLR	0.93	0.32	23.08	0.75	0.46	6.34
MLR-GA	0.65	0.65	2.42	0.60	0.57	2.83
MLR-SPA1	0.83	0.49	8.31	0.63	0.50	3.37
MLR-SPA2	0.63	0.64	2.60	0.45	0.78	1.26

Selected descriptors by MLR, GA, and SPA methods were separately used as SVM inputs. Statistical parameters obtained from SVM modeling are reported in table 2. The results clearly indicated that the SVM-SPA2 model had a higher correlation coefficient than the other models. Thus, the SVM-SPA2 model with good values of R and SE had more power and validity to model and predicted the inhibitory activity of PDE 10A inhibitors of these compounds. The predicted values of SVM, experimental and residual $LogIC_{50}$ for the best model are reported in table 3. Finally, the cross-validation test was used to evaluate the validity of this model and the Q2 and PRESS statistics were 0.88 and 1.11, respectively.

Table 2 Statistical parameters of SVM models				
Model	Training Set		Test Set	
	R	SE	R	SE

SVM-MLR	0.91	0.28	0.89	0.47
SVM-GA	0.99	0.12	0.55	0.63
SVM-SPA1	0.80	0.45	0.78	0.45
SVM-SPA2	0.99	2.45×10^{-4}	0.91	0.35

Table 3 Experimental, predictive and residual logIC₅₀ values of the best model

Number	Experimental values	QSAR Residual values	QSAR predicted values	Docking Residual values	Docking predicted values
1	2.5888	0.2914	2.2974	1.6351	0.9537
2	1.7118	0.0001	1.7117	0.0309	1.6809
3	1.9605	0.0001	1.9604	0.0264	1.9341
4	1.8096	0.5535	1.256	0.3337	1.4758
5	1.6776	0.0004	1.6772	0.1121	1.5655
6	1.6618	0.0002	1.6616	0.579	1.0828
7	1.1173	-0.6445	1.7617	-0.1811	1.2984
8	1.1761	0.0002	1.1759	0.6387	0.5374
9	1.1173	0	1.1173	0.0368	1.0805
10	1.7251	0.0001	1.725	0.5236	1.2015
11	0.9926	-0.0002	0.9928	-0.0222	1.0147
12	2.0569	0.1304	1.9265	0.7155	1.3414
13	3.3856	0.0004	3.3852	0.7509	2.6347
14	1.6274	0.0003	1.6271	-0.9612	2.5886
15	1.0043	0.0002	1.0042	-0.0083	1.0126
16	1.0934	0	1.0934	-0.039	1.1324
17	0.5185	0.1672	0.3513	-0.4121	0.9306
18	0.382	0	0.3821	-0.1446	0.5266
19	-0.5528	-0.0006	-0.5523	-0.9556	0.4027
20	0.7101	0.0005	0.7096	0.4098	0.3003
21	0.525	0	0.525	-0.091	0.616
22	0.1703	-0.1292	0.2994	-0.3534	0.5237
23	0.6191	-0.0001	0.6192	0.1949	0.4242
24	0.3139	-0.0005	0.3143	-0.5475	0.8613
25	0.1847	0.0001	0.1846	-0.5494	0.734
26	-0.0706	-0.0007	-0.0699	0.2894	-0.36
27	0.601	0.1655	0.4354	-0.0266	0.6276
28	0.8149	0	0.815	0.0025	0.8124
29	0.9595	0	0.9595	-0.4688	1.4283
30	0.5502	0.0001	0.5502	-0.3222	0.8725
31	1.4048	0.0002	1.4046	0.9353	0.4696
32	0.8573	0.2938	0.5636	0.4015	0.4558
33	0.8943	-0.0001	0.8944	0.113	0.7813
34	1.1173	0	1.1173	0.5478	0.5695
35	0.2175	-0.0002	0.2177	-0.0111	0.2286
36	1.0682	0	1.0682	-0.0415	1.1097
37	0.1703	-0.0001	0.1704	-0.6096	0.7798
38	0.7084	0.3259	0.3826	-0.6829	1.3914
39	0.2672	-0.0001	0.2673	-0.443	0.7102
40	0.8451	-0.0001	0.8452	0.181	0.6641
41	0.9206	-0.0001	0.9208	0.3444	0.5763
42	0.9745	0.0002	0.9743	0.039	0.9743

