Pergamon

Bioorganic & Medicinal Chemistry 8 (2000) 2755–2770

BIOORGANIC & MEDICINAL CHEMISTRY

In Silico Studies for the Rational Discovery of Anticonvulsant Compounds

Ernesto Estrada [a,*] and Alfredo Peña [b,†]

[a]*Faculty of Pharmacy, Department of Organic Chemistry, University of Santiago de Compostela, Santiago de Compostela 15706, Spain*
[b]*Fachbereich Chemie, Universität Rostock, D-18051 Rostock, Germany*

Received 3 April 2000; received in revised form 17 July 2000; accepted 27 July 2000

Abstract—Theoretical models to virtual screening and rational design of anticonvulsant compounds based on a topological substructural molecular design (TOSS-MODE) approach are developed. These models, developed on the basis of data sets of great structural variability, permit the classification of compounds as active/inactive anticonvulsants and predict the quantitative anticonvulsant potency of such compounds. The classification model is applied to a virtual screening of anticonvulsant compounds by analyzing a data set of molecules reported in the literature. More than 88% of them were well classified by the current model. Active and inactive fragments are identified by using the present approach. Some of the active fragments are identified in anticonvulsant molecules as potential pharmacophores and one of them is analyzed in detail. The three-dimensional (3-D) features of this fragment are investigated in a series of five anticonvulsant compounds. Some structure–anticonvulsant activity relationships are derived on the basis of the 3-D structure of this fragment and some findings reported in the literature that indicate that it is an important pharamacophore are outlined.

Introduction

An alternative to the 'real' world of synthesis and screening of compounds in the laboratory is an in silico 'virtual' world of data, analysis, hypothesis and design that reside inside a computer. By this means, "the expensive commitment to actual synthesis and bioassay is made only after exploring the initial concepts with computational models and screens."[1] This in silico approach will be used here in order to find predictive models for virtual screening and rational design of anticonvulsant compounds.

The main usefulness of anticonvulsant drugs is the prevention and control of epileptic seizures.[2] However, several epileptic seizures cannot be controlled by currently available anticonvulsants. For instance, before the introduction in clinics of vaiproic acid and carbamazepine in the 1970s, the marketed anticonvulsant drugs protected only 50% of epileptics. Approximately 30–40% of patients with chronic epilepsies are not well-controlled or are largely unresponsive to treatment with carbamazepine.[3] This seizure protection has been improved in last years by the introduction of the 'second generation' of antiepileptic drugs, and good prospects exist with the 'third generation' of such drugs now emerging.[4] However, even today 25% of patients do not respond to current therapeutic agents and novel anticonvulsant compounds with more selectivity and less toxicity are needed for the future.[5,6] The necessity for novel specific and low-toxicity anticonvulsant drugs is evidenced by the several harmful side effects that the individuals under treatment with the currently marketed drugs must tolerate.[7–11] Even some of the most recently marketed antiepileptic drugs have significant adverse effects.[9–11]

A similar situation is observed for the potency of the antiepileptic compounds recently developed. For instance, felbamate, lamotrigine, gabapentin, topiramate, vigabatrin and tiagabine have been shown to be effective in reducing seizures in a number of patients.[4] However, it has been shown that their efficacy is not superior to that of the first generation drugs, such as carbamazepine.[12] For these reasons it is necessary to find a way of maximizing the efficiency of anticonvulsant drugs by 'discovering' novel compounds with improved anticonvulsant potencies and minimal side effects.[13–17]

*Corresponding author. Tel.: +34-981-571764; fax: +34-981-594912; e-mail: estrada66@yahoo.com
†On leave from the Faculty of Chemistry, University of Granma, Bayamo, Cuba.

