J Comput Aided Mol Des (2007) 21:527–538
DOI 10.1007/s10822-007-9136-9

A successful virtual screening application: prediction of anticonvulsant activity in MES test of widely used pharmaceutical and food preservatives methylparaben and propylparaben

Alan Talevi · Carolina L. Bellera · Eduardo A. Castro · Luis E. Bruno-Blanch

Received: 21 June 2007 / Accepted: 8 September 2007 / Published online: 25 October 2007

Abstract A discriminant function based on topological descriptors was derived from a training set composed by anticonvulsants of clinical use or in clinical phase of development and compounds with other therapeutic uses. This model was internally and externally validated and applied in the virtual screening of chemical compounds from the Merck Index 13th. Methylparaben (Nipagin), a preservative widely used in food, cosmetics and pharmaceutics, was signaled as active by the discriminant function and tested in mice in the Maximal Electroshock (MES) test (i.p. administration), according to the NIH Program for Anticonvulsant Drug Development. Based on the results of Methylparaben, Propylparaben (Nipasol), another preservative usually used in association with the former, was also tested. Both methyl and propylparaben were found active in mice at doses of 30, 100, and 300 mg/kg. The discovery of the anticonvulsant activities in the MES test of methylparaben and propylparaben might be useful for the development of new anticonvulsant medications, specially considering the well-known toxicological profile of these drugs.

Keywords Virtual Screening · Anticonvulsants · MES · Methylparaben · Propylparaben

A. Talevi · E. A. Castro
Instituto de Investigaciones Fisicoquímicas Teóricas y Aplicadas (INIFTA), Department of Chemistry, Faculty of Exact Sciences, Universidad Nacional de La Plata, B1900AVV La Plata, Buenos Aires, Argentina

A. Talevi · C. L. Bellera · L. E. Bruno-Blanch (✉)
Department of Biological Sciences, Faculty of Exact Sciences, Universidad Nacional de La Plata, 47 y 115, B1900AVV La Plata, Buenos Aires, Argentina
e-mail: lbb@biol.unlp.edu.ar

Introduction

Virtual screening: a key methodology to identify new leads

A drug discovery cycle demands an estimated average of 9–12 years and at least $800 million economic cost (including capital cost, the theoretical calculation of what Research and Development expenditures might be worth if they were invested elsewhere). In developing countries, the costs and risks of the development of new drugs by pharmaceutical industry are out of balance with the perceived limited profits and the long payback period [1–3]. Since the rapidly rising cost of pharmaceutical Research and Development is due mainly to the increased cost of animal testing and conducting clinical trials [3], the development and application of computational methodologies in the search of new leads arises then as a key strategy to reduce the probability of unsatisfactory results in pre-clinic and clinic evaluation stages.

Virtual screening (VS) comprises a range of computational methods to prioritize the selection and testing of large chemical databases so as to ensure that those drugs which have the highest a priori probabilities of activity are bioassayed first in the drug discovery program [4]. Virtual screening arises as an interesting option for researchers to identify promising new therapeutic agents. The ability of VS methodologies to capture non-obvious requirements for a drug to elicit a specific therapeutic activity allows the identification of new leads with distinctive pharmacological, pharmacokinetic and toxicological profiles.

In this study we have applied in VS a descriptor-based, machine learning methodology based in the application of Linear discriminant analysis (LDA) to derive a discriminant function (DF) with the ability to classify

anticonvulsant and non-anticonvulsant compounds in the MES test.

Motivation

The word epilepsy usually describes a group of common chronic neurological disorders characterized by recurrent (two or more) unprovoked seizures due to excessive neuronal firing or synchronous neuronal activity in the brain [5, 6]. Seizures may vary from the briefest lapses of attention or muscle jerks to severe and prolonged convulsions [7]. Recently, the International League Against Epilepsy (ILAE) and the International Bureau for Epilepsy (IBE) have proposed a definition of epilepsy that does not require two or more seizures for its diagnosis, but at least one seizure together with an enduring alteration in the brain that increases the likelihood of future seizures [6].

It is one of the most common among brain disorders that disturbs human condition, affecting around 1% of world population. Current chemotherapy remains ineffective in up to 30% of the patients. In non-developed countries up to 80% of epilepsy sufferers receive no treatment [8]. Moreover, even the new generation of anticonvulsant agents presents important side effects such as headache, drowsiness, diplopia, ataxia, dizziness, nausea, allergies, blood dyscrasias, and hepatotoxicity [9]. Thus, new medications for the treatment of epilepsy are urgently needed.

Linear discriminant analysis

Linear discriminant analysis is a statistical technique that aims to find the linear combination of features which best separates two or more classes of objects or events. This linear combination of features (the DF) has the general formula:

$$\text{Class} = a_0 + a_1X_1 + a_2X_2 + \ldots + a_nX_n \quad (1)$$

where X_1, X_2,..., X_n represent the chosen descriptors and a_1,...,a_n represent the coefficients of the classification function determined by the least-squares method. Linear discriminant analysis is closely related to regression analysis, which also attempts to express a dependent variable as a linear combination of other features or measurements. In regression analysis, however, the dependent variable is a continuous quantity while for LDA it is a categorical variable (the class label).

Linear discriminant analysis has been applied extensively in drug discovery [10–13]. Our group has also applied this methodology in the identification of new antichagasic and anticonvulsant agents [14–16], recently reporting the anticonvulsant activity of abietic acid [16].

For the application of LDA to drug discovery, the DF should be capable of distinguishing between compounds with and without a biological activity of interest. For this purpose, a training set containing known active and inactive molecules is used to obtain the classification algorithm. After proper validation, the DF is applied to the classification of other chemical structures not included in the training set, selecting those with highest probability of being active at biological tests.

In this paper we describe the discovery of the anticonvulsant activity in the Maximal Electroshock (MES) test of two drugs widely used in pharmaceutical formulations, through application in VS of a DF based in topological descriptors. The discovery is significant since both drugs have a well-known toxicological profile: they present no toxicity in acute and chronic-toxicity studies; they are not carcinogenic or mutagenic [17, 18]. Moreover, they are well absorbed in the gastrointestinal tract. These features make them good candidates to develop new medications. The abundant chemical and biological data on both compounds may help to achieve new anticonvulsants medicines in a relatively short time. This underlines another advantage of VS: discovering new applications of drugs which are already used in therapeutics may shorten the research and development time of new medicines, allowing faster response to overcome urgent therapeutic needs and difficulties.

