Application of molecular topology to the prediction of the phenoloxidase inhibition by a group of benzaldehyde thiosemicarbazone and their derivatives

Ramón García-Domenech*, María L. Calvo-Chamorro, Angélica Y. Cuervo-Arias, Leysa J. Gómez-Sucerquia, Verónica Ortega-Chávez, Eva Pérez-Torrado, Jorge Gálvez.

Dept. Química Física, Facultad de Farmacia, Universitat de Valencia, Avd. V.A. Estellés, s/n, 46100-Burjassot, Valencia, Spain.

Aplicación de la topología molecular para la predicción de la inhibición de la feniloxidasa de un grupo de benzaldehydo tiosemicarbazonas y derivados

Aplicació de la topologia molecular per a la predicció de la inhibició de la feniloxidasa d'un grup de benzaldehydo tiosemicarbazonas i derivats

Recibido: 23 de abril de 2008; revisado: 25 de julio de 2008; aceptado: 29 de julio de 2008

RESUMEN

Se ha desarrollado un modelo topológico-matemático para investigar nuevos derivados de benzaldehydo tiosemicarbazonas y compuestos relacionados que inhiban la feniloxidasa. Utilizando el análisis de regresión multilineal se seleccionó una función con dos descriptores, $^{1}\chi^{v}$, $^{4}\chi_{p}^{v}$ y r^2 = 0.940 capaz de predecir satisfactoriamente la concentración inhibitoria cincuenta, CI_{50}, para cada compuesto. Basado en el modelo seleccionado, se realizó un barrido molecular virtual y se propusieron nuevas estructuras potencialmente activas frente al enzima feniloxidasa.

Palabras clave: **Topología Molecular. Inhibidores de la feniloxidasa. Análisis de regresión multilineal.**

SUMMARY

A topological-mathematical model has been arranged to search for new derivatives of benzaldehyde thiosemicarbazone and related compounds acting as phenoloxidase inhibitors. By using multilinear regression analysis a function with two descriptors, $^{1}\chi^{v}$, $^{4}\chi_{p}^{v}$ and r^2 = 0,940 was capable to predict adequately the IC_{50} for each compound. After carrying out a virtual screening based upon such a model, new structures potentially active against the enzyme are proposed.

Key words: **Molecular topology. Phenoloxidase inhibitors. Multilinear regression analysis.**

RESUM

S'ha desenrotllat un model topologico-matemàtic per a investigar nous derivats de benzaldehydo tiosemicarbazonas i compostos relacionats que inhibisquen la feniloxidasa. Utilitzant l'anàlisi de regressió multilineal es va seleccionar una funció amb dos descriptors, $^{1}\chi^{v}$, $^{4}\chi_{p}^{v}$ y r^2 = 0.940 capaç de predir satisfactòriament la concentració inhibitòria cinquanta, CI_{50}, per a cada compost. Basat en el model seleccionat, es va realitzar un agranat molecular virtual i es van proposar noves estructures potencialment actives enfront de l'enzim feniloxidasa.

Mots clau: **Topologia Molecular. Inhibidors de la feniloxidasa. Anàlisi de regressió multilineal.**

* Corresponding author. Tel. +34 963544291.
E-mail address: ramon.garcia@uv.es

1. INTRODUCTION

Phenoloxidase, PO, also known as tyrosinase, is a key enzyme in different metabolic processes of microorganisms and other animals and plants[1, 2]. In insects, PO is related to three important biochemical functions, including cuticule sclerotization, defensive encapsulation and the mechanization of alien organisms[3].

It is likely that PO inhibition may lead to suppression of the defense mechanisms of the insects or to an abnormal softening up of the insect body thereby making possible controlling the pest.[4]. Moreover, PO has been recently reported to be linked to Parkinson's and other neurodegenerative diseases[5, 6].

The prediction of biological properties of organic compounds is one of the main objectives of the quantitative structure-activity relationships (QSAR) and quantitative structure-property relationships (QSPR) methods. The success of these methods is closely depending of an appropriate characterization of the molecular structure as well as an adequate selection of the molecular descriptors to be correlated[7].

Molecular topology has widely demonstrated its ability for a straightforward and efficient characterization of molecular structure by means of the so-called topological indices, TIs[8]. When these indices are adequately selected, it is possible to obtain a precise mathematical characterization of each chemical compound, what allows the development of the QSAR and QSPR models[9, 10, 11, 12].