0968-0896/00/$ - see front matter
PII: S0968-0896(00)00204-2

At present, it is not possible to carry out a rigorous classification of the antiepileptic drugs according to their mechanisms of action.[18,19] However, there exist at least three cellular mechanisms that can be severely affected for the action of these drugs.[2] They are: (i) the use-dependent blockade of voltage-sensitive sodium channels, (ii) the increment of the inhibitory activity of some neurotransmitters, especially the GABA systems, and (iii) the alteration of the calcium channels. The classification of anticonvulsant drugs according to their mechanisms of action is very difficult because some of them act by several mechanisms, new modes of action are discovered, such as for D23129,[20] and in some cases the mechanisms of anticonvulsant action are incompletely described, as for gabapentin.[18,19,21] All these facts make the mechanism-based selection of anticonvulsant drugs a difficult task.[22]

Consequently, most of the antiepileptic drugs actually known have been discovered by empirical ways. There are two common approaches for the 'discovery' of novel drugs: they are the 'design' and the 'screening'.[23] Both approaches have been rapidly developed in recent years in the form of rational drug design methods,[24] combinatorial chemistry,[25] high-throughput biochemical assays,[26] and so forth. The main objective of the present work is to find rationality in the search of novel anticonvulsant drugs. This approach will permit the classification of candidate compounds as anticonvulsant/non-anticonvulsant and will give quantitative estimates of their potencies previous to the pharmacological screenings, identifying the best candidates to be evaluated from thousand of chemicals. This approach will also permit the rational design of novel compounds using the information on the fragment contributions to the anticonvulsant activity.

Methods

TOSS-MODE approach

The Topological Sub-Structural Molecular Design (TOSS-MODE) approach will be used here for the virtual screening and rational design of anticonvulsant compounds. The general principles of this approach have been explained in some detail elsewhere.[27–32] However, an overview of this approach will be given in this work.

The TOSS-MODE approach is based on the computation of the spectral moments of the so-called *bond matrix*.[33,34] The bond matrix is defined as a square and symmetric matrix whose entries are ones or zeros if the corresponding bonds are adjacent or not. The order of this matrix is m, that is the number of bonds in the molecular graph. We say that two bonds are adjacent if they are incident to a common atom. The spectral moments of the edge adjacency matrix are defined as the traces, i.e., the sum of the main diagonal, of the different powers of such a matrix.

In applying the current approach, the following series of steps are carried out:

i select a training data set of active and inactive compounds with great structural diversity (see Data set selection section for explanation of how this data set is selected),
ii draw the hydrogen-depleted molecular graphs for each molecule in the training set,
iii use appropriate bond weights in order to differentiate the molecular bonds (see Bond weights section for explanation of how to select the bond weights),
iv compute the spectral moments of the bond matrix with the appropriate weights for each molecule in the training data set generating a table in which the rows correspond to molecules in the data set and columns correspond to the spectral moments of the bond matrix,
v find a classifier function by using any discrimination statistical technique, such as linear discriminant analysis (see Statistical analysis section for details):

$$P = a_0\mu_0 + a_1\mu_1 + a_2\mu_2 + a_3\mu_3 + a_4\mu_4 + a_5\mu_5 + a_6\mu_6 + b \quad (1)$$

where P is the biological property, μ_k is the kth spectral moment, and the a_ks are the coefficients obtained by the LDA,
vi test the predictive capacity of the classifier model by applying it to an external prediction set,
vii compute the contribution of the different fragments of interest in order to determine if they have positive or negative contributions and their magnitude (see Computation of fragment contributions section for details).