Methods

Calculation of molecular descriptors

Dragon computer software was used for the computation of 636 constitutional and topological molecular descriptors, and 241 functional group counts and atoms-centered fragments [19]. Topological descriptors are molecular descriptors derived from Graph Theory and account for structural information contained in two-dimensional representation of molecules called graphs, in which atoms are represented as vertices and covalent chemical bonds are represented as edges. Using molecular graphs, the chemical structure of an organic compound may be expressed through graph matrices, polynomials, spectra, spectral moments, counts of paths and walks, and topological indices derived from algebraic operations on any of these elements. They offer a simple way of measuring molecular size, shape, branching and complexity, and molecular similarity.

3D-descriptors were not considered in this study. In this way we ensured that the DF generated would not include

3D features of the training set and that the VS results would not depend on the conformation of the molecules from the database being screened.

Linear discriminant analysis: generation and validation of the DF

The first step in the generation of a good DF is the generation of a dataset with good structural diversity. To ensure this point we selected a training set of 48 compounds: 21 anticonvulsant compounds with proven activity in the MES test and 27 compounds with other biological activities. The 21 anticonvulsants chosen for the model generation are either market drugs or in clinical stage of development and were defined as the ACTIVE category; since they have all been reported as active in the MES test, their mechanism of action includes sodium channels blockade. The 27 compounds with other bioactivities compose the INACTIVE category and were selected following two criterions:

(a) They should belong to a wide range of therapeutic categories and have no previous report of anticonvulsant activity. Among them we have included: antiulceratives, antiinflammatories, antibacterials, antivirals, antiparasitaries, antihypertensives, bronchodilators, antitusives, and others. Although some of the non-active compounds of the training set belong to the same therapeutic category, they do not share the same mechanism of action. For example, enalapril elicits its antihypertensive action through angiotensin-converting enzyme inhibition while candesartan is an angiotensin type I receptor blocker; ranitidine and lansoprazol are both antiulcerative agents, but while ranitidine blocks histamine H_2 receptors that mediate gastric secretion, lansoprazol is a proton pump inhibitor.

(b) They should be structurally diverse compounds. Chemical diversity was checked through calculation of the Tanimoto inter-molecular distances for every pair of molecules in the training set. For this purpose we have used PowerMV software from the National Institute of Statistical Sciences [20]. For the similarity comparison we have used three different molecular substructures included in the software package: atom pairs, Carhart atom pairs and pharmacophore fingerprints. Average inter-molecular distances were computed within the actives, within the inactives and between actives and inactives. Table 1 presents the average Tanimoto inter-molecular distances (along with their standard deviations). Note that, since two identical structures present an inter-molecular distance of 0 and two structures with no substructure in common present an inter-molecular distance of 1, the average inter-molecular distances are quite high and the standard deviations relatively low, which indicates the training set can be considered structurally diverse. The average inter-molecular distance is slightly smaller within the active category (this may be a consequence of the fact that the molecules within the active class share the same action mechanism, thus having more structural features in common) but it is still quite high. The inactive category is even more diverse. The average inter-molecular distance within the inactive category is almost identical (no matter which substructures it is calculated from) than the average distance between the actives and the inactives.

Both categories were codified through a dummy variable (Active compound = 1; Inactive compound = –1). Figure 1 presents the structures of the 21 molecules with anticonvulsant activity in the training set. Figure 2 shows the structures of the 27 non-anticonvulsant compounds in the training set.

Linear discriminant analysis was applied to determine which subset of constitutional and topological descriptors contributes the most to the separation of the ACTIVE and the INACTIVE categories. For this purpose we used

Table 1 Average Tanimoto inter-molecular distances within the active category, within the inactive category and between the active and the inactive classes of the training set

	Within the active category	Within the inactive category	Between the active and the inactive categories
Average Tanimoto Inter-molecular distance—AP	0.73 (0.17)	0.63 (0.14)	0.67 (0.12)
Average Tanimoto Inter-molecular distance—CAP	0.79 (0.13)	0.80 (0.10)	0.82 (0.08)
Average Tanimoto Inter-molecular distance—FP	0.86 (0.23)	0.95 (0.09)	0.94 (0.09)

Three different substructures sets included in PowerMV were used for molecular comparisons: atom pairs (AP), Carhart atom pairs (CAP) and pharmacophore fingerprints (PF). Between parenthesis, the standard deviation for each average inter-molecular distance

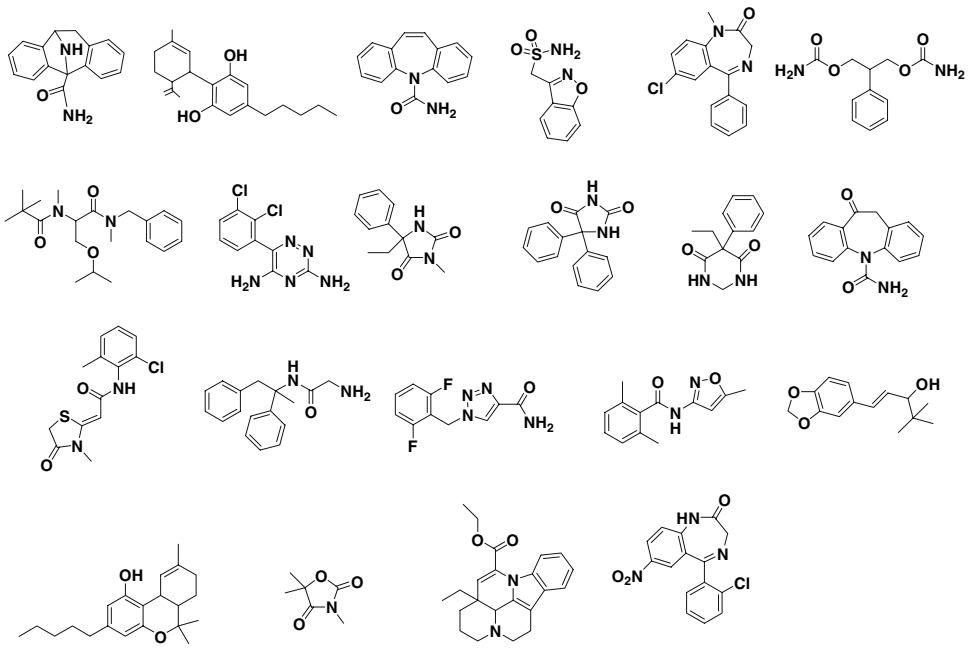


Fig. 1 Structures of the 21 anticonvulsants that compose the active category of the training set

STATISTICA statistical package [21]. Several DFs were generated and the best DF was selected according to the following criteria:

- We search for a DF with low value of the statistical Wilk's U (U = 0 means perfect discrimination between the classes considered; U = 1 means no discriminating ability at all).
- We looked for a DF with high percentage of good classifications in the training set. We paid special attention to the percentage of good classifications in the inactive category, because a misclassification in this group (a false positive) means that an inactive compound will be sent to bioassays, with consequent loss of time and resources.
- Between two DFs with similar performance we preferred that with fewer descriptors, following the principle of parsimony.

