Original Article

Application of MIA-QSAR Method for Study of Benzamidin Derivatives and Design of New Compounds

Hooshang Nasiri[1], Ali Niazi[2]*, Atisa Yazdanipour [2], Tahereh Momeni-Isfahania[1]

[1] Department of Chemistry, Arak Branch, Islamic Azad University, Arak, Iran. [2]Department of Chemistry, Central Tehran Branch, Islamic Azad University, Tehran, Iran.

Abstract

Malaria is known as an infectious disease that is transmitted from the parasite anopheles mosquitoes from the plasmodium species. Nowadays, this disease is a serious problem around the world, which imposes many costs for treatment to countries. Therefore, the design of new drugs for this disease is one of the serious concerns of chemical researchers and pharmacists. In this paper, quantitative structure-activity relationship (QSAR) analysis has been carried out for the prediction of inhibitory activity of a set of Benzamidin compounds. In this paper, we studied the effect of variable selection by application of genetic algorithms (GAs) for the PLS model. The GAs is very useful in the variable selection in modeling and selecting the subset of pixels with the low prediction error. We applied to orthogonal pixel correction (OPC) as pre-processing methods. These models were applied for the prediction of the molecules inhibition, which were not in the modeling procedure. The resulting model showed a high predictive ability with the root mean square error of prediction (RMSEP) of 0.3464 for PLS and 0.1986 for GA-PLS and 0.7550 for OPC-GA-PLS models, respectively. Finally, the proposed QSAR model with the OPC-GA-PLS method was used to predict the inhibitory activity of the new compounds.

Keywords: Multivariate image processing, Orthogonal pixel correction, Genetic Algorithms, Partial least squares, Benzamidin Compounds

Introduction

Malaria is known as an infectious disease in the world [1,2]. Nowadays, this disease imposes many costs for treatment to countries [3,4]. Therefore, the design of new drugs for this disease is one of the serious concerns of chemical researchers and pharmacists [5,6]. To achieve this goal, quantitative structure-activity relationship (QSAR) is one of the most useful methods for designing and developing new antimalarial drugs [7,8].

QSAR-method reports a mathematical relationship between the chemical structure of the compounds and their physical, chemical, or biological properties. Then, surveying the reaction between ligand and receptor, it designs a novel molecule. Biological effects of compounds with similar physico-chemical properties are similar. For drug design, QSAR method makes a correlation between structural properties of potential drug candidates and their potency in inhibiting a specific biological function. The concept of a quantitative structure-activity relationship (QSAR) was first introduced in 1964. In fact, QSAR is trying to find a balanced and coordinated relationship between molecular structures and chemical activity, then, based on these relationships, it evaluates the activity of new compounds. Most QSAR applications are in areas such as drug design and estimation of environmental damage of chemical compounds [9]. In this research, we have studied benzamidine derivatives as an antimalarial drug [10].

Several QSAR studies have been done successfully on antimalarial drugs [11,12]. Using the topologic descriptor of the Szeged (SZ) index, Agrawl et al. reported a QSAR model for a series of sulfonamide derivatives of anti-malarial properties. Helping QSAR model, Jay Ann et al. studied anti-malarial properties of anti-artemisin derivatives [13]. Based on multiple linear regression model, Gupta et al surveyed quinoline antimalarial agents [14]. Montgado and colleagues carried out QSAR studies on benzamidine derivatives as an antimalarial drug. In this work, the QSAR predictive model

Address for correspondence: Mr. Ali Niazi, Department of Chemistry, Central Tehran Branch, Islamic Azad University, Tehran, Iran.
E-mail: Ali.niazi@gmail.com

How to cite this article: Nasiri, H., Niazi, A., Yazdanipour, A., Momeni-Isfahania, T. Application of MIA-QSAR Method for Study of Benzamidin Derivatives and Design of New Compounds. Arch Pharma Pract 2020;11(S1):161-8.