The graph of the predictive LogIC₅₀ values obtained by the SVM-SPA2 model is plotted according to the experimental values in fig. 1, representing a good correlation between the prediction and experimental results. Moreover, the difference between the experimental values and the predictions (residuals) is plotted in fig. 2 in

terms of experimental values so that the uniform and random distribution of residuals around the zero axis demonstrates that there is no systematic error in the SVM-SPA2 model.

Fig. 1 Predicted values graph based on experimental values in SVM-SPA2 model

Fig. 2 Residual values graph based on experimental values in SVM-SPA2 model

The second part of the present study, docking descriptors and the test data set were used to measure the predictive power of models made by MLR method. The statistical parameters of the model created for the training and test series at its best mode are presented in table 4. In continuous tests, four descriptors selected by the MLR method were used as SVM inputs. Also Table 4 illustrates the statistical parameters obtained by the SVM model, representing the superiority of the SVM model over the MLR model.

Table 4 Statistical parameters of MLR and SVM models				
Model	Test set		Training set	
	SE	R	SE	R
MLR-MLR	0.70	0.34	0.74	0.40
SVM-MLR	0.69	0.75	0.35	0.88

The results represents in table 4 shows that there is a non-linear relationship between the molecular descriptors of the compounds and PDE 10A enzyme inhibitors. Hence, SVM models with good values of R and SE had more power and validity to predict the inhibitory activity of PDE 10A inhibitors of these compounds.

Subsequently, the cross-validation test was applied to investigate the generality and validity of the constructed SVM model as well as the statistical parameters of Q2 and PRESS was 0.3 and 0.8, respectively. The predicted values of SVM, experimental and residual LogIC50 for the best model are expressed in table 3.

The graph of the predicted LogIC50 values obtained by the SVM-MLR model is plotted according to the experimental values in fig. 3, indicating a good correlation between the prediction and experimental results.

Furthermore, the difference between the experimental values and the predictions (residuals) is plotted in fig. 4 in terms of experimental values so that the uniform and random distribution of residuals around the zero axis exhibits that there is no systematic error in the SVM model.

Fig. 3 Predicted values graph based on experimental values in SVM-MLR model

Fig. 4 Residual values graph based on experimental values in SVM-MLR model

The Y-scrambling test resulting from the randomization of the model response was done to ensure that there was no random correlation in the data matrix. In the first part, the R average value after 20 times of the Y-scrambling test presented in table 5 is 0.38, representing that finding the probability of chance correlation is very low in the data matrix.

Table 5 Y-scrambling test of SVM-SPA2 model					
Replication	SE	R	Replication	SE	R
1	0.77	0.30	11	0.70	0.5
2	0.76	0.34	12	0.77	0.3
3	0.79	0.24	13	0.78	0.26
4	0.75	0.37	14	0.68	0.53
5	0.74	0.40	15	0.70	0.50
6	0.74	0.40	16	0.76	0.34
7	0.78	0.26	17	0.71	0.48
8	0.75	0.39	18	0.74	0.40
9	0.75	0.37	19	0.73	0.42
10	0.72	0.45	20	0.74	0.40

In the second part, the R average value after 20 times of the Y-scrambling test given in table 6 is 0.29, indicating that finding the probability of chance correlation is very low in the data matrix.