Validation of the DF was carried through internal and external validation methodologies [22]. We used Leave-group-out (LGO) cross-validation to assure the model robustness and predictive ability (randomly removing 6 compounds through 8 LGO runs) and 48 randomization tests. In each one of the LGO runs, a group of 6 compounds is extracted from the training set to the test set. The model is generated from the remaining compounds and the new model is used to predict the category of those structures that have been excluded from the model generation.

External validation was performed on a test set of 48 structurally diverse compounds that were not used in the model generation step (20 active compounds with reported activity in the MES test; the remaining 28 belonging to the inactive category and having other bioactivities). The active compounds are in preclinical stage of development and were extracted from different literature sources [23–30]. External validation is the ultimate and more important step to assess model predictive capability.

The ability of the DF to separate anticonvulsants from non-anticonvulsants was also assessed through the generation of a Pharmacological distribution diagram (PDD) [31], which are frequency distribution diagrams of a dependent variable (in our case, value of the DF for each compound of the training set) in which the ordinate represents the expectancies of this variable for every interval. The expectancies of the DF are defined as the probability that a compound will be active or inactive for a range of values of the DF. They are obtained by means of the expressions indicated below. The 100 that appears in the denominator prevents from dividing by zero.

$$Ea = (\text{percentage of activities})/(\text{percentage of inactivities} + 100) \quad (2)$$

$$Ei = (\text{percentage of inactivities})/(\text{percentage of activities} + 100) \quad (3)$$

Pharmacological distribution diagram allows the visual determination of the intervals of the DF where there is a

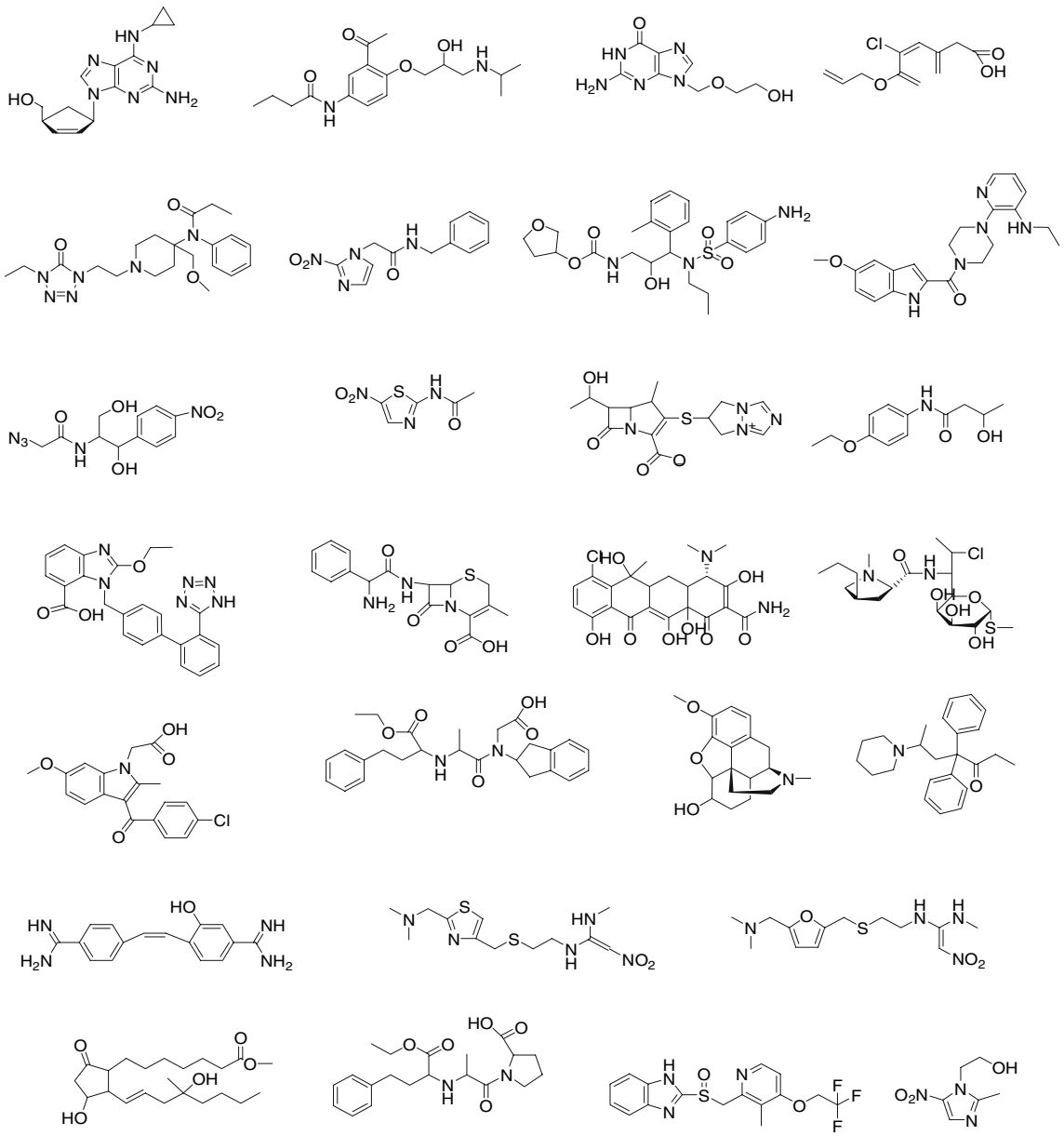


Fig. 2 Structures of the 27 non-anticonvulsants that compose the INACTIVE category of the training set

maximum probability of finding new active compounds and a minimum of encountering inactive ones.

The DF was finally applied in the VS of 10,250 compounds (with exclusion of inorganic compounds) from Merck Index 13th [32]. We have checked that the selected compounds fall within the chemical space defined by the training set compounds. That is, the descriptors included in the DF assume, for the selected compounds, values that fall within the range of values assumed by the descriptors for the training set molecules (this procedure has been referred as applicability domain estimation based on descriptors ranges).

Biological tests

The research was conducted in accordance with the internationally accepted principles for laboratory animal use and care as recommended in the Guide for the Care and Use of Laboratory Animals NIH publication No. 85-23, revised in 1985 and 1996. The pharmacological tests were performed according to standard procedures provided by the Antiepileptic Drug Development (ADD) Program of the National Institute of Neurological and Communicative Disorders and Stroke (NINCDS) [33].