was developed based on the GETAWAY descriptors. R2 and Q2 in this method was 0.78 and 0. 69, respectively. The results showed that this descriptor is suitable for predicting of the anti-malarial power of base benzamidine derivatives [15] In 2010, Saho and colleagues conducted QSAR studies on 4-amino-7-chloroquine derivatives for anti-malarial activity [16]. In 2011, Ojha.et al. conducted a new QSAR study on endocrin (derived from quinoline) comparing several QSAR methods consisting of FA- MLR, FA-PLS and GFA-PLS [17]. In 2012, Ibizim et al. performed their QSAR studies on Ariel Piperazine derivatives as antimalaria and reported R2= 0.998 for QSAR model [18]. QSAR is one of the most important applications of Chemometrics that gives useful information for the design of new compounds acting on a specific target and with desired properties. Nowadays, the QSAR technique was applied in pharmaceutical chemistry, toxicology, drug design, geology, and remote sensing extensively [19-22].

In comparison with different QSAR methods, multivariate image analysis (MIA-QSAR) offers a rapid analysis and reliable results similar to the most sophisticated methodologies available today. Besides, it is cheap and easy to deal with and it is capable of predicting any modeled response for a congeneric series of chemical structures with no 3D alignment and conformational analysis. MIA-QSAR method was established by Esbensen and Geladi [23] and other researchers used it afterwards [22,24,25]. In this technique, 2D pixels show the topo-chemical features of the compounds; it is built a model between molecules and y-block consisted of independent variables. MIA-QSAR is a non-invasive analysis deals with a lot of information eliminating inessential time and expensive measures. MIA-QSAR technique aims at correlating several columns of independent variables to one column dependent variable. In MIA-QSAR approach, many coordinate of pixels in the molecular drawing display the structural changes, which were used to show the variance in bioactivity for a congeneric group in drug-like compounds. In modeling step, substitution pattern along with the congeneric series of compounds use and prediction of the bioactivities of similar compounds can be feasible.

MIA-QSAR method involves the following steps, in the respective order:

- Drawing the molecule structures, image producing and their alignment
- De-noising and then unfolding images to a two-way array and descriptors generation
- Feature selection and regression modeling

One of the major steps here is modeling. It is possible to use different methods for modelling, such as multiple linear regression (MLR) [26,27], partial least squares (PLS) [28] and artificial neural network (ANN) [22].

QSAR investigations have used MLR, extensively; however, it led to relatively poor accuracy. Additionally, MLR is effective when the amount of rows surpasses the number of columns. ANN shows enough accuracy in most cases; however, there is a probability of overfitting the training data that might not be capable of extrapolating suitable information of data consequently. The partial least squares (PLS) regression analysis is the most commonly used method for this purpose, which is based on the factor analysis that in a chemical manner applied by Joreskog and Wold [29]. The PLS transformed X and Y matrices to a sum of latent variable simultaneously [30]. Several groups have applied PLS [27]. The success rate of modeling is dependent on the correct selection of molecular descriptors. Therefore, the method of variable selection is very effective; it should show the most information on activity variations and less collinearity among them. Genetic Algorithms (GAs) is capable of addressing this need [27]. GAs is a stochastic technique explained by Leardi and his associate.26 This algorithm has simulated theory of evolution that believes the best genome is more likely to survive and to be transferred by reproduction. GAs uses different fitness criterion and genetic functions [31,32]. It was proved to preprocessing before PLS regression removes unwanted information and provides suitable input for PLS, therefore enhancing model quality. In 1998, orthogonal signal correction (OSC) was introduced by Wold at el as preprocessing [33]. Previous studies have dealt with OSC [34,35]. The key idea of OSC (in this study orthogonal pixel correction) is deleting extraneous and systematic variance from the matrix X that are unrelated and orthogonal to the matrix Y indeed structured noise in X35. This study aims at generating a MIA-QSAR model for benzamidine and then predicting their EXP IC50 with OPC -GA-PLS model and finally, designing the new compounds based on this model. The half-maximal inhibitory concentration (IC50) is one of the parameters for clarifying a measure of the potency of a substance to inhibit a specific biological or biochemical function revealing the minimum molar concentration leading to 50% of the enzyme inhibition.

Materials and Methods

The EXPIC50 data of benzamidines for inhibitory were received by review of literature [33]. The structure of each compound and its matching EXP IC50 are shown Table 1.