This way, TIs have demonstrated their usefulness in the prediction of diverse physical, chemical and biological properties for different types of compounds[13, 14], as well as in the selection and design of new active compounds[15, 7, 16, 17, 18]. In some cases, the predicted structures can be regarded as new lead drugs.

In the present study, the PO inhibitory activities of 57 compounds were considered to obtain predictive models using molecular topology combined with multilinear regression analysis. Furthermore, a molecular screening was performed to select new compounds with theoretical higher bioactivity.

2. MATERIALS AND METHODS

2.1. Analysed compounds

A group of benzaldehyde thiosemicarbazones, benzaldehyde, benzoic acid and their derivatives was selected as phenoloxidase inhibitors. The IC_{50} (mmol/L) experimental data reported in ref.[4] have been input to perform the correlation. Table 1 shows the chemical structure and the activity for each compound.

2.2. Molecular descriptors

In Chemical Graph Theory, molecular structures are normally represented as hydrogen-depleted graphs, whose vertices and edges act as atoms and covalent bonds, respectively. Chemical structural formulas can be then assimilated to undirected and finite multigraphs with labeled vertices, commonly known as molecular graphs. Graph-theoretical indices, also known as topological indices, are descriptors characterizing molecular graphs which enable to give account of their structural properties.

Desmol11 program[19] was used to calculate the set of graph-theoretical indices included in this work. The indices belong to the families of *Randić-Kier-Hall* sub-graph connectivity indices ${}^{m}\chi_{t}$[20] up to the fourth order, and their corresponding valence indices[21].

2.3. Multilinear regression analysis

The IC_{50} was linearly correlated with the aforementioned descriptors to obtain the connectivity function. The *Furnival-Wilson* algorithm[22] was used to obtain subsets of descriptors, and equations with the least Mallows Cp[23] were selected. This algorithm estimates regression equations for «best» subsets of predictor variables and performs extensive residual analysis. It identifies efficiently the subsets while computing only a small fraction of all possible regressions. Several criteria are able to define «best», for instance, the sample r^2, adjusted r^2 or Mallow's C_p.

The predictive ability of the selected equation was measured through the percentage of mean relative error, defined as:

$$MRE\ (\%) = \frac{100}{N} \sum \left| \frac{X_{Exp} - X_{Cal}}{X_{Exp}} \right|$$

The stability of the mathematical model selected was evaluated through a cross-validation by leave-one-out[24]: To do that, one compound of the set is extracted, and the model is recalculated using as training set the remaining N-1 compounds. The property is then predicted for the removed element. This process is repeated for all the compounds of the set, obtaining a prediction for every one. The correlation coefficient Q^2 of the prediction is determined. This procedure also makes easy the disclosing of outliers data.

In order to unveil the possible existence of fortuitous correlations, the randomization test is adopted in this paper[25]. To make so, the values of the property of each compound are randomly permuted and linearly correlated with the aforementioned descriptors. This process is repeated several times. The usual form to represent the results of a randomization test is plotting the correlation versus the prediction coefficients, r^2 and Q^2 respectively.

2.4. Molecular screening

Molecular topology is an efficient tool showing some advantages over molecular mechanics or quantum chemistry. The most remarkable advantage is perhaps the calculation speed. Hundreds of compounds can be analyzed within a few minutes time frame.

For this reason, molecular topology is well suited to evaluate possible biological activities of compounds represented in large databases or virtual libraries.

If the predictive power of the QSAR model obtained is satisfactory, it can be used to record and optimize the activity.

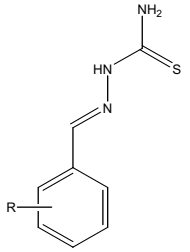
In this paper, we have designed a library of thiosemicarbazone derivates using the scaffold illustrated in Figure 1, by varying the position and number of substituents in the aromatic ring. All the compounds selected have been obtained from the Scifinder Scholar molecular database. Since the theoretical number of structures selected was too large, we just selected a cut-off for compounds with molecular weight between 150-250 g/mol being also commercially available.

3. RESULTS AND DISCUSSION

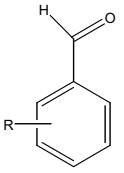
The first goal was to find the best QSAR model for describing the pIC_{50} of the group of compounds analyzed. There is a narrow interval of pIC_{50} values, ranging from 0.97 to 6.59. It makes it advisable to use a small test group, containing 20% of the sample, to validate the prediction pattern in addition to the cross-validation and randomization test of the selected function.