Data set selection

The overall performance of the current method critically depends on the selection of compounds for the training series used to build the classifier model. The most critical aspect of the construction of the training series is to warranty a great molecular diversity in this data set. In order to ensure this molecular diversity, we have selected a data set of 235 compounds, 87 of them used as anticonvulsants in clinic and the rest having a series of other pharmacological uses. The data set of active compounds was selected by considering representatives of most of the different structural patterns and action mechanisms of anticonvulsant activity. For instance, it includes derivatives of phenylpropionamide, valeramide, 1,4-benzodiazepin-2-one, methylcaproic acid, succinimide, benzylcarbamate, hydantoin, barbituric acid, acetylurea, pyrrolo[2,3-d]pyrimidine, cyclohepten-5,10-imine, oxo-morphanthridine, triazolo[4,3-a][1,4]diazepine, phenyl-1,3-indandione, propionohydrazine, imidazoles, and so forth. In order to demonstrate the structural diversity of this data set, we carried out a hierarchical cluster analysis of the set of active compounds. A dendrogram illustrating the results of this cluster analysis developed by using the euclidean distance and the complete linkage is given in Figure 1. It is observed by visual examination of the dendrogram that a great number of different subsets exist, which prove the molecular variability of the compounds in this data set.

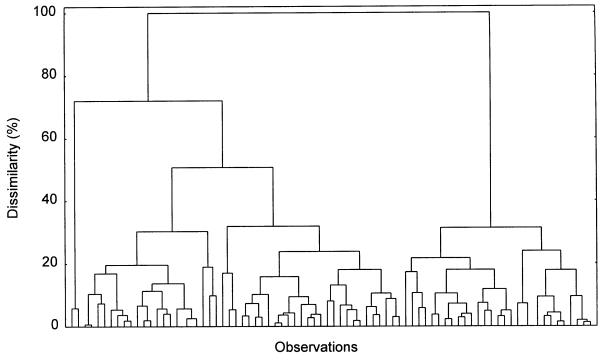


Figure 1. A dendrogram illustrating the results of the hierarchical cluster analysis of the set of active compounds used in the training and prediction sets of the current work.

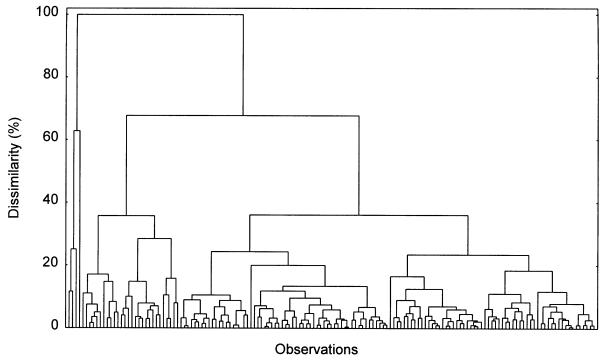


Figure 2. A dendrogram illustrating the results of the hierarchical cluster analysis of the set of inactive compounds used in the training and prediction sets of the current work.

The set of inactive compounds was built as follows. We selected at random 148 drugs with different pharmacological uses. These drugs include antibiotic, antifungal, antiviral, antibacterial, antihypertensive, vasodilator, antilipidepic, antineoplastic, cardiotonic, muscle relaxant, antihistaminic, immunostimulant, antineuralgic, anorexic, sedative, tranquilizer, anxiolytic, antidepressive, etc. It is clear that the declaration of these compounds as 'inactive' anticonvulsant per se does not guarantee anticonvulsant side-effects for some of these drugs that have been left undetected so far. This problem can be reflected in the results of classification for the series of inactive compounds. However, some of these compounds can be detected when they are classified as active by the classifier function. For instance, in developing a classification function for sedative/hypnotic compounds,[29] eucaliptol (1,8-cineol) was used as a member of the inactive set for the development of the classifier function. This compound was 'erroneously' classified as active by the model developed. However, it was found experimentally to be a sedative/hypnotic one.

The molecular diversity of the inactive set of compounds can be observed in the dendrogram obtained after the hierarchical cluster analysis carried out on this set. As can be observed in Figure 2, there are several subsets of structurally different compounds in the inactive data set.

The data set of 235 compounds was randomly divided into two subsets, one containing 57 active and 112 inactive compounds used as training set for developing the classifier function, and the other set containing 30 active and 36 inactive was used as an external prediction set. The compounds in the external prediction set were never used in the development of the quantitative model. All compounds were taken from the Negwer Handbook,[35] where their names, synonyms and structural formulas can be found.