The data set is divided into the parts of the training and prediction set based on the Kennard-Stones algorithm [34,35]. This algorithm, as one of the best techniques to build training and prediction sets in QSAR way, involves the total space occupied by the original data set.

Hardware and Software

The Asus personal (8 GB RAM) equipped with the Windows 10 operating system and MATLAB (Version 10.0) was applied for calculations. ChemOffice package (Version 2010) was used for drawing the molecular structures. Kennard–Stones program was written in MATLAB based on the previous algorithm[34,35].

Preparation of Chemical Images and Data Matrix

In the MIA-QSAR modeling, there is an association between independent variables or descriptors that are the pixels of the 2D or 3D images and the dependent variables or the specific property. First, the ChemOffice was used to draw the structures of each compound in Table 1, were drawn in same size; next, they were converted into bitmaps in 150 × 83 pixels windows, then built congruent the common structural scaffold of all structures so that, they were fixed at the 36 × 44 coordinate that is visible in Fig. 1. This pixel was selected as a reference in the alignment step. Each 2D image was read and transformed to double array in MATLAB. In this matrix, 0 and 756, in the respective order, match the black and white pixels based on RGB color composition. Alignment of the 29 images gave a three-way array of 29×150×83 dimension. In unfolding step, this array was turned into a two-way array (matrix) of 29×37350 dimension. The columns that are precisely similar to each other were removed.

Results and Discussion

Principal Component Analysis is a variable reduction procedure. Principal Components (PCs) are capable of detecting internal relations between features of a set of objects, therefore allowing a drastic reduction of the dimensionality of the original raw data. This decrease is realised through the transformation of the original matrix to a new one, whose set of variablest ermed PCs seem to be orthogonal to each other (uncorrelated) and ordered so that the first few, with descending significance, keeps most of the variance content from the total set of original variables [33]. For initial data analysis, a Chemometrics tool, namely principal component analysis (PCA) was carried out on 29 compounds. PCA aims at searching the distributions in chemical space, studying clusters and exploring outliers. The PCA results show that 73.76% of the overall variances related to three PCs as follows: PC1=54.54%, PC2=10.71% and PC3=8.51% (Fig. 2). Since the most of the variables can be accounted for the first three PCs, their score plot is a reliable.

Presentation of the spatial distribution of the points in the data set. It is not seen clear clustering between compounds in Fig. 2 which is very important in the generation of efficient QSAR models. For regression analysis, the data set was separated into two classes, a training set (24 data) and a prediction set (5 data) based on Kennard-Stones algorithm. As Fig. 2 shows, these 2 groups were chosen in balanced way from the whole of the space of the principal components. The Kennard–Stone algorithm is a method to select a set of compounds from an original data set. Thus, the first two molecules are chosen from two distant points and the third molecule the farthest from the first two molecules, etc. PLS Analysis:

PLS is a linear modelling technique in which the information in the descriptor matrix X is projected onto a small number of underlying ('latent') variables called PLS components or latent variables. The matrix Y is concurrently applied to estimate the 'latent' variables in X, which will be the most relevant for the Y variables prediction. Pixel changes in descriptors of image of molecules, their principal components or latent variables can be caused by independent or predictor variables.

In multivariate calibration.

It is possible to develop the relationship between the matrix of pixel as latent variable and activity matrix by implementing Partial least-square (PLS) modeling, which is a potent multivariate statistical tool. To this aim, the Kennard - Stone algorithm is applied for sorting data into training and prediction sets and then building the PLS model. For determining the number of optimum latent variables in modeling, cross validation (leave-one-out) method was used. This method works based on minimum in prediction error variance or RMSECV assuming that RMSECV for the model is not considerably greater than the minimum RMSECV. Thus, only one compound at a time was deleted and the remaining of the training set was applied in modeling; ultimately, the activity of the eliminated compound was predicted using this model. In the next steps, this was repeated for the other compounds. RMSEP values match the optimum number of factors shown in Table 2.