Swiss mice weighing between 18 and 23 g at time of testing, were used as experimental animals. Mice are housed in colony cages on a 12-h light/dark cycle, and provided with food and water ad libitum. The compounds were dissolved in 30% PEG 400. A maximal volume of 10 mL/kg of the freshly made solutions was injected intraperitoneally (i.p.).

Maximal electroshock seizures were elicited in mice by delivering a 60 Hz/50 mA electrical stimulus for 0.2 s via ear clip electrodes. A drop of saline solution was applied on each ear before placing the electrodes ensured adequate electrical contact. In these conditions, maximal seizures are produced in virtually all normal mice. The maximal seizure typically consists of a short period of tonic flexion followed by a longer period of tonic extension of the hind limbs and a final clonic episode. Blockade of the hind limbs tonic extensor component due to the drug treatment is taken as the end point. The tonic component is considered abolished if the hindleg tonic extension does not exceed a 90° angle with the trunk. Positive controls were conducted with two known anticonvulsant drugs, phenytoin and valproic acid, at the time of peak effect of each of them (2 h for phenytoin and 15 min for valproic acid). Negative controls were performed with mice to which no drug was administered.

The RotoRod test was used to determine the possible neurotoxic effects of the drugs tested. A normal mouse can maintain its equilibrium on a rotating rod (6 rpm) for long periods of time. Neurological deficit is indicated by failure to maintain balance on a rotating rod during 1 min.

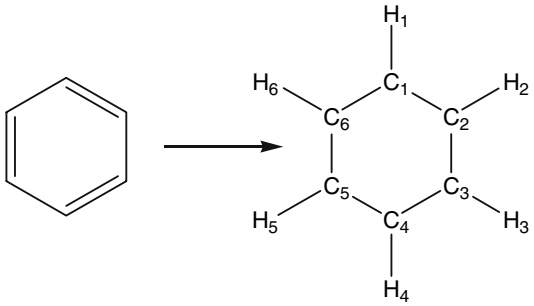
Results

The following DF was selected:

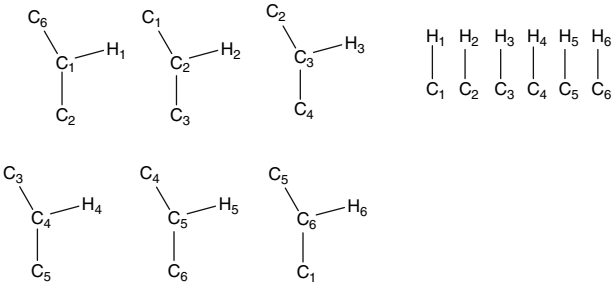
$$\begin{aligned} \mathrm{DF} &= 8011 - 2.206 \times \mathrm{HVcpx} - 4.277 \times \mathrm{BIC2} \\ &\quad + 0.443 \times \mathrm{GATS7e} + 1.089 \times \mathrm{GATS8p} \\ &\text{Wilks' U}: 0.32530 \qquad F(4,43) = 22.297 \\ &\qquad p < 0.0000 \qquad\quad n = 48 \end{aligned} \tag{4}$$

HVcpx represents the graph vertex complexity index [34], BIC2 symbolizes the Bond Information Content (neighborhood symmetry of second order) [35], GATS7e denotes Geary autocorrelation –lag 7, weighted by atomic Sanderson electronegativities and GATS8p stands for Geary autocorrelation –lag 8, weighted by atomic polarizabilities [36].

HVcpx and BIC2 belong to a group of topological indices known as information indices, which are derived from both Graph Theory and Shannon's Information Theory [37]. This family of topological descriptors can be obtained through simple operations on Information Content (IC, also known as Shannon's Entropy). Consider, for example, the hydrogen-explicit graph of benzene:



We can obtain a set of "neighborhoods" of order n; the neighborhood of order n of the vertex v_i is a subgraph of the graph that includes vertex v_i and the set of vertex separated from v_i by a topological distance equal or smaller than n. For example, the set of subgraphs of first order that can be derived from benzene is:



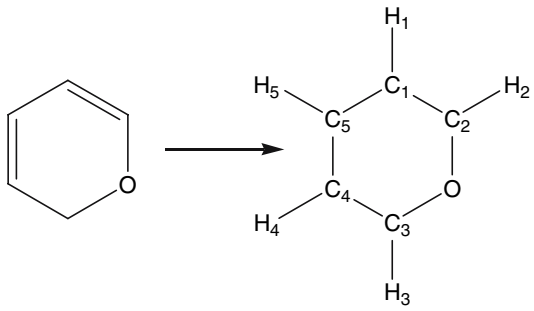
The information content of first order IC_1 is computed as:

$$\mathrm{IC}_1 = -\sum_i p_i \log_2 p_i \tag{5}$$

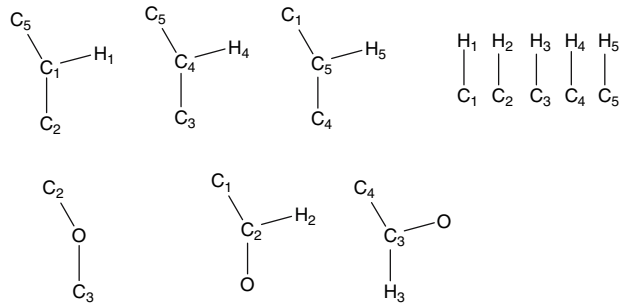
where the sum runs over the i types of neighborhoods of first order. p_i represents the probability of finding the subgraph i in the set of subgraphs of first order derived from the graph. In the case of benzene, since all the neighborhoods of carbon atoms are equivalent, all the neighborhoods of hydrogen atoms are equivalent and the total number of subgraphs of order one is 12, the probability of finding a subgraph representing the neighborhood of a carbon atom is 6/12 and the probability to find a subgraph representing the neighborhood of a hydrogen atom is also 6/12, thus having an IC_1 equal to:

$$\mathrm{IC}_1 = -(6/12 \log_2 6/12 + 6/12 \log_2 6/12) = 1 \tag{6}$$

Let's consider, now, the graph of 2H-pyran:



The set of neighborhoods of order one is now composed by four types of neighborhoods:



and the IC_1 is now calculated as:

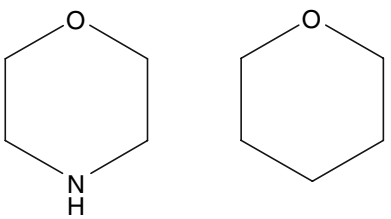
$$IC_1 = -(5/11\log_2 5/11 + 2/11\log_2 2/11 + 1/11\log_2 1/11 + 4/11\log_2 4/11) = 1.81 \quad (7)$$

Information indices, then, give information of the complexity of the graphs, quantifying the diversity of subgraphs of order n that can be derived from a given graph from a given molecular structure. The fact that, according to the selected DF (4) an increment of the value of any of the two information indices is unfavorable to activity may reflect that the anti-MES anticonvulsant drugs from Fig. 1 seem to be structurally less complex than those molecules in the inactive category. For example, anticonvulsant apparently have less heteroatoms than the compounds from other therapeutic categories considered here, thus leading to less diversity within the set of neighborhoods. It may be argued that there are, at least for some of the therapeutic categories included in the inactive group, less complex compounds that could be classified as anticonvulsants by the DF. For example, aspirin is a structurally simpler analgesic than those included in the inactive category of the training set. However, aspirin has reported anticonvulsant activity and could not been included in the training set inactive category [38, 39].