Bond weights

When we are concerned with molecules containing heteroatoms, as most drugs are, an approach based on the use of diagonal weights in the bond matrix is necessary in order to distinguish the different bonds. These weights correspond to different bond properties for organic molecules. For instance, bond lengths were used for describing antifungal activity of benzyl alcohols, boiling points and diamagnetic susceptibilities of alkyl halides.[31] Other bond weights, such as standard bond dipole moments, standard bond polarizabilities, electronegativities, etc., can also be used in this sense.[28] The use of the standard bond dipole moments was first introduced for the discrimination of the sedative/hypnotic activity of a large data set of organic molecules.[29] In this previous work, we showed that this magnitude produced better discrimination than the standard bond distances for the case of large data sets. Consequently, we will use the same values of standard dipole moments in the current work. Some of these values are given in Table 1.

Statistical analysis

In spite of several chemometric techniques to find good discriminant functions exist, such as SIMCA or neural networks, we select the linear discriminant analysis (LDA) in order to generate the classifier function on the basis of the simplicity of the method. On the other hand, the fragment contribution to the pharmacological activity studied can be easily derived from the linear discriminant function, which is difficult or even impossible from highly non-linear discriminant functions such as those derived from neural networks.

All the statistical analyses are carried out by using the STATISTICA package. The discriminant function is obtained by using the stepwise discriminant function analysis, in which a model is 'built' step-by-step.

Table 1. Standard dipole moments for bonds used in the present study

Bond	Dipole (D)[a]	Bond	Dipole (D)[a]	Bond	Dipole (D)[a]
C–C	0.00	C–Br	1.42	C=O	2.40
C–N	0.45	C–S	0.80	C=S	2.00
C–O	0.70	C=C	0.00	C≡C	0.00
C–F	1.39	C=N	1.40	C≡N	3.10
C–Cl	1.47				

[a]Taken from Potapov, V. M. *Stereochemistry*; Mir: Moscow, 1978. It is assumed that the dipole moment of C–H bond is equal to 0.4 D with positive charge on the hydrogen atom.

Specifically, at each step STATISTICA will review all variables and evaluate which one will contribute most to the discrimination between groups. That variable will then be included in the model, and STATISTICA will proceed to the next step. The following statistics are used to 'control' this selection procedure: F to enter, F to remove and tolerance. The stepwise procedure is 'guided' by the respective F to enter and F to remove values. The F value for a variable indicates its statistical significance in the discrimination between groups, that is, it is a measure of the extent to which a variable makes a unique contribution to the prediction of group membership. These parameters were maintained with the default values of 1 and 0, respectively. The tolerance parameter is used to guard against matrix ill-conditioning. It is the proportion of variance that is unique to the respective variable. Here we always used the default value in discriminant analysis for the minimum acceptable tolerance, which is 0.01. The linear regression analysis uses similar parameters to generate the QSAR model.

In order to test the quality of the discriminant function derived we used the Wilks λ and the Mahalanobis distance. The Wilks λ statistic for the overall discrimination can take values in the range of 0 (perfect discrimination) to 1 (no discrimination). The Mahalanobis distance indicates the separation of the respective groups. It shows whether the model possesses an appropriate discriminatory power for differentiating between the two respective groups. The classification of cases was carried out by means of the posterior classification probabilities. Using the Mahalanobis distances to do the classification, we can now derive probabilities. The probability that a case belongs to a particular group is basically proportional to the Mahalanobis distance from that group centroid (it is not exactly proportional because we assume a multivariate normal distribution around each centroid). Because we compute the location of each case from our prior knowledge of the values for that case on the variables in the model, these probabilities are called posterior probabilities. In summary, the posterior probability is the probability, based on our knowledge of the values of other variables, that the respective case belongs to a particular group.

In developing this classification function, the values of 1 and -1 were assigned to active and inactive compounds. By using this model one compound is classified as active if $C > 0.2$ and as inactive if $C \leq -0.2$. The interval between -0.2 and 0.2 is left for not classified compounds (see below). However, we will use the a posteriori probabilities instead of these cut-off values in order to classify the compounds as active/inactive.