Table 6 Y-scrambling test of SVM- MLR model					
Replication	SE	R	Replication	SE	R
1	0.75	0.27	11	0.65	0.54
2	0.69	0.45	12	0.72	0.39
3	0.75	0.25	13	0.76	0.22
4	0.76	0.16	14	0.75	0.26
5	0.76	0.20	15	0.75	0.25
6	0.77	0.13	16	0.73	0.33

7	0.76	0.21	17	0.75	0.26
8	0.68	0.48	18	0.72	0.38
9	0.76	0.20	19	0.76	0.21
10	0.75	0.23	20	0.74	0.31

Applicability domain analysis

An important step in developing the QSAR model is to determine the applicability domain (AD) of this model. In the QSAR study, the AD is validation of the biological, structural or physico-chemical space [28].

The AD is delineated by calculating the leverage threshold (h^*) for each compound whose activity is predicted by the model. The leverage is the measurement of distance from the center of the training series. Using the leverage, it is possible to know whether or not a molecule falls within the structural domain of the model. Reliable measurement of the AD of this model through leverage (h_i) is defined below:

$$h_i = x_i (X^T X)^{-1} x_i^T \quad (i = 1, 2, \dots, n) \quad (1)$$

Where x_i is the descriptor raw vector of targeted molecule and X is the matrix of the $n \times k$ obtained from the descriptor values of the training set (n is the number of training set molecules and k is the number of model descriptors).

The critical value of the threshold h^* is generally fixed at $3p/n$ where p is the number of model and n is the number of training set compounds [29]. The AD of the model presented for the first part of the study is shown in fig. 5.

Fig. 5 AD graph of the SPA2-SVM model

The black horizontal lines represent and the vertical dashed line displays the critical value of the threshold h^* equal to 0.73. As shown in fig. 5, the molecule 36 is out of, called outlier data and this may be due to an error in the experimental response.

The AD of the model presented for the second part of the study is illustrated in fig. 6.

Fig. 6 AD graph of MLR-SVM model

The horizontal lines represent and the vertical dashed line represents the critical value of the threshold h^* equal to 0.43. As represented in fig. 6, there are no outlier data in the dataset. As a result, the obtained SVM model could be used with a high degree of validity in this AD model.

Descriptor Interpretation

Sensitivity analysis is used to find the most important descriptors that have good correlation with the biological activity of the molecules. The principle of sensitivity analysis is to eliminate the sequence of descriptors in the model. After each descriptor has been sequentially removed, a new model is constructed and the root mean square error ($RMSE_i$) is calculated for the model. Differences between $RMSE$ for the QSAR model and $RMSE_i$ of omitted from the trained model indicate the importance of i_{th} input descriptor.

$$Rmdiff_i = RMSE_i - RMSE \quad (2)$$

In the first part of the current study, the GATS7i of Geary autocorrelation descriptors (GATS) was more important than the other descriptors. The importance of the other descriptors of this model is shown in fig. 7.

Thus, , the least importance was related to the relative negative charged surface (RNCS) descriptor, which belongs to the family of *electronic* charged partial surface area (CPSA) descriptors. Some of these descriptors are related to autocorrelation type of topological descriptors, and others are associated with electron descriptors so that they combine structural and electronic information to identify molecules and indicate the factors responsible for showing the polar interactions between molecules.

In the second part, the most effective descriptor was associated with $E_{FreeBind}$. The importance of the other descriptors of this model is presented in fig. 8.

Fig. 7 Differences between RMSE and $RMSE_i$ for QSAR model

Fig. 8 Differences between RMSE and $RMSE_i$ for docking model

The least importance was related to the E_{Tors} descriptors. Binding free energy (quantum mechanics descriptor) is the stretching criterion for ligand-protein complex, which is equal to the difference between the energy of the complex and the sum of the free energies of each molecule separately. (Kollman et al. [30] , Arba et al. [31] , Balasubramanian et al. [32]). Also the role of electrostatic energy is important in understanding intramolecular interaction, because they are long-term and affect charged molecules (Maleki et al., [33]). Another descriptor is final total internal energy (quantum mechanics descriptor). Internal energy is considered as a receptor (docking descriptors), which is calculated in the receptor to create a receptor model. And torsional free energy (quantum mechanics descriptor) is associated with the total number of

rotatable bonds. It also depends on the number of torsions of each molecule and how they are arranged and positioned in a site.