Geary autocorrelations describe the distribution of a given atomic property in the molecule. They are defined as:

$$\mathrm{GATSL}(w) = \frac{1/2K\sum_i \sum_j (w_i - w_j)\delta(L, d_{ij})}{1/(n-1)\sum_i (w_i - \overline{w})^2} \quad (8)$$

where w_i represents the value of the atomic property under consideration for atom i, w_j represents the value of the atomic property under consideration for atom j, L is the lag or order of the autocorrelation and $\delta(L,d_{ij})$ is the krönecker delta between L and d_{ij} (which means only pairs of atoms i and j separated by a topological distance equal to L will be considered for the computation of the descriptor). K is the total number of atom pairs separated by a distance L. The DF includes two Geary autocorrelations considering electronegativity and polarizability as atomic properties, of lag 7 and 8 in that order. According to the DF, the anti-MES anticonvulsant activity increases when the value of GATS7e and GATS8p increases. Note that Geary autocorrelations will assume larger values when the molecule includes atoms with differences in the atomic property under consideration at a topological distance of L. The DF might be reflecting that (a) anticonvulsants, compared to the inactive structures in the training set, tend to possess rather compact structures (they present lower eccentricities, the eccentricity of a vertex v_i defined as the topological distance of the longest path that begins in v_i). Thus, anticonvulsants present less pairs of atoms separated by long inter-atomic distances, therefore presenting a low value of K and larger values of Geary autocorrelations of superior lags and, (b) anticonvulsants tend to present, in their molecular structure, one or two polar regions, with the other part of the molecule being rather hydrophobic, therefore presenting relatively large (w_i–w_j) terms in (8), compared to other structures with a more uniform charge distribution. In the inactive molecules some of those terms may be particularly small due to the fact that heteroatoms are more uniformly distributed in the molecule (then being a higher probability of finding atoms with similar electronegativity or polarizability at a topological distance of 7 or 8 from each other). Consider, for illustrative purposes, structures A and B:



A **B**

GATS3v (Geary autocorrelation of lag 3 weighted by atomic electronegativities) will assume a higher value in the case of B, because the difference between the

electronegativities of oxygen and carbon atoms in B (separated by a topological distance of 3) is quite larger than the difference between the electronegativities of oxygen and nitrogen atoms in A (also separated by 3 edges). GATS3v, effectively, has a value of 0.878 for A and 0.972 for B.

Table 2 presents the values of the DF and the posterior probabilities of belonging to the active category for the 48 compounds which compose the training set. Note that with only four independent variables the DF correctly classifies 93.75% of the compounds from the training set. The PDD confirms the low superposition of the DF values for the two considered categories (Fig. 3). The PDD presents a narrow but rather long tail of the inactive region that extends into the active region. This indicates the model does not present a perfect discrimination between the two classes, and that there is some (likely small) probability of misclassification. Although this is also reflected in the Wilk's U value, which in spite of being small is not 0, therefore indicating no perfect separation of the classes, PDD allows to understand what type of error is more likely to happen when the model is applied. In this case, false positives seem more likely than false negatives.

Table 3 presents the correlation matrix for the four descriptors included in the DF. Observing the principle of non-redundancy, no pair of descriptors is correlated above 0.55. The tolerance of all the variables included in the model (defined as 1 minus the squared multiple correlation of this variable with all other independent variables in the equation) is above 0.5, which corroborates that the included variables are not redundant.

Table 4 presents the results of the LGO cross-validation, as well as the names of the compounds that were removed in each run. Figure 4 shows the results of the randomization test. The original DF always performed much better than the ones obtained through randomization of the dependent, dummy variable (in the randomized models, the Wilk's U statistical assumes values similar to 1, meaning non-discriminating power, and both the F statistical and the percentage of hits decrease considerably). The results of the internal validation indicate that the model is robust and

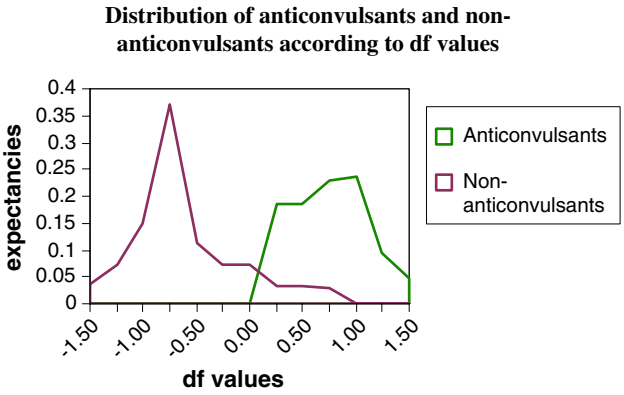


Fig. 3 PDD showing the separation by the DF of the anticonvulsant from the non-anticonvulsants compounds of the training set

Table 2 Discriminant function values for the 48 compounds that compose the training set