Computation of fragment contributions

Each of the μ_k spectral moments given in eq (1) contains structural information on the molecules that can be directly obtained by the following computational approach.[27,28,36] In this approach, we calculate the spectral moments for all the fragments contained in a given substructure, and by difference of these moments we obtain the contribution of the substructure.

The general algorithm followed in this computational approach is as follows. First, we select the substructure whose contribution to the moments we would like to determine. Then we generate all the fragments (subgraphs) which are contained in the corresponding substructure, and calculate the spectral moments for both the substructure and all their fragments. The contribution of the substructure to the spectral moments is finally obtained as the difference between the spectral moments of the substructure and all those from their fragments.

Having the contributions of the different structural fragments in which we are interested, we only need to substitute these contributions into the quantitative model developed to describe the property studied, e.g., model (1), and we obtain the quantitative contribution of the different fragments to P. The creation of a fragment database is then recommended for those fragments that are of interest for the research group using the approach.

Results and Discussion

Development of the classification model

The classification model obtained is given below together with the statistical parameters of the LDA:

$$\begin{aligned} \text{Class} &= 2.040\mu_0 + 2.780\mu_1 - 0.917\mu_2 - 1.590\mu_3 \\ &\quad + 0.370\mu_5 - 4.844 \times 10^{-2}\mu_7 + 8.294 \times 10^{-3}\mu_8 \\ &\quad - 2.016 \times 10^{-8}\mu_{15} - 5.523 \\ N &= 169 \quad \lambda = 0.439 \quad D^2 = 5.69 \quad F(8, 162) = 25.8 \end{aligned} \tag{2}$$

where λ is the Wilks statistic, D^2 is the squared Mahalanobis distance and F is the Fisher ratio. These statistics indicate that model (2) is appropriate for the discrimination of active/inactive compounds studied here. It classifies correctly 80.7% of active and 93.8% of inactive compounds in the training set, for a global good classification of 89.3%. The percentages of false actives and false inactives in the training set are 4.1% (7/169) and 6.5% (11/169), respectively. False actives are those inactive compounds that the model classifies as actives, and the false inactives are active compounds classified as inactives by the model. In Table 2 we give the classification of compounds in the training set together with their posterior probabilities calculated from the Mahalanobis distance.

The most important criterion for the acceptance or not of a discriminant model, such as model (2), is based on the statistics for the external prediction set. Model (2) classifies correctly 76.6 and 97.2% of active and inactive compounds in the prediction set, respectively, for a global classification of 87.9%. The percentages of false actives and false inactives are 1.5% (1/66) and 10.6% (7/66), respectively. In Table 3, we give the classification of compounds in the external prediction set.

A plot of the probability of active for each compound in the training and prediction sets is illustrated in Fig. 3, where the good classification results obtained with the current approach can be observed.

Quantitative structure–anticonvulsant potency study

The second step in the search of antiepileptic drugs is to find a way to predict the anticonvulsant potency of such