GA-PLS modeling

The appropriate and effective descriptor makes the model more precise; therefore, variablesel ection results in robust model. In this study, the GA was run for descriptor selection. One of the key features of GA is that it studies numerous possible solutions concurrently that eac.

one explores different regions in space of input variable [22]. The GA produces many random sets of variables defined by a chromosome. The number of genes or variables in each chromosome equals the number of descriptors. Each subset of chromosome is checked by its fitness for predicting the inhibitory activity values. After running the GAs, pixels were chosen according to the best prediction of EXP IC50. RMSEP values for this model are shown in Table 2. According to our findings, it seems that the GAs can be a good method for pixel selection in image analysis.

OSC-GA-PLS Modeling

OPC algorithms were applied for eliminating systematic variation from the response pixel that is unrelated or orthogonal, to the property matrix Y and extraction of important information regarding the data. After preprocessing, GA was used as descriptor selection and optimized variation and selection of the fitness values, and then calibration model provided by PLS regression. Table 2 displays the optimum number of RMSEP values. For estimating the models used in this study (PLS, GA-PLS and OPC GA-PLS), the inhibitory activities were tested with a set of tests. Table 2 shows the results. As shown, the OPC treated data give considerably lower RMSEP values than two other models.

Model Validation and Prediction of Inhibitory Activity

To assess the predictive ability of different models (PLS, GA-PLS, and OPC-GA-PLS), it is possible to use the root mean square error of prediction (RMSEP) and relative standard error of prediction (RSEP) and cross validation coefficient (Q2 and R2):

$$RMSEP = \sqrt{\frac{\sum_{i=1}^{n}(y_{pre} - y_{exp})^2}{n}} \quad (1)$$

$$RSEP(\%) = 100 \times \sqrt{\frac{\sum_{i=1}^{n}(y_{pre} - y_{exp})^2}{\sum(y_{exp})^2}} \quad (2)$$

$$Q^2_{abs} = 1 - \frac{\sum_Y \left(Y_{exp} - Y_{LOO}\right)^2}{\sum_Y \left(Y_{exp} - \bar{Y}_{exp}\right)^2} \quad (3)$$

$$R^2_{abs} = 1 - \frac{\sum_Y \left(Y_{exp} - Y_{pred}\right)^2}{\sum_Y \left(Y_{exp} - \bar{Y}_{exp}\right)^2} \quad (4)$$

For this object, prediction of inhibitory activity of 5 molecules (Table 1 shows their structures) was investigated, and the statistical parameters were estimated (see Table 2) revealing the good statistical qualities.

Molecular Design

One of the drug design methods is computational approaches. As an application of the proposed method, OPC-GA-PLS model was used in this paper for predicting the inhibitory activity of five new compounds for which the biological tests were not performed yet. EXPIC50= 5 shows potent inhibitory. According to this EXP IC50 value, new compounds were designed. The inhibitory activity of these five compounds, which were manually designed, was computed by this proposed method as shown in Table 3.

Conclusion

A fitting and robust MIA-QSAR/OPC-GA-PLS model was developed for benzamidines compounds offering a simple 2D image-based approach for measurement of inhibitory. The results of the table showed the power of pixels in the prediction of inhibitory activity of benzamidines. The QSAR model developed in this study is capable of offering an effective tool for predicting the activity of new compounds and designing new compounds with increased activity.

Acknowledgements

The authors gratefully acknowledge the support of this work by the Islamic Azad University, Arak Branch.