Conclusion

The aim of this study was to investigate the QSAR relation to model and predict the inhibitory activity of PDE 10A inhibitors. In all parts of the current study, efforts were made to develop a proper model for predicting and calculating IC₅₀ of PDE 10A inhibitors to provide potential drugs for the treatment of schizophrenia.

In the first section of the present study, the inhibition activity of 42 PDE 10A inhibitors was predicted using QSAR modeling. In this section, stepwise MLR, GA and SPA were used as variable selection methods as well as MLR and SVM were applied as linear and non-linear modeling methods. The results clearly suggested the superiority of the SVM-SPA2 model over the rest of the models and proved there was a non-linear relationship between the molecular descriptors and inhibitory activity of PDE 10A enzyme.

Hence, the SVM-SPA2 model with good values of R and SE had more power and validity to model and predicted inhibitory activity of PDE 10A inhibitors of these compounds.

The second aim of the present study was to investigate the effective ligand-receptor interactions and to calculate the drug-protein E_{FreeBind} using a combined docking and QSAR method. Like the previous section, the inhibition activity of 42 PDE 10A inhibitors was predicted in this section. The used descriptors were obtained using AutoDock software. The most important descriptors which were associated with the inhibitory activity of PDE 10A inhibitors and provided a correct understanding of their activity were selected by the variable selection method.

Further investigations indicated the good ability of the obtained descriptors to encode different properties of the molecules.

The QSAR studies were performed using linear modeling (MLR) and non-linear (SVM) methods; the results demonstrated that the SVM model was superior to MLR model and proved that there was a non-linear relationship between the molecular descriptors and inhibitory activity of PDE 10A enzyme.

Declarations

Funding: Not applicable

Conflicts of interest/Competing interests: Not applicable

Availability of data and material: All used data and materials were transparent and available for our research. The raw dataset extract from Malamas et al. [1]. Also the dataset included 42 compounds of PDE 10A whose inhibitory activity was exhibited by IC₅₀ and its range was from 0.28 to 2430.

Code availability: The software packages used in this research are *SPSS 21*, *HyperChem version 7.0*, *Padel-descriptor version 2.21*, *WinMopac version 2.0*, *Arguslab version 4.0.1*, *LigPlus version 1.4.5* and *Aoutodock version 1.5.6* and also we used coded software in *MATLAB version 2014a* for applying GA, SPA and SVM algorithms.

Authors' contributions: the author's contributions are as follows:

Zahra Khosravi: data analysis, applying methods, used software to achieve research goals and get final results.