Compound	DF value	Prob (Act)	Compound	DF value	Prob (Act)	Compound	DF value	Prob (Act)
ADCI	0.80	99.2	Stiripentol	0.17	73.9	*Bucetin*	–0.71	0.01
Cannabidiol	0.33	88.0	THC	0.30	85.8	*Candesartan*	–0.97	0.00
Carbamazepine	1.21	99.9	Trimetadione	0.79	99.1	*Cephalexin*	–0.82	0.01
Clonazepam	0.16	72.5	Vinpocetine	0.74	98.8	*Clortetraciclin*	–0.48	0.05
Diazepam	1.10	99.9	Zonisamide	0.32	87.2	*Clindamicine*	–1.05	0.00
Felbamate	0.96	99.7	*Abacavir*	–0.47	5.7	*Clometacine*	–0.70	0.01
Lacosamide	0.99	99.7	*Acebutolol*	–1.22	0.1	*Delapril*	–1.04	0.00
Lamotrigine	0.96	99.7	*Acyclovir**	0.21	77.6	*Dihydrocodeine**	0.50	0.95
Mephenytoin	0.51	95.5	*Alclofenac*	–0.83	0.7	*Dipipanone**	0.65	0.98
Oxcarbacepine	0.75	98.9	*Alfentanil*	–1.01	0.2	*Enalapril*	–0.83	0.01
Phenytoin	1.39	100.0	*Aminitrozole*	–0.75	1.1	*Hydroxystilbamidine*	–0.16	0.28
Primidone	0.41	92.1	*Amprenavir*	–1.27	0.1	*Lansoprazol*	–0.99	0.00
Ralitoline	0.07	60.1	*Atevirdine*	–1.54	0.0	*Metronidazol*	–0.17	0.27
Remacemide	0.65	98.0	*Azidamphenicol*	–1.44	0.0	*Misoprostol*	–0.94	0.00
Rufinamide	0.01	52.0	*Benznidazole*	–0.62	2.3	*Nizatidine*	–1.00	0.00
Soretolide	0.68	98.3	*Biapenem*	–0.92	0.4	*Ranitidine*	–0.77	0.01

Compounds from the INACTIVE category are written in italic. An asterisk indicates misclassifications

Table 3 Correlation matrix for the independent variables included in the DF

	HVcpx	BIC2	GATS7e	GATS8p
HVcpx	1.00	0.05	0.12	0.55
BIC2	0.05	1.00	–0.21	–0.02
GATS7e	0.12	–0.21	1.00	0.09
GATS8p	0.55	–0.02	0.09	1.00

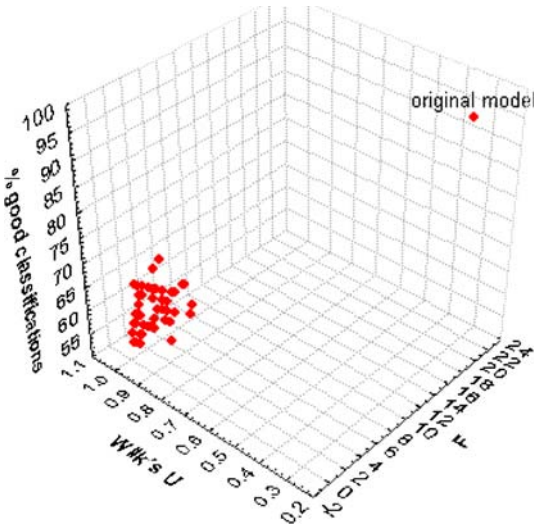


Fig. 4 Results from the randomization test. The real DF clearly outperforms the models obtained by randomization of the dependent, dummy variable

the probability of chance correlation between the independent variables of the DF and the dependent variable is low. The results of the external validation are shown in Table 5.

Table 6 shows the results of the VS for 40 compounds from Merck Index 13th (20 predicted as anticonvulsants and 20 as non-anticonvulsants). Abietic acid anticonvulsant activity has been already demonstrated in mice by our research group [16]. On the basis of the DF value and the extensive use given to methylparaben (MPB) as preservative in food, medicines and cosmetics, we decided to test this compound in the MES test, a preclinic assay which selects anticonvulsant agents whose mechanism of action is related to sodium channels blockade. Results of the biological test are presented in Table 7. Based on the positive results we also decided to test Propylparaben (PPB), a compound structurally close-related to MPB which is usually used in association with it as preservative system. The only previous report that we found of anticonvulsant activity of *p*-hydroxybenzoic acid esters is its ability to control cocaine-induced cramps, through i.v. and intra-arterial administration, which the authors of the report have related to vasodilator effects [40]. Bibliographic search through SciFinder has not pointed out other reports of anticonvulsant activity of MPB and PPB [41]. Results are showed in Table 7. Results of positive controls performed with known anticonvulsant drugs phenytoin and

Table 4 Results of the LGO cross-validation

Removed compounds	*N*	Wilk's U	% Hits training set	% Hits test set
None (original DF)	48	0.33	94	75
Rufinamide, THC, Vinpocetine, Benzmidazole, Bucetina, Clometacine	42	0.33	93	72
Clonazepam, Lacosamide, Oxcarbacepine, Biapenem, Candesartan, Delapril	42	0.36	90	72
ADCI, Mephenytoin, Primidone, Alfentanil, Atevirdine, Enalapril	42	0.35	90	74
Cannabidiol, Diazepam, Phenytoin, Aminitrozole, Amprenavir, Nizatidine	42	0.35	93	74
Carbamazepine, Felbamate, Lamotrigine, Acebutolol, Cephalexin, Dipipanone	42	0.3	93	72
Ralitoline, Soretolide, Stiripentol, Azidamphenicol, Lansoprazol, Ranitidine	42	0.32	90	72
Clonazepam, Lamotrigine, Amprenavir, Clindamicine, Clometacine, Misoprostol	42	0.35	90	80
Remacemide, Trimetadione, Zonisamide, Alclofenac, Clortetracicline, Dihydrocodeine	42	0.29	95	74

The performance of the DFs obtained after random removal of six compounds in each of 8 LGO runs is similar to that of original DF

Table 5 Results of the external validation

Compound	DF value	Compound	DF value	Compound	DF value
534U87	0.26	*N-Phenyl-N′-(3,5-dimethylpyrazole-4-yl)urea*	0.55	*Caromoxirole*	–1.52
N-benzyl-2-ethylamino-3-methoxypropionamide*	–0.18	3,3,3-Trifluoro-2-hydroxy-2-phenyl-propionamide	0.48	*Ceftezole*	–1.58
N-benzyl-2-(1-oxo-ethylamino)-2-phenylacetamide	0.05	3,3,3-Trifluoro-2-hydroxy-N-methyl-2-phenylpropionamide	0.96	*Chlorobutanol**	1.62
N-(2-fluorobenzyl)-2-azaspiro[4.4]nonane-1,3-dione	0.73	THIQ-10c*	–0.07	*Cimetidine*	–1.35
N-(3-fluorophenyl)-2-azaspiro[4.5]decano-1,3-dione	0.30	*Acetoxolone*	–0.30	*Cloxyquin*	–0.37
N-(3-trifluoromethyl phenyl)-2-azaspiro [4.4]nonane-1,3-dione	0.27	*Alacepril*	–1.37	*Cycloserine**	0.45
CFM11	0.29	*Albuterol*	–0.02	*Desomorphine**	0.20
CFM2*	–0.13	*Amoxicillin*	–1.02	*Didanosine*	–0.22
CFM2S	0.20	*Aranidipine*	–0.03	*Dioxaheptyl butyrate**	0.43
1-(4-methylpiperazin-1-yl)-3-(3-chlorophenyl)-pyrrolidine-2,5-dione [78]*	–0.13	*Atenolol*	–1.09	*Doxofylline*	–0.95
1-(4-methylpiperazin-1-yl)-3-(4-chlorophenyl)-pyrrolidine-2,5-dione [78]	0.05	*Atorvastatin*	–1.01	*Ebrotidine*	–1.79
GYKI 53655	0.00	*Balofloxacin*	–0.48	*Efavirenz**	0.54
NBQX*	–0.43	*Bambuterol*	–0.02	*Emtricitabine*	–0.12
4-amino-1,3,5-trimethylpyrazole	0.68	*Benorylate*	–0.80	*Ephedrine**	0.21
N-Cylohexyl-N′-(3,5-dimethylpyrazole-4-yl) thiourea	0.20	*Bromfenac*	–0.44	*Ertapenem*	–2.09
N-(4-Methoxyphenyl)-N′-(3,5-dimethylpyrazole-4-yl) thiourea*	–0.54	*Capravirine*	–0.56	*Famciclovir*	–0.48