References

1. Snow RW, Guerra CA, Noor AM, Myint HY, Hay SI. The global distribution of clinical episodes of Plasmodium falciparum malaria. Nature. 2005;434(7030):214-7.
2. Bell DR, Jorgensen P, Christophel EM, Palmer KL. Malaria risk: estimation of the malaria burden. Nature. 2005;437(7056):E3-4.
3. LaCount DJ, Vignali M, Chettier R, et al. A protein interaction network of the malaria parasite Plasmodium falciparum. Nature. 2005;438(7064):103-7.
4. Brady RL, Cameron A. Structure-based approaches to the development of novel anti-malarials. Current drug targets. 2004 Feb 1;5(2):137-49.
5. Brady RL, Cameron A. Structure-based approaches to the development of novel anti-malarials. Curr Drug Targets. 2004;5(2):137-49.
6. Wiesner J, Ortmann R, Jomaa H, Schlitzer M. New antimalarial drugs. Angew Chem Int Ed Engl. 2003;42(43):5274-93.
7. Gonzalez-Diaz H, Cruz-Monteagudo M, Vina D, Santana L, Uriarte E, De Clercq E. QSAR for anti-RNA-virus activity, synthesis, and assay of anti-RSV carbonucleosides given a unified representation of spectral moments, quadratic, and topologic indices. Bioorg Med Chem Lett. 2005;15(6):1651-7.
8. Santana L, Uriarte E, Gonzalez-Diaz H, Zagotto G, Soto-Otero R, Mendez-Alvarez E. A QSAR model for in silico screening of MAO-A inhibitors. Prediction, synthesis, and biological assay of novel coumarins. J Med Chem. 2006;49(3):1149-56.
9. Kubinyi H, Wolff M, Burgeros E. Medicinal Chemistry and Drug Discovery, Principles and Practice. John Wiley & Sons, New York; 1995.
10. Leban J, Pegoraro S, Dormeyer M, Lanzer M, Aschenbrenner A, Kramer B. Sulfonyl-phenyl-ureido benzamidines; a novel structural class of potent antimalarial agents. Bioorg Med Chem Lett. 2004;14(8):1979-82.
11. Xie A, Sivaprakasam P, Doerksen RJ. 3D-QSAR analysis of antimalarial farnesyltransferase inhibitors based on a 2,5-diaminobenzophenone scaffold. Bioorg Med Chem. 2006;14(21):7311-23.
12. Geladi P, Esbensen K. Can image analysis provide information useful in chemistry?. Journal of Chemometrics. 1989 Mar;3(2):419-29.
13. Chen M, Theander TG, Christensen BS, et al. Licochalcone A, a new antimalarial agent, inhibits in vitro growth of the human malaria parasite Plasmodium falciparum and protects mice from P. yoelii infection. Antimicrob Agents Chemother. 1994;38(7):1470-5.
14. Xie A, Sivaprakasam P, doerksen RJ. 3D-QSAR analysis of antimalarial farnesyltransferase inhibitors based on a 2,5-diaminobenzophenone scaffold. Bioorg Med Chem. 2006;14(21):7311-23.
15. Cruz-Monteagudo M, Borges F, Perez González M, Cordeiro MN. Computational modeling tools for the design of potent antimalarial bisbenzamidines: overcoming the antimalarial potential of pentamidine. Bioorg Med Chem. 2007;15(15):5322-39.
16. Sahu NK, Sharma MC, Mourya V, Kohli DV. QSAR studies of some side chain modified 7-chloro-4-aminoquinolines as antimalarial agents. Arabian Journal of Chemistry. 2014 Nov 1;7(5):701-7.
17. Ojha PK, Roy K. Chemometric modeling, docking and in silico design of triazolopyrimidine-based dihydroorotate dehydrogenase inhibitors as antimalarials. European journal of medicinal chemistry. 2010 Oct 1;45(10):4645-56.
18. Ibezim E, Duchowicz PR, Ortiz EV, Castro EA. QSAR on aryl-piperazine derivatives with activity on malaria. Chemometrics and Intelligent Laboratory Systems. 2012 Jan 15;110(1):81-8.
19. Goodarzi M, Freitas MP. On the use of PLS and N-PLS in MIA-QSAR: azole antifungals. Chemometrics and Intelligent Laboratory Systems. 2009 Mar 15;96(1):59-62.
20. Sarkhosh M, Ghasemi JB, Ayati M. A quantitative structure-property relationship of gas chromatographic/mass spectrometric retention data of 85 volatile organic compounds as air pollutant materials by multivariate methods. Chemistry Central Journal. 2002; 6:1–8
21. Cui W, Yan X. Adaptive weighted least square support vector machine regression integrated with outlier detection and its application in QSAR. Chemometrics and Intelligent Laboratory Systems. 2009 Oct 15;98(2):130-5.
22. Sarkhosh M, Khorshidi N, Niazi A, Leardi R. Application of genetic algorithms for pixel selection in multivariate image analysis for a QSAR study of trypanocidal activity for quinone compounds and design new quinone compounds. Chemometrics and Intelligent Laboratory Systems. 2014 Dec 15;139:168-74.
23. Niazi A, Ghasemi N. PARAFAC and PLS applied to determination of captopril in pharmaceutical preparation and biological fluids by ultraviolet spectrophotometry. Annali di Chimica: Journal of Analytical, Environmental and Cultural Heritage Chemistry. 2007 Aug;97(9):845-58.
24. Akrami A, Niazi A. Application of MIA for a QSAR Study of Inhibitory Activity of DHP Derivatives and Design of New