TABLE I

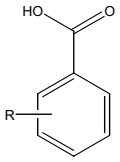
Chemical structures, activity and prediction of pIC50 for each compound using Eq. 1 selected through multilinear regression analysis.



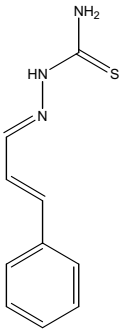
N1-N21



Q1-Q17



S1-S18



N22

Compound	R	IC_{50}exp(mM)[a]	pIC_{50}exp[a]	pIC_{50}calc[b]	Residual[b]	Residual(cv)[c]
			Training set			
N01	-	0.00083	6.08	5.74	0.34	0.37
N03	3-OH	0.00052	6.28	5.42	0.86	0.92
N05	2,4-OH	0.00105	5.98	5.70	0.28	0.30
N06	3,4-OH	0.00029	6.54	5.96	0.58	0.63
N07	2,5-OH	0.00224	5.65	5.63	0.02	0.02
N08	2-OMe	0.02138	4.67	4.97	–0.30	–0.31
N09	3-OMe	0.00191	5.72	5.70	0.02	0.02
N10	4-OMe	0.00060	6.22	6.23	–0.01	–0.01
N12	3-OH;4-OMe	0.00026	6.59	5.42	1.17	1.23
N13	4-OH;3-OMe	0.00331	5.48	5.63	–0.15	–0.16
N14	2,4-OMe	0.01549	4.81	5.44	–0.63	–0.66
N16	2,4,5-OMe	0.09999	4	4.67	–0.67	–0.76
N17	4-Propyl-i	0.00035	6.46	6.39	0.07	0.08
N18	4-Butyl	0.00028	6.55	6.86	–0.31	–0.36
N19	4-Butyl-t	0.00056	6.25	6.45	–0.20	–0.22
N21	4-$N(CH_3)_2$	0.00036	6.44	6.33	0.11	0.12
N22	-	0.00065	6.19	7.02	–0.83	–0.94
Q01	-	5.89	2.23	1.97	0.26	0.28
Q02	2-OH	4.17	2.38	1.85	0.53	0.56
Q03	3-OH	9.55	2.02	1.64	0.38	0.41
Q05	2,4-OH	4.90	2.31	1.88	0.43	0.45
Q06	3,4-OH	8.91	2.05	2.18	–0.13	–0.14
Q08	2-OMe	107.15	0.97	1.15	–0.18	–0.19
Q09	3-OMe	42.66	1.37	1.92	–0.55	–0.58
Q11	2-OH;3-OMe	5.75	2.24	1.36	0.88	0.93
Q13	3-OH;4-OMe	25.12	1.6	1.64	–0.04	–0.04
Q14	4-OH;3-OMe	50.12	1.3	1.85	–0.55	–0.58
Q15	4-Propyl-i	0.85	3.07	2.62	0.45	0.47
Q16	4-Butyl	0.74	3.13	3.09	0.04	0.04
Q17	4-Butyl-t	0.72	3.14	2.68	0.46	0.49

TABLE I (cont.)

S02	2-OH	33.88	1.47	1.61	-0.14	-0.15
S04	4-OH	16.22	1.79	2.42	-0.63	-0.66
S05	2,4-OH	30.20	1.52	1.61	-0.09	-0.10
S06	2,5-OH	32.36	1.49	1.70	-0.21	-0.22
S08	3-OH;4-NH2	25.70	1.59	2.38	-0.79	-0.83
S09	4-OH;3-NH2	3.80	2.42	2.20	0.22	0.22
S10	3-OMe	13.18	1.88	2.12	-0.24	-0.25
S11	4-OMe	5.50	2.26	2.49	-0.23	-0.24
S12	3-OH;4-OMe	47.86	1.32	1.81	-0.49	-0.51
S13	4-OH;3-OMe	69.18	1.16	2.01	-0.85	-0.89
S14	4-Propyl	1.74	2.76	1.79	0.97	1.07
S16	4-Butyl	1.05	2.98	3.12	-0.14	-0.15
S17	4-Butyl-t	2.09	2.68	2.70	-0.02	-0.03
S18	4-Hexyl	1.17	2.93	2.62	0.31	0.45
			Test set			
N02	2-OH	0.00219	5.66	5.67	-0.01	
N04	4-OH	0.00044	6.36	6.16	0.20	
N11	2-OH;3-OMe	0.00209	5.68	5.19	0.49	
N15	3,4-OMe	0.00269	5.57	5.22	0.35	
N20	3,5-Butyl-i;2-OH	0.07244	4.14	5.40	-1.26	
Q04	4-OH	3.09	2.51	2.39	0.12	
Q07	2,5-OH	1.05	2.98	1.81	1.17	
Q10	4-OMe	1.91	2.72	2.46	0.26	
Q12	2-OH;4-OMe	2.00	2.7	2.14	0.56	
S01	-	14.12	1.85	2.03	-0.18	
S03	3-OH	16.22	1.79	1.84	-0.05	
S07	3,4-OH	29.51	1.53	2.35	-0.82	
S15	4-Propyl-i	1.91	2.72	2.65	0.07	