Table 2. Results of the classification of compounds in the training set

Training active set								
Compound	C	Probabilities	Compound	C	Probabilities	Compound	C	Probabilities
Atrolactamide	+	66.7	Tetrantoin	+	98.1	Nonapyrimine	+	57.5
Chlorphenacemid	+	65.3	Metphenobarbital	+	91.5	SL 75102	+	93.6
Valnocta mide	—	10.5	Mesuximide	+	56.1	Etizolam	—	26.3
Tenazepam	+	68.8	Valpromide	—	31.7	Aphendion	+	99.5
Atrolactamide	+	66.7	Trifenitoine	+	99.3	Dioxamate	+	97.2
MECap	+	63.1	SQ 10996	+	79.6	Diphoxazide	+	100.0
Brosuccimide	—	47.3	Aminglutethimide	+	98.2	Piperine	+	85.5
Aconv. 1467	NC	52.5	Hibital	+	81.4	Valproic acid	—	44.6
Beclamide	—	41.0	Stiripentol	—	0.6	Bagrosin	+	99.8
Buramate	+	84.7	Oxcarbazepine	+	98.6	RO 5-5807	+	97.9
Aconv. 1894	NC	50.2	Phenytoin	+	100.0	Atolide	+	87.4
Phensuccimide	+	63.9	Estazolam	—	47.4	Paramethadione	—	4.0
Ethotoin	+	79.4	Methylphenitoine	+	99.6	Benzbarbital	+	99.9
Nirvanol	+	98.8	MK-801	+	98.8	Oxitriptyline	+	96.6
Phenobarbital	+	98.7	Cyhetamide	+	99.5	Cipenam	+	100.0
Fenaclon	+	59.0	ICI 45337	+	92.1	Delorazepam	+	76.0
Metetoin	+	90.9	Meprophendiol	—	39.3	Doxenitoin	+	99.9
Pheneturide	+	94.2	IL-40	+	99.6	Cloronafimidone	—	20.5
Hexetal	+	98.1	Antiepilepsirinum	+	56.2	Flufenisal	+	85.5
Training inactive set								
Furanomycin	—	98.1	Vitacampher	—	98.6	Fumaric acid	—	83.3
Oleum tropac.	—	85.1	Oxocampher	—	91.8	Ammoidin	—	92.4
Flucytosine	—	99.2	Dorodosine	—	99.7	Orazamide	—	99.8
Mycosid	—	78.8	Furidanone	—	99.7	Pamabrom	—	99.8
Fuberidazole	—	81.1	Isobromin	—	99.7	Colestipol	—	88.1
Budosamide	—	61.0	Trapidil	—	99.6	Xanthine	—	99.2
Protiofate	—	79.0	Amiquinsin	—	96.5	Cyacetacide	—	98.5
Trichofytocid	—	97.4	Cromodil	—	98.4	Guanamine	—	99.8
LK 274	—	95.4	Bicordin	—	98.0	MJ 10459-2	—	98.5
Ribavirin	—	99.9	Euclidan	—	84.5	Metirosine	—	77.8
BVDU	—	96.6	Bromthymol	—	99.6	Regutensin	—	97.0
Cutison	—	95.6	Diamide	—	98.6	Metildopa	—	91.6
Acedoben	—	75.9	Noxytiolin	—	94.9	Hydralazine	—	91.0
Carbodine	—	97.5	Antibrucellin	—	90.5	Depreton	—	99.7
Dimepranol	—	100.0	Azaserine	—	98.8	Dihydralazine	+	44.1
Riodoxol	—	99.6	Isopropicillin	+	40.9	Pildralazine	—	99.2
Ketoxal	—	99.8	Cefotiam	—	81.7	Oxdralazine	—	95.9
Citenazone	—	99.9	Cefamandole	+	3.8	Clonidine	—	90.8
Aciclovir	—	99.9	Pyracinamide	—	99.7	Lofemizole	—	98.1
Metoprine	—	98.7	Omnasteine	—	98.7	Chlortenoxazine	—	88.9
Guanazol	—	98.6	Valine	—	99.2	Tianafac	—	92.0
Fluoxidin	—	99.8	Chloranil	—	99.9	GOBAB	—	96.2
Canfazolinum	—	99.3	Glutarienol-calcium	—	94.1	Ag 307	—	99.8
BA-1	—	99.1	Toxopyrimidine	—	99.4	Abbot 29590	—	99.6
Bromebric Ac.	—	76.0	Desmethylmethalibur	—	89.8	Tizoprolic acid	—	94.5
Lysepsina	—	95.2	Tioxolone	—	90.1	Triclacetamol	—	96.8
Novenbitol	—	89.4	Redimyl	—	98.7	Tetridamine	—	89.1
Formycin A	—	99.8	Procodazole	—	68.0	Geraniol	—	97.4
Thyaguanosine	—	99.9	Mg Ferulate	NC	54.9	VUFB-7904	—	98.9
Dobutamine	+	35.7	Baclofen	+	30.6	Diethylcarbamazin	—	92.5
Bromotheamin	—	98.5	Morinamide	—	99.6	Tetramizole	—	75.5
Milrinone	—	97.8	Phentermine HCl	—	95.6	Ciclobendazol	—	86.5
Quazodine	—	90.5	Ortetamine	—	95.7	Bitoscanate	—	98.8
D-4026	—	90.6	Benserazide	—	81.2	Antienite	—	98.0
Vindeburnol	+	16.2	Fenchone	—	96.1	Wormin	—	64.8
Aminodal	—	99.5	Eucalyptol	—	84.3	Tiabendazole	—	98.9
Etofilline	—	99.5	Natrii fenbutyras	+	13.4	Planomycin	—	99.6
Amrinone	—	97.3						