Elham Baher: Main adviser and manuscript writer

Sajjad Gharaghani: research reviser

Salma Ehsani: research reviser

References

1. Malamas, M. S., Stange, H., Schindler, R., Lankau, H. J., Grunwald, C., Langen, B., ... & Hoefgen, N. (2012). Novel triazines as potent and selective phosphodiesterase 10A inhibitors. *Bioorganic & medicinal chemistry letters*, 22(18), 5876-5884.
2. Amin, S. A., Bhargava, S., Adhikari, N., Gayen, S., & Jha, T. (2018). Exploring pyrazolo [3, 4-d] pyrimidine phosphodiesterase 1 (PDE1) inhibitors: a predictive approach combining comparative validated multiple molecular modelling techniques. *Journal of Biomolecular Structure and Dynamics*, 36(3), 590-608.
3. Prasad, V. R., Reddy, G. D., Appaji, D., Peters, G. J., & Mayur, Y. C. (2013). Chemosensitizing acridones: In vitro calmodulin dependent cAMP phosphodiesterase inhibition, docking, pharmacophore modeling and 3D QSAR studies. *Journal of Molecular Graphics and Modelling*, 40, 116-124.
4. Bhansali, S. G., & Kulkarni, V. M. (2014). Combined 2D and 3D-QSAR, molecular modelling and docking studies of pyrazolodiazepinones as novel phosphodiesterase 2 inhibitors. *SAR and QSAR in Environmental Research*, 25(11), 905-937.
5. Kjellgren, E. R., Glue, O. E. S., Reinholdt, P., Meyer, J. E., Kongsted, J., & Poongavanam, V. (2015). A comparative study of binding affinities for 6, 7-dimethoxy-4-pyrrolidylquinazolines as phosphodiesterase 10A inhibitors using the linear interaction energy method. *Journal of Molecular Graphics and Modelling*, 61, 44-52.
6. Halder, A. K., Amin, S. A., Jha, T., & Gayen, S. (2017). Insight into the structural requirements of pyrimidine-based phosphodiesterase 10A (PDE10A) inhibitors by multiple validated 3D QSAR approaches. *SAR and QSAR in Environmental Research*, 28(3), 253-273.
7. Goodarzi, M., Saeys, W., Deeb, O., Pieters, S., & Vander Heyden, Y. (2013). Particle swarm optimization and genetic algorithm as feature selection techniques for the QSAR modeling of imidazo [1, 5-a] pyrido [3, 2-e] pyrazines, inhibitors of phosphodiesterase 10 A. *Chemical biology & drug design*, 82(6), 685-696.
8. Dym, O., Xenarios, I., Ke, H., & Colicelli, J. (2002). Molecular docking of competitive phosphodiesterase inhibitors. *Molecular pharmacology*, 61(1), 20-25.
9. Fossa, P., Giordanetto, F., Menozzi, G., & Mosti, L. (2002). Structural basis for selective PDE 3 inhibition: a docking study. *Quantitative Structure-Activity Relationships*, 21(3), 267-275.
10. Mpamhanga, C. P., Chen, B., McLay, I. M., Ormsby, D. L., & Lindvall, M. K. (2005). Retrospective docking study of PDE4B ligands and an analysis of the behavior of selected scoring functions. *Journal of chemical information and modeling*, 45(4), 1061-1074.
11. Oliveira, F. G., Sant'Anna, C. M., Caffarena, E. R., Dardenne, L. E., & Barreiro, E. J. (2006). Molecular docking study and development of an empirical binding free energy model for phosphodiesterase 4 inhibitors. *Bioorganic & medicinal chemistry*, 14(17), 6001-6011.
12. Kang, N. S., Jhon, D. J., Song, J. H., & Yoo, S. E. (2007). Docking and 3-D QSAR studies of dual PDE4-PDE7 inhibitors. *Molecular Simulation*, 33(14), 1109-1117.
13. Martins, D. R., Pazini, F., de Medeiros Alves, V., de Moura, S. S., Lião, L. M., de Magalhães, M. T. Q., ... & Rocha, M. L. (2013). Synthesis, docking studies, pharmacological activity and toxicity of a novel Pyrazole derivative (LQFM 021)—Possible effects on phosphodiesterase. *Chemical and Pharmaceutical Bulletin*, 61(5), 524-531.
14. Cichero, E., D'Ursi, P., Moscatelli, M., Bruno, O., Orro, A., Rotolo, C., ... & Fossa, P. (2013). Homology modeling, docking studies and molecular dynamic simulations using graphical processing unit architecture to probe the type-11 phosphodiesterase catalytic site: A computational approach for the rational design of selective inhibitors. *Chemical biology & drug design*, 82(6), 718-731.
15. Du, Q. S., Huang, R. B., & Chou, K. C. (2008). Recent advances in QSAR and their applications in predicting the activities of chemical molecules, peptides and proteins for drug design. *Current protein and peptide science*, 9(3), 248-259.
16. Roy, K., & Ojha, P. K. (2010). Advances in quantitative structure–activity relationship models of antimalarials. *Expert opinion on drug discovery*, 5(8), 751-778.
17. Release, HyperChem. "7.5 for windows, molecular modeling system, Hypercube." Inc. <http://www.hyper.com> (2002).
18. Yap, C. W. (2011). PaDEL-descriptor: An open source software to calculate molecular descriptors and fingerprints. *Journal of computational chemistry*, 32(7), 1466-1474.
19. Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., & Olson, A. J. (2009). AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of computational chemistry*, 30(16), 2785-2791.
20. Stewart, J. J. (1990). MOPAC: a semiempirical molecular orbital program. *Journal of computer-aided molecular design*, 4(1), 1-103.