[a] From ref.[4] [b] From the mathematical model, Eq. 1. [c] From leave-one-out cross-validation analysis.

The best linear regression equation obtained, including its statistical parameters, was:

$$pIC_{50} = -3.132(\pm 0.328) + 5.716(\pm 0.257)\ {}^{1}\chi^{v} - 16.581(\pm 0.971)\ {}^{4}\chi_{p}^{v} \quad \text{Eq. 1}$$

$N=44\ \ r^2=0.940\ \ Q^2=0.931\ \ MRE=15.3\%\ \ SEE=0.500\ \ F=321.1\ \ p<0.00001$

The value of $r^2 = 0.940$ as well as the low SEE=0.500 (less than 9% of the average value of the property) account for the quality of the selected model.

The presence of the ${}^{1}\chi^{v}$ and ${}^{4}\chi_{p}^{v}$ indices in the equation reflects the influence of branching and position of the substituent in aromatic ring, respectively. Table 1 and Fig 1 show the predicted results obtained with each one of the compounds in the training set. The mean error in predictions was: MRE = 15.3%, which reveals the good quality of the selected model .

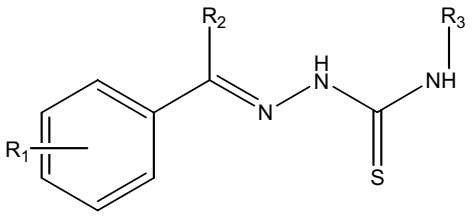
Similar results were obtained in the cross-validation study (column 7, Table 1) with a mean error of 16.3%. The graphic representations of the residuals obtained by using the selected equation versus the residuals of the cross-validation, Fig. 2a, and the correlation coefficient, r^2, vs. prediction coefficient, Q^2, Fig. 2b, illustrate the quality of the selected model. The discrepancies between the two residuals are small for most of the studied compounds. The results of the randomness tests, Fig. 2b, suggest a high stability of the model (all regressions were rather poor except for the selected equation).

Table 1 and Fig 1 show the results of the prediction obtained for the test group. The mean error is 14.2%, which is similar to the achieved for the training group.

The results obtained exclusively using graph-theoretical descriptors are comparable in quality to the ones obtained by *Chao-Bin Xue et al* [4] in the same property and compounds set by means of CoMFA and CoMSIA approach. This fact confirms the performance of graph-theoretical indices as useful tools for the structural characterization and prediction of biological properties of thiosemicarbazone derivates.

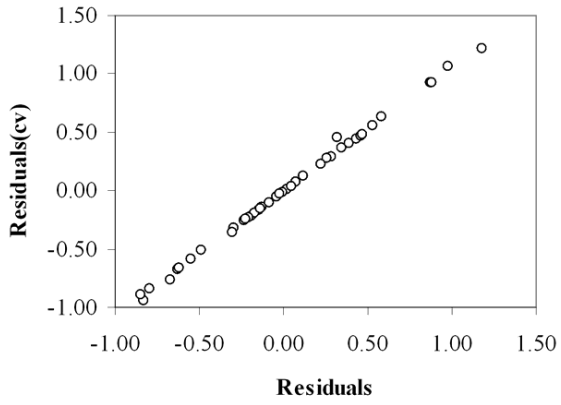
TABLE II

Computational screening from Eq.1 applied to thiosemicarbazones analogues obtained of Scifinder Scholar base.