Table 3. Results of the classification of compounds in an external prediction set

Test active set								
Compound	C	Probabilities	Compound	C	Probabilities	Compound	C	Probabilities
Phenacemide	+	79.1	Mexitilene	–	0.7	Ethadione	–	1.1
Metharbital	–	28.2	Metindion	+	92.2	Fletazepam	+	100.0
IL-16	–	25.2	Primidone	+	98.7	Progabide	+	89.5
Albutoin	–	1.6	Acetylpheneturide	+	95.9	CGS 9096	+	91.4
Heptobarbital	+	88.9	Diphendionum	+	100.0	Benzamyl	+	89.8
Phenythilone	+	98.0	Premazeoam	–	36.0	Denzimol HCl	+	94.5
Zebromal	–	39.1	Carbamazepine	+	99.1	TV 1901	+	52.68
Tiletamine	+	96.7	Clotiazepam	+	57.2	Clobazam	+	91.97
Felbamate	+	91.3	Remacemide	+	99.9	Oxcarbazepine	+	98.55
D-23129	+	78.4	Dezinamide	+	99.9	Gabapentin	+	93.45
Test inactive set								
Glycarbylamide	–	99.3	Nimazone	–	95.2	Acluracil	–	67.5
Aminopicoline	–	99.5	Fenamole	+	49.1	IMPY	–	99.8
Bijosal	–	96.6	Strinoline	–	97.7	Moroxidine	–	99.5
APPA	–	62.7	R 8231	–	94.0	Guanazole	–	99.6
Progallin-p	–	92.1	Certuna	–	95.2	Tenuazonic Ac.	–	95.9
Isoverin	–	85.0	Drazidox	–	99.6	Tioguanine	–	99.9
Piperazine	–	98.1	Fumigatin	–	98.0	Enoxinone	–	68.6
Aminothiazole	–	99.3	Selectan	–	98.8	Heptaminol	–	99.8
Guanfacine	–	86.5	Thienamycin	–	66.9	Protheobromine	–	100.0
Debrisoquin	–	82.4	Patulin	–	97.3	Nicotinic acid	–	97.0
Alarmine	–	64.4	Azaconazole	–	94.0	Manozodil	–	98.2
Praxadine	–	99.7	Desderman	–	99.8	Salsolidine	–	87.1

compounds. With this objective we conform a data set of 62 compounds reported by different authors in anticonvulsant studies.[37–49] The chemical structures of these compounds are given in Table 4. The logarithm of the oral effective dosis (log ED_{50}) in the maximum electroshock test was used here as the criterion of anticonvulsant potency.