Compounds Using WT and GA for Pixel Processing. Polycyclic Aromatic Compounds. 2017 Oct 20;37(5):442-55.
25. Veyseh S, Hamzehali H, Niazi A, Ghasemi JB. Application of multivariate image analysis in QSPR study of pKa of various acids by principal components-least squares support vector machine. Journal of the Chilean Chemical Society. 2015 Sep;60(3):2985-7.
26. Narasimhan B, Judge V, Narang R, Ohlan R, Ohlan S. Quantitative structure-activity relationship studies for prediction of antimicrobial activity of synthesized 2,4-hexadienoic acid derivatives. Bioorg Med Chem Lett. 2007;17(21):5836-45.
27. Zareh MK. A QSPR study of boiling point of saturated alcohols using genetic algorithm. Acta Chim. Slov. 2003;50:259-73.
28. Niazi A, Bozorghi SJ, Shargh DN. Prediction of acidity constants of thiazolidine-4-carboxylic acid derivatives using ab initio and genetic algorithm-partial least squares. Turkish Journal of Chemistry. 2006 Dec 15;30(5):619-28.
29. Jöreskog KG, Wold H. North-Holland, Amsterdam, 1982;68:108.
30. Niazi A, Ghasemi N. Parafac and PLS applied to determination of captopril in pharmaceutical preparation and biological fluids by ultraviolet spectrophotometry. Ann Chim. 2007;97(9):845-58.
31. Niazi A, Leardi R. Genetic algorithms in chemometrics. Journal of Chemometrics. 2012 Jun;26(6):345-51.
32. Leardi R, in: Genetic algorithms in molecular modeling, Elsevier, 1996; 67-86.
33. Wold S, Antti H, Lindgren F, Öhman J. Orthogonal signal correction of near-infrared spectra. Chemometrics and Intelligent laboratory systems. 1998 Dec 14;44(1-2):175-85.
34. Pauli ED, Bruns RE, Scarminio IS. UV-Vis spectral fingerprinting and chemometric method applied to the evaluation of Camellia sinensis leaves from different harvests. Analytical Methods. 2016;8(41):7537-44.
35. Niazi A, Azizi A. Orthogonal Signal Correction--Partial Least Squares Method for Simultaneous Spectrophotometric Determination of Nickel, Cobalt, and Zinc. Turkish Journal of Chemistry. 2008 Mar 25;32(2):217-28.