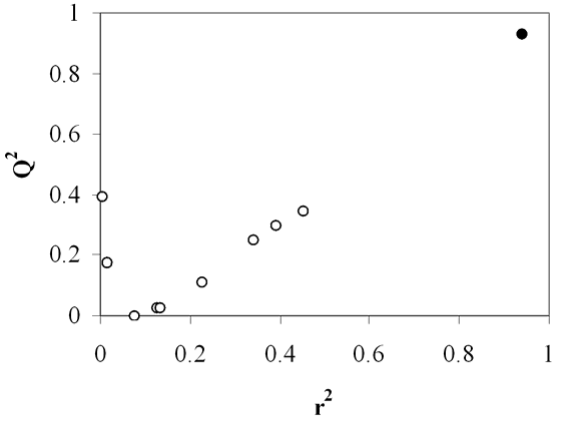


Registry Number	R1	R2	R3	$pIC_{50}calc^{a}$.
18099-70-2	4-OPropyl-i	Me	H	7.90
20718-93-8	4-OPropyl-i	H	H	8.28
22043-09-0	2-OCH_2COOH	H	H	7.34
22043-11-4	4-OCH_2COOH	H	H	8.27
342631-63-4	4-OMe	H	Me	8.02
342631-65-6	4-OMe	H	Ethyl	7.09
342631-67-8	3-OMe	H	Me	7.49
386254-97-3	4-COOH	H	Me	8.04
51146-65-7	2-OH;3-OMe	H	Me	6.98
51236-92-1	2-Me	H	Me	6.81
51236-93-2	2-OH	H	Me	7.46
749210-48-8	3-OH	H	Propyl-n	8.35
866926-11-6	2-OH;3-OMe	H	Me	6.98
869948-13-0	3-OCH_2CHCH_2	H	H	7.73

[a] From the mathematical model, Eq. 1.



a) Cross-validation study



b) Randomization test study

Figura 2. Validation of the mathematical model obtained for the pIC_{50}. a) Residuals obtained with the best regression versus Residuals obtained by cross-validation. b) Correlation coefficients, r^2, versus prediction coefficients, Q^2, obtained by randomization test.

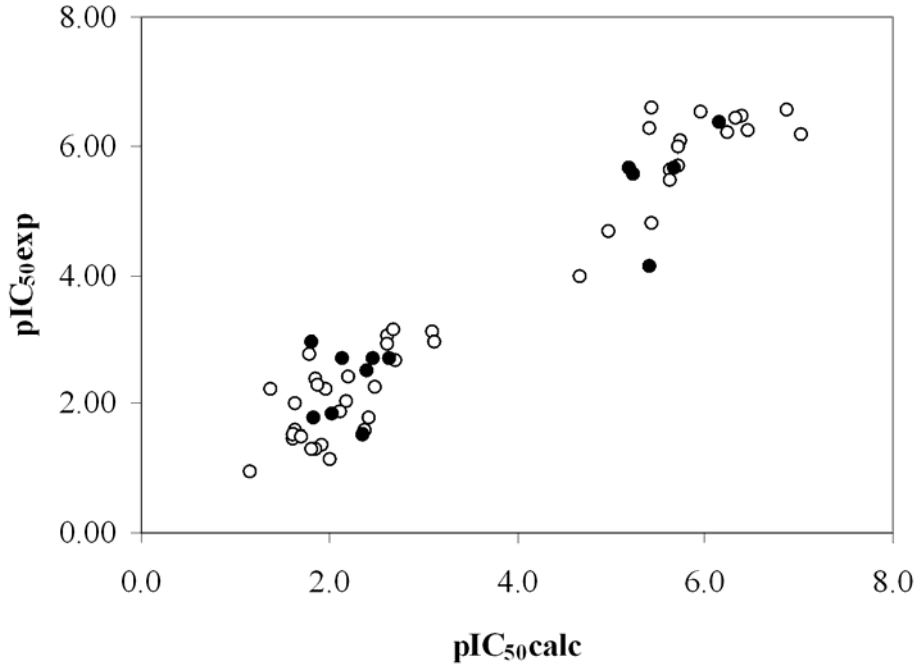


Figura 1. Relationship of pIC_{50exp} with pIC_{50calc} from prediction function obtained using multilinear regression analysis, Eq. 1. Open circles represent predictions for the training set; solid circles represent predictions for the test set.