Compounds from the INACTIVE category are written in italic. An asterisk beside the compound name indicates misclassifications. Chemical nomenclature is used when generic name is not available

valproic acid are also shown; reference drugs were evaluated at their reported time of peak effect in [42]. Negative control mice showed no protection against MES test. Chemical structures of MPB and PPB are represented in Fig. 5.

Discussion

The anticonvulsant activity of MPB and PPB in the MES test was successfully predicted through a VS methodology and confirmed through biological tests. The two compounds have showed activity at 30 mg/kg, the minimum dose tested according to the NIH Anticonvulsant Drug Development Program. Besides, due to their extensive use as preservatives in cosmetics, foods and medicines, both substances have well-known pharmacokinetic and toxicological profiles (having showed no chronic toxicity and having neither mutagenic nor carcinogenic activity). They have proven long term safety in much higher doses than dose administered in the present study for evaluation in the MES test. Neither carcinogenicity nor long-term toxicity were observed after oral administration of MPB and PPB at doses of 0.9–1.2 and 5.5–5.9 g/kg/day in rats during 96 weeks). Administration of 550 mg/kg/day of methylparaben to pregnant mice and rats had no effect on nidation or maternal and fetal survival, and produce no abnormalities on skeletal tissue. Neither paraben has showed evidence of mutagenesis in the Ames test [17, 18]. These and other well-documented facts may shorten the development of new anticonvulsant medications based in these drugs.

The confirmation of anticonvulsant activity in the MES test of MPB and PPB together with the results of the internal and external validation procedures arises as new evidence of the efficacy of LDA based in topological descriptors in the rational search of new leads through VS.

Table 6 Results of the VS for forty of the screened compounds from Merck Index 13th

Compound	Therapeutic category/use	DF value	Compound	Therapeutic category/use	DF value
Abietic acid	Manufacture of ester gums	0.37	Abecarnil	Anxiolytic	–1.14
Buformin	Antidiabetic	0.16	Abikoviromycin	Antiviral	–0.31
Bufan	Insecticide	0.68	Acediasulfone	Antibacterial	–0.26
Bupivacaine	Anesthetic	0.22	Acetamidoeugenol	Anesthetic	–0.43
Diphenidol	Antiemetic	0.39	Buparvaquone	Antiprotozoal	–0.05
Demegestone	Progestogen	0.43	Cilostazol	Antithrombotic	–1.40
Dropropizine	Antitussive	0.03	Droxidopa	Antiparkinsonian	–0.47
Elenolide	–	0.39	Dulcin	Sweetener	–0.46
Eltoprazine	Serenic	0.26	Eicosapentaenoic acid	Antihyperlipoproteinimic	–1.40
Guanadrel	Antihypertensive	0.19	Histamine	Antineoplastic	–0.23
Hexestrol	Estrogen; antineoplastic	0.15	Hydralazine	Antihypertensive	–0.24
Idazoxan	Antiparkinsonian	0.38	Hydroxyamphetamine	Adrenergic, mydriatic	–0.07
Idrocilamide	Muscle relaxant	0.31	Levobunolol	Antiglaucoma	–0.30
Isosorbide	Diuretic	0.68	Mephenesin	Muscle relaxant	–0.11
Mequitazine	Antihistaminic	0.60	Nialamide	Antidepressant	–1.39
Metapramine	Antidepressant	0.64	Perazine	Antipsychotic	–0.08
Methylparaben	Preservative	0.39	Succinylsulfathiazole	Antibacterial	–1.36
Stanolone	Androgen	0.59	Zalcitabine	Antiviral	–0.11
Tetraethylphthalamide	Analeptic	0.74	Zoledronic acid	Bone resorption inhibitor	–0.03
Zotepine	Antipsychotic	0.13	Zomepirac	Analgesic	–0.65

Therapeutic categories or uses (according to Merck Index) and DF values are presented for each of the compounds. Twenty of them (on the left half of the table) have been classified as potential anticonvulsants, with DF value above zero; the other 20 (on the right half), as non-anticonvulsants

Table 7 Biological data determined (mice, i.p.) for MPB and PPB and the reference anticonvulsants phenytoin and valproic acid

Drug	Dose[a] (mg/kg)	MES test		RotoRod test	
		0.5 h	4 h	0.5 h	4 h
MPB	30	1/3	1/3	0/3	0/3
	100	2/3	0/3	0/3	0/3
	300	1/3	1/3	0/3	0/3
PPB	30	2/3	1/3	0/3	0/3
	100	2/3	1/3	0/3	0/3
	300	1/3	3/3	0/3	0/3
		0.25 h	2 h	0.25 h	2 h
Phenytoin	5	–	3/4	–	0/4
	7	–	4/4	–	0/4
	11	–	4/4	–	0/4
Valproic acid	30	–	–	–	–
	100	3/4	–	0/4	–
	300	3/3	–	2/3	–

[a] Groups of three or four mice are tested at each time and each dose. The table indicates how many of the total number of animals tested at each dose and concentration was protected against electrical induced convulsions by the drug on the left

Possible interactions of MPB and PPB in anticonvulsant medicines which may use them as anticonvulsants (such as carbamazepine pediatric suspension and syrup) should be investigated, although MPB and PPB are present in those medications at lower doses than those administered in the present study for assessment of anti-MES activity.