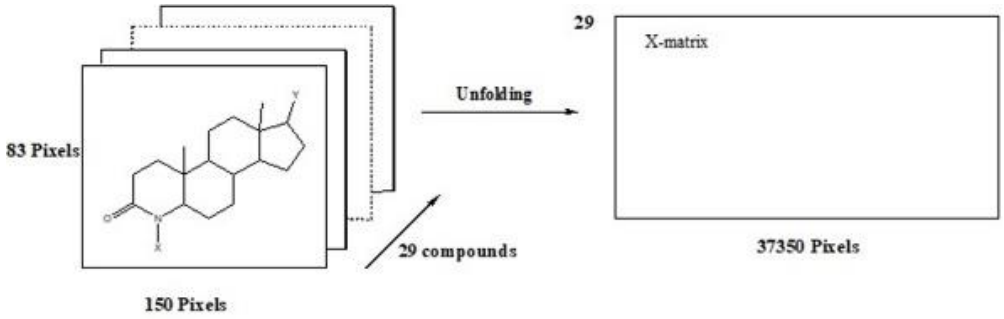
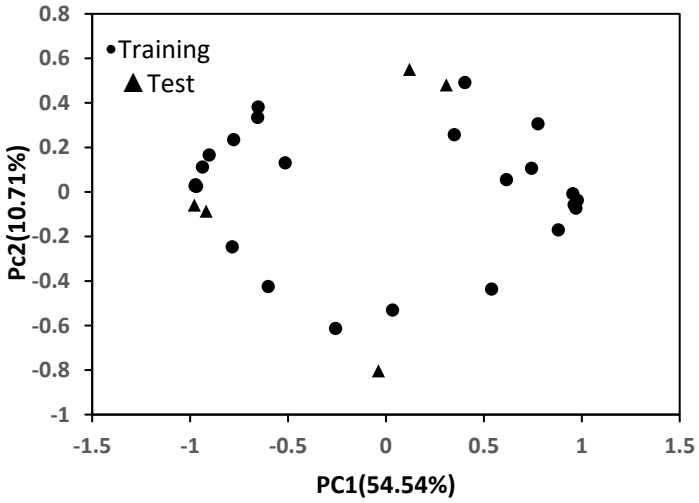
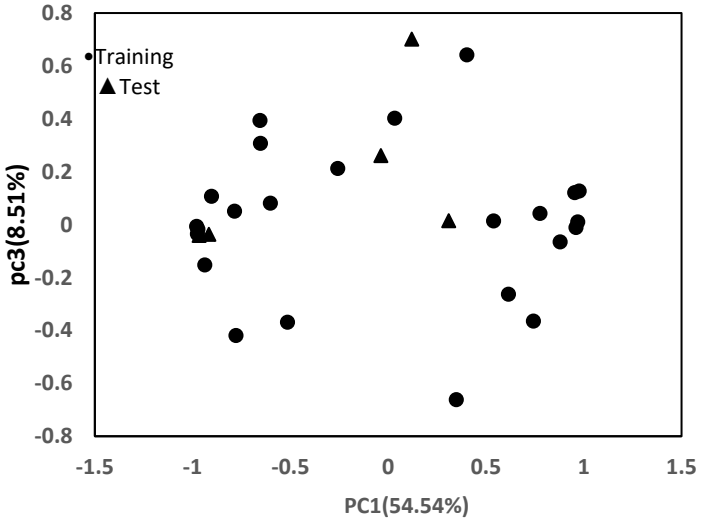


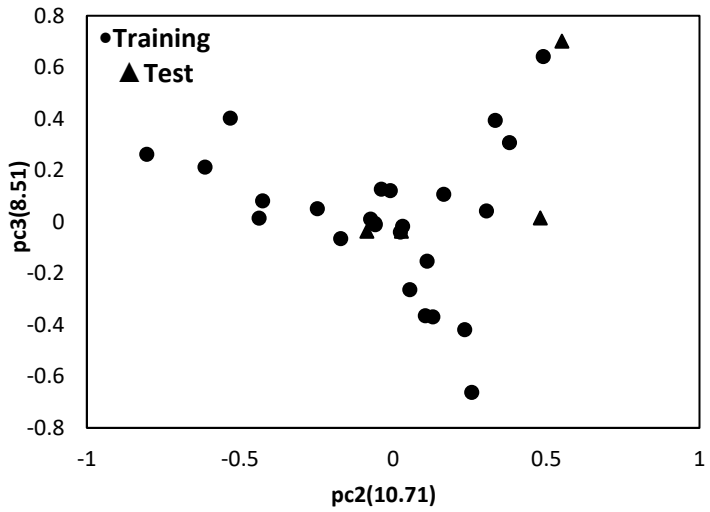
Figure 1: 2D images and unfolding step of the 29 chemical structures to give the X-matrix. The arrow in structure indicates the coordinate of a pixel in common among the whole series of compounds, used in the 2D alignment step.



(A)



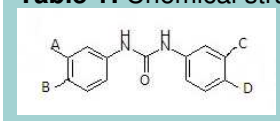
(B)



(C)

Figure 2: Principal components analysis of the 2D image descriptors for the data set; (A) PC1 versus PC2, (B) PC1 versus PC3 and (C) PC2 versus PC3.