The second objective of this work was to take advantage of the huge predictive capability of the obtained model, Eq. 1, to predict and optimize the PO inhibitory activity of thiosemicarbazone analogous.

After arranging a molecular library made up of hundreds of compounds (modifying the number and position of the substitutes in R_1, R_2 and R_3, from the base structure, Table 2), we carried out a virtual screening based upon Eq. 1 with the aim of selecting those molecules showing a better activity. Table 2 discloses some of the most significant results. Some interesting comments can be stated from results in Table 2:

- The PO inhibitory activity is enhanced by the addition of groups such as 4-Opropyl-i with pIC_{50}(Nº 20718-93-8)=8.28, 2-OCH_2COOH with pIC_{50}(Nº 22043-09-0)=7.34, 4-OCH_2COOH with pIC_{50}(Nº 22043-11-4)=8.27 or 3-OCH_2CHCH_2 with pIC_{50}(Nº 869948-13-0)=7.73 on position R_1 of the aromatic ring.
- Attaching groups such as methyl, ethyl or n-propyl on R_3, increases significantly the PO inhibitory capability (compare for example: pIC_{50}(N03)=6.28 and pIC_{50}(Nº 749210-48-8)=8.35; pIC_{50}(N10)=6.22 and pIC_{50}(Nº 342631-63-4)=8.02 or pIC_{50}(N09)=5.72 and pIC_{50}(Nº 342631-67-8)=7.49).

These suggestive results need to be corroborated by the corresponding PO activity assays, which should allow the validation or evaluation of the model proposed and serve as a powerful tool for the search of novel compounds with a higher activity on PO.

4. CONCLUSIONS

Molecular topology has been successfully used for finding a QSAR model useful to predict the inhibitory activity of phenoloxidase of a group of benzaldehyde thiosemicarbazone and their derivatives. All the molecular descriptors used herein were graph-theoretical ones. The mathematical model employed in this work retains the main structural features that involve the correlated property, pIC_{50}, and, therefore, can be applicable to the search of new active compounds by virtual screening throughout different databases. We have built up a virtual library with several hundreds of benzaldehyde thiosemicarbazone analogous for seeking within and optimizing the PO activity. Significant improvements in the activity have been obtained.

5. ACKNOWLEDGEMENTS

We thank the Fondo de Investigación Sanitaria, Ministerio de Sanidad, Spain (project SAF2005–PI052128) and Master Internacional en Enfermedades Parasitarias Tropicales, Universitat de Valencia, for support of this work.

6. BIBLIOGRAPHY

[1]. A. Sanchez-Ferrer; J.N. Rodríguez-López; F. García-Cánovas; F. García-Carmona. Tyrosinase: a comprehensive review of its mechanism, *Biochim. Biophys. Acta*, **1247**, 1-11, 1995.

[2]. M.R. Chase; K. Raina; J. Bruno; M. Sugumaran. Purification, characterization and molecular cloning of prophenoloxidases from *Sarcophaga bullata*. *Insect Biochem. Molec.*, **30**, 953-967, 2000.

[3]. M. Ashida; P.T. Brey. Role of the integument in insect defense: pro-phenol oxidase cascade in the cuticular matrix. *Proc. Natl. Acad. Sci. U.S.A.*, **92**, 10698-702, 1995.

[4]. C.B. Xue; L. Zhang; W.C. Luo; X.Y. Xie; L. Jiang; X.T. Lin. 3D-QSAR and molecular docking studies of benzaldehyde thiosemicarbazone, benzaldehyde, benzoic acid, and their derivatives as phenoloxidase inhibitors. *Bioorg. Med. Chem.*, **15**, 2006-2015, 2007.

[5]. Y. Xu; A.H. Stokes; R.Jr. Roskoski; K.E. Vrana. Dopamine, in the presence of tyrosinase, covalently modifies and inactivates tyrosine hydroxylase. *J. Neurosci. Res.*, **54**, 691-697, 1998.

[6]. M. Asanuma; I. Miyazaki; N. Ogawa. Dopamine- or L-DOPA-induced neurotoxicity: the role of dopamine quinone formation and tyrosinase in a model of Parkinson's disease. *J. Neurotox. Res.*, **5**, 165-176, 2003.

[7]. R. Garcia-Domenech; J. Galvez; J.V. de Julián-Ortiz; L. Pogliani. Some new trends in chemical graph theory. *Chem. Rev.*, **108**, 1127-1169, 2008.