By using the TOSS-MODE approach and the multivariate linear regression analysis we developed the following QSAR model to describe the oral $ED_{50}(MES)$ for these compounds:

$$\log ED_{50}(MES) = 2.459(\pm 0.087) - 3.545 \times 10^{-4}(\pm 2.703 \times 10^{-5})\,\mu_1\mu_2 + 1.83 \times 10^{-9}\mu_1\mu_{12}(\pm 3 \times 10^{-10}) - 0.100(\pm 0.017)HB1$$
$$N = 62 \quad R = 0.964 \quad s = 0.164 \quad F(3, 58) = 258.5$$
$$RMSECV = 0.170 \qquad (3)$$

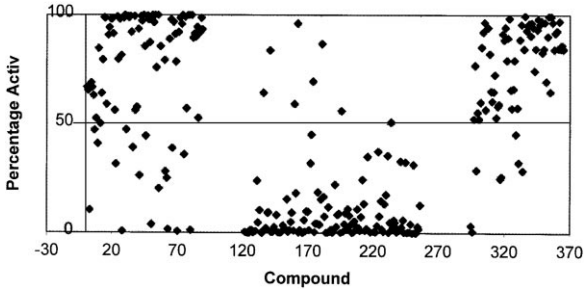


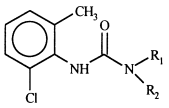
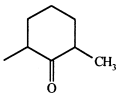
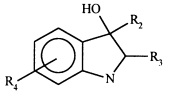
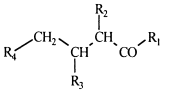
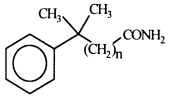
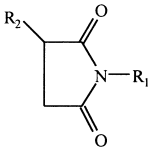
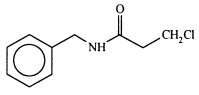
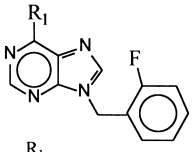
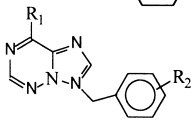
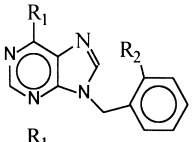
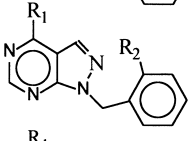
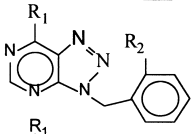
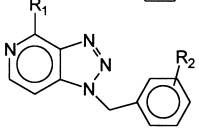
Figure 3. Plot of the probability of active for each compound in the training and prediction sets, as well as those selected from the virtual screening approach. Compounds **1**–**87** are active in training and prediction sets; compounds **122**–**256** are inactive in both training and prediction sets; compounds with numbers over **270** are those selected from virtual screening.

where N is the size of the data set, R is the regression coefficient, s is the standard deviation of the regression, F is the Fisher ratio and $RMSECV$ is the root of the mean squared error of the cross validation carried out by the leave-one-out procedure. The variable $HB1$ included in model (3) is a hydrogen bonding counter. It corresponds to the number of centers in the molecule which are able to form hydrogen bonds as donors or acceptors of the proton, e.g., $HB1 = 1$ for carbonyl group, ethers, tertiary amines, etc., $HB1 = 2$ for nitro group, -OH group, secondary amines, and so forth. The cause for which it was necessary to include this hydrogen bonding indicator variable into the QSAR model describing the anticonvulsant potency is not so clear. This variable was introduced as a structural indicator complementary to a large set of topological indices to describe correctly the partition coefficient of a great data set of organic compounds.[50] By this means we might speculate that this parameter serves for modulating the ADME parameters of the drugs which are closely related to the partition coefficient and have a significant influence on the anticonvulsant potency. In the training series used to develop the classifier function (2), there are many drugs with hydrogen acceptor and donor groups. However, the absence of this indicator variable, $HB1$, in the classifier function (2) makes us think that it does not influence the mechanistic aspects of such drugs, but only influences their ADME parameters related to their potency.

In Table 5 we give the values of the $ED_{50}(MES)$ observed and calculated by model (3) for the 62 compounds used in the development of this QSAR model.

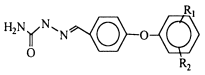
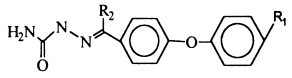
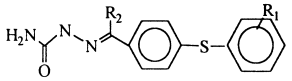