Table 1: Chemical structure and EXPIC50 Data of benzamidines



Compound	A	B	C	D	EXPIc50
*1	CN_2H_3	H	H	$SO_2NHC_6H_5$	5.045
2	CN_2H_3	H	H	$SO_2NHC_{10}H_7$	5.086
3	CN_2H_3	H	$SO_2NHCH_2C_6H_5$	H	5.207
*4	H	CN_2H_3	$HSO_2NHCH_2C_6H_5$	H	6.397
5	CN_2H_3	H	H	$SO_2NHC_2H_4C_6H_5$	5.602
6	H	CN_2H_3	$SO_2NHCH_2C_6H_4SO_2NH_2$	H	5.602
7	CN_2H_3	H	H	$SO_2NHCH_2C_6H_3F_2$	5.954
8	CN_2H_3	H	$SO_2NHCH_2C_6H_4SO_2NH_2$	H	6.008
9	CN_2H_3	H	H	$SO_2NHCH_2CH_2OH$	6.050
*10	CN_2H_3	H	H	$SO_2NHCH_2C_6H_4F$	6.301
11	CN_2H_3	H	H	$SO_2NHCH_2C_6H_5$	6.346
12	H	CN_2H_3	$SO_2NHCH_2C_6H_5$	H	6.387
13	CN_2H_3	H	$SO_2NHCH_2C_6H_2F_3$	H	6.602
14	H	CN_2H_3	H	$SO_2NHCH_2C_6H_2F_3$	6.677
15	H	CN_2H_3	$SO_2NHCH_2C_6H_2F_3$	H	6.721
16	CN_2H_3	H	H	$SO_2NHCH_2C_6H_2F3$	7.769
17	CN_2H_3	H	H	$SO_2NHCH_2C_6H_4SO_2NH_2$	8.154
18	CN_2H_3	H	CN_2H_3	H	8.301
19	CN_2H_3	H	H	$SO_2NHC_{10}H_7$	4.769
20	CN_2H_3	H	H	SO_2NH_2	5.309
*21	CN_2H_3	H	H	$SO_2NHC_4H_7$	5.408
22	CN_2H_3	H	H	$SO_2NHCH_2C_6H_2F_3$	5.744
23	CN_2H_3	H	H	$SO_2NHCH_2C_6H_3F_2$	5.954
24	CN_2H_3	H	H	$SO_2NHCH_2C_6H_4F$	6.207
25	CN_2H_3	H	H	$SO_2C_{10}H_7$	6.214
*26	CN_2H_3	H	H	$SO_2NHCH_2C_6H_5$	6.397
27	CN_2H_3	H	H	$SO_2NHCH_2C_6H_4F$	6.494
28	H	CN_2H_3	H	$SO_2NHCH_2C_6H_4SO_2NH_2$	7.045
29	CN_2H_3	H	H	$SO_2NC_9H_8$	7.769

*Compounds selected for test set.

Table 2: Observation and calculation values of (EX9IC50) using PLS, GA-PLS and OPC-GA-PLS models.

Number of compounds (Table 1)	Observation activity	PLS		GA-PLS		OPC-GA-PLS	
		Predicted	Error (%)	Predicted	Error (%)	Predicted	Error (%)
1	5.045	5.1183	1.45	5.1624	2.33	5.029	-0.32
4	6.397	6.683	4.47	6.4399	0.67	6.353	-0.68
10	6.301	6.192	-1.73	6.432	2.08	6.274	-0.43
21	5.408	5.4103	0.04	5.1727	-4.35	5.4562	0.89
26	6.397	5.6293	-3.63	6.0668	-5.16	6.2445	-2.38
LVs		12		9		4	
Q^2		0.8231		0.9029		0.9884	
R^2		0.631		0.8788		0.9825	
RMSEP		0.3464		0.1986		0.755	
RSEP (%)		2.66		1.517		0.5751	

Table 3: Chemical structures with the observed values of the inhibitory activity for new compounds.

No.	A	B	C	D	Observed
1	CN_2H_3	H	H	SO_2NHCH_3	5.45
2	H	CN_2H_3	H	SO_2NHCH_3	6.02
3	CN_2H_3	H	SO_2NHCH_3	H	6.178
4	CN_2H_3	H	SO_2NH_2	H	6.07
5	H	CN_2H_3	$SO_2NHCH_2C_6H_4SO_2NHCH_3$	H	5.65