[8]. L.B. Kier; W.J. Murray; M. Randić; L.H. Hall. Molecular connectivity V: connectivity series concept applied to density. *J. Pharm. Sci.*, **65**, 1226-1230, 1976.

[9]. O. Ivanciuc; T. Ivanciuc; A.T. Balaban. Quantitative structure-property relationship study of normal boiling points for halogen-/ oxygen-/ sulfur-containing organic compounds using the CODESSA program. *Tetrahedron*, **54**, 9129-9142, 1998.

[10]. H. Hosoya; M. Gotoh; M. Murakami; S. Ikeda. Topological Index and Thermodynamic Properties. 5. How Can We Explain the Topological Dependency of Thermodynamic Properties of Alkanes with the Topology of Graphs? *J. Chem. Inf. Comput. Sci.*, **39**, 192-196, 1999.

[11]. H. Gao; C. Williams; P. Labute; J. Bajorath. Binary quantitative structure-activity relationship (QSAR) analysis of estrogen receptor ligands. *J. Chem. Inf. Comput. Sci.*, **39**, 164-168, 1999.

[12]. C. de Gregorio; L.B. Kier; L.H. Hall. QSAR modeling with the electrotopological state indices: corticosteroids, *J. Comput. Aid. Mol. Des.*, **12**, 557–561, 1998.

[13]. J.V. de Julian-Ortiz; C. de Gregorio Alapont; I. Rios-Santamarina; R. Garcia-Domenech and J. Galvez. Prediction of properties of chiral compounds by molecular topology. *Journal of Molecular Graphics & Modelling*, **16**, 14-18, 1998.

[14]. J. Galvez; J.V. de Julian-Ortiz and R. Garcia-Domenech. General topological patterns of known drugs. *Journal of Molecular Graphics & Modelling*, **20**, 84-94, 2001.

[15]. J. Galvez; R. Garcia-Domenech; J.V. de Julian-Ortiz and R. Soler. Topological Approach to Drug Design. *Journal of Chemical Information and Computer Sciences*, **35**, 272-284, 1995.

[16]. J.V. de Julian-Ortiz; J. Galvez; C. Munoz-Collado; R. Garcia-Domenech and C. Gimeno-Cardona. Virtual Combinatorial Syntheses and Computational Screening of New Potential Anti-Herpes Compounds. *Journal of Medicinal Chemistry*, **42**, 3308-3314, 1999.

[17]. N. Mahmoudi; J.-V. de Julian-Ortiz; L. Ciceron; J. Galvez; D. Mazier; M. Danis; F. Derouin and R. Garcia-Domenech. Identification of new antimalarial drugs by linear discriminant analysis and topological virtual screening. *Journal of Antimicrobial Chemotherapy*, **57**, 489-497, 2006.

[18]. M.J. Duart; G.M. Anton-Fos; P.A. Aleman; J.B. Gay-Roig; M.E. Gonzalez-Rosende; J. Galvez and R. Garcia-Domenech. New Potential Antihistaminic Compounds. Virtual Combinatorial Chemistry, Computational Screening, Real Synthesis, and Pharmacological Evaluation. *Journal of Medicinal Chemistry*, **48**, 1260-1264, 2005.

[19]. DESMOL11, pp. Unidad de Investigación de Diseño de Fármacos y Conectividad Molecular. Facultad de Farmacia, Universitat de Valencia.

[20]. L.B. Kier; T. Di Paolo; L.H. Hall. Structure-activity studies on odor molecules using molecular connectivity. *J. Theor. Biol*., **67**, 585-595, 1977.
[21]. L.B. Kier; L.H. Hall. General definition of valence delta-values for molecular connectivity. *J. Pharm. Sci*., **72**, 1170-1173, 1983.
[22]. G.M. Furnival; R.W. Wilson. Regression by leaps and bounds. *Technometrics*, **16**, 499-511, 1974.
[23]. R.R. Hocking. Criteria for selection of a subset regression: Which one should be used? *Technometrics*, **14**, 967-970, 1972.
[24]. D.M. Allen. The Relationship between Variable Selection and Data Agumentation and a Method for Prediction. *Technometrics*, **16**, 125-127, 1974.
[25]. L.E. S. Wold Statistical validation of QSAR results. In: Van de Waterbeemd H (ed). Chemometric methods in molecular design, VCH, New York, 1995.