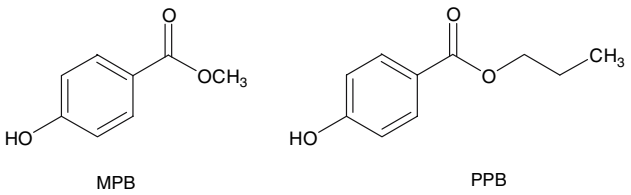


Fig. 5 Structures of MPB and PPB

Acknowledgments We want to thank Guillermina Estiú (University of Notre Dame) and Luciana Gavernet for their valuable help in bibliographic search and Rubén Greco for his valuable assistance in biological tests. A. Talevi thanks CONICET for his Type I Post grade Fellowship. E.A. Castro is a member of CONICET. L.E. Bruno-Blanch thanks the Facultad de Ciencias Exactas de la Universidad Nacional de La Plata (Incentivos UNLP) and the Agencia Nacional de Promoción Científica y Tecnológica (PICT BIB 1728/06-11985).

References

1. Frank RG (2003) J Health Econ 22:325
2. DiMasi JA, Hansen RW, Grabowski HG, (2003) J Health Econ 22:151
3. Dickson M, Gagnon JP (2004) Discov Med 4:172
4. Willett P (2006) Drug Discov Today 11:1046
5. Comission on Epidemiology, Prognosis, International League Against Epilepsy (1993) Epilepsia 34:592
6. Fisher RS, van Emde Boas W, Blume W, Elger C, Genton P, Lee P, Engel J Jr (2005) Epilepsia 46:470
7. World Health Organization (2001) Fact sheet no. 165: Aetiogy, epidemiology and prognosis
8. World Health Organization (2001) Fact sheet no. 265: Mental and neurological disorders
9. Bialer M, Johannessen SI, Kupferberg HJ, Levy RH, Loiseau P, Perucca E (2002) Epilepsy Res 51:31
10. Mahmoudi N; de Julián-Ortiz JV, Ciceron L, Gálvez J., Mazier D, Danis M, Derouin F, García-Domenech R (2006) J Antimicrob Chemother 75:489
11. García-García A, Gálvez J, de Julián-Ortiz JV, García-Domenech R, Muñoz C, Guna1 R, Borrás R (2004) J Antimicrob Chemother 53:65
12. Marrero-Ponce Y, Machado Tugores Y, Pereira DN, Escario JA, Barrio AG, Nogal-Ruiz JJ, Ochoa C, Aran VJ, Martínez-Fernández AR, García-Sánchez RN, Montero-Torres A, Torrens F, Meneses-Marcel A (2005) Curr Drug Discov Technol 2:245
13. Montero-Torres A, García-Sánchez RM, Machado-Tugores Y, Nogal-Ruiz JJ, Martínez-Fernández AR, Aran VJ, Ochoa C, Meneses-Marcel A, Torrens F (2006) Eur J Med Chem 41:483
14. Prieto JJ, Talevi A, Bruno-Blanch LE (2006) Mol Divers 10:361
15. Talevi A, Bellera CL, Castro EA, Bruno-Blanch LE (2006) Drugs Future 31(Suppl. A):XIXth Int Symp Med Chem, p 188
16. Talevi A, Sella-Cravero M, Castro EA, Bruno-Blanch LE (2007) Bioorg Med Chem Lett 17:1684
17. Soni MG, Taylor SL, Greenberg NA, Burdock GA (2002) Food Chem Toxicol 40:1335
18. Soni MG, Burdock GA, Taylor SL, Greenberg NA (2001) Food Chem Toxicol 39:513
19. Talete srl DRAGON for Windows (Software for Molecular Descriptors Calculation) (2003) Version 4.0, http://www.talete.mi.it
20. National Institute of Statistical Sciences. PowerMV Version 0.61, http://www.niss.org/PowerMV
21. Statsoft, Inc. STATISTICA (1996) Version 5.1, http://www.statsoft.com
22. Yasri A, Hartsough D (2001) J Chem Inf Comput Sci 41:1218
23. Obniska J, Kaminski K, Zagorska A, Dzierzawska-Majewska A, Karolak-Wojciechowska J (2006) J Fluorine Chem 127:417
24. Bialer M, Johannessen SI, Kupferberg HJ, Levy RH, Loiseau P, Perucca E (1999) Epilepsy Res 34:1
25. De Sarro G, Ferreri G, Gareri P, Russo E, De Sarro A, Gitto R, Chimirri A (2003) Pharmacol Biochem Behav 74:595
26. Ferreri G, Chimirri A, Russo E, Gitto R, Gareri P, De Sarro A, De Sarro G (2004) Pharmacol Biochem Behav 77:85
27. Obniska J, Jurczyk S, Zejc A, Kaminski K, Tatarczynska E, Stachowicz K (2005) Pharmacol Rep 57:170
28. Schenck HA, Lenkowski PW, Choudhury-Mukherjee I, Ko S, Stables JP, Patel MK, Brown ML (2004) Bioorg Med Chem 12:979
29. Beguin C, LeTiran A, Stables JP, Voyksner RD, Kohn H (2004) Bioorg Med Chem 12:3079
30. Kaymakçıoğlu B, Rollas S, Körceğez E, Arıcıoğlu F (2005) Eur J Pharm Sci 26:97
31. Gálvez J, García-Domenech C, de Alapont G, de Julián-Ortiz JV, Popa L (1996) J Mol Graph 14:272
32. The Merck Index (2001) Encyclopedia of chemicals, drugs and biologicals, 13th edn. Merck and Co. Withehouse Station, New Jersey
33. Stables JP, Kupferberg HJ, Gladding R (1997) In: Avanzini G, Regesta G, Tanganelli P, Avoli M (eds) Molecular and cellular targets for antiepileptic drugs, John Libbey & Company Ltd., London, England, pp 191–198
34. Raychaudhury C, Ray SK, Ghosh JJ, Roy AB, Basak SC (1984) J Comput Chem 5:581
35. Magnuson DR, Harris VK, Basak SC (1983) In: King RB (ed) Studies in physical and theoretical chemistry. Elsevier, Amsterdam, Netherlands, pp 178–191
36. Geary RC (1954) Incorp Stat 5:115
37. Basak SC (1999) In: Devillers J, Balaban AT (eds) Topological indices and related descriptors in QSAR and QSPR. Gordon and Breach Science Publishers, Amsterdam, pp 563–593
38. Srivasta AK, Gupta YK (2001) Indian J Physiol Pharmacol 45:475
39. Walil RS, Patil PA (1995) Indian J Physiol Pharmacol 39:77
40. Bubnoff MV, Schnell D, Vogt-Mycoff J (1957) Arzeimittel Forschung 7:340
41. SciFinder Chemical Abstracts Service (American Chemical Society) (2006) Bibliographic database. http://www.cas.org/SCIFINDER/
42. Levy RH, Dreifuss FE, Mattson RH, Meldrum BS, Kiffin Penry J (1989) Antiepileptic drugs. Raven Press, New York