21. Bos, M., & Weber, H. T. (1991). Comparison of the training of neural networks for quantitative X-ray fluorescence spectrometry by a genetic algorithm and backward error propagation. *Analytica Chimica Acta*, 247(1), 97-105.
22. Lucasius, C. B., & Kateman, G. (1991). Genetic algorithms for large-scale optimization in chemometrics: an application. *TrAC Trends in Analytical Chemistry*, 10(8), 254-261.
23. Goudarzi, N., Goodarzi, M., Chamjangali, M. A., & Fatemi, M. H. (2013). Application of a new SPA-SVM coupling method for QSPR study of electrophoretic mobilities of some organic and inorganic compounds. *Chinese Chemical Letters*, 24(10), 904-908.
24. Cristianini, N., & Shawe-Taylor, J. (2000). *An introduction to support vector machines and other kernel-based learning methods*. Cambridge university press.
25. Gunn, S. R. (1998). Support vector machines for classification and regression. *ISIS technical report*, 14(1), 5-16.
26. Lu, H., Goren, A. C., & Zhan, C. G. (2010). Characterization of the structures of phosphodiesterase 10 binding with adenosine 3', 5'-monophosphate and guanosine 3', 5'-monophosphate by hybrid quantum mechanical/molecular mechanical calculations. *The Journal of Physical Chemistry B*, 114(20), 7022-7028.
27. Morris, G. M., Goodsell, D. S., Halliday, R. S., Huey, R., Hart, W. E., Belew, R. K., & Olson, A. J. (1998). Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *Journal of computational chemistry*, 19(14), 1639-1662.
28. Sheridan, R. P., Feuston, B. P., Maiorov, V. N., & Kearsley, S. K. (2004). Similarity to molecules in the training set is a good discriminator for prediction accuracy in QSAR. *Journal of chemical information and computer sciences*, 44(6), 1912-1928.
29. Eriksson, L., Jaworska, J., Worth, A. P., Cronin, M. T., McDowell, R. M., & Gramatica, P. (2003). Methods for reliability and uncertainty assessment and for applicability evaluations of classification-and regression-based QSARs. *Environmental health perspectives*, 111(10), 1361-1375.
30. Kollman, P. A., Massova, I., Reyes, C., Kuhn, B., Huo, S., Chong, L., ... & Cheatham, T. E. (2000). Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Accounts of chemical research*, 33(12), 889-897.
31. Arba, M., Ihsan, S., Wahyudi, S. T., & Tjahjono, D. H. (2017). Molecular modeling of cationic porphyrin-anthraquinone hybrids as DNA topoisomerase II β inhibitors. *Computational biology and chemistry*, 71, 129-135.
32. Balasubramanian, P. K., Balupuri, A., Kang, H. Y., & Cho, S. J. (2017). Receptor-guided 3D-QSAR studies, molecular dynamics simulation and free energy calculations of Btk kinase inhibitors. *BMC systems biology*, 11(2), 1-11.
33. Maleki, M., Vasudev, G., & Rueda, L. (2013). The role of electrostatic energy in prediction of obligate protein-protein interactions. *Proteome science*, 11(1), 1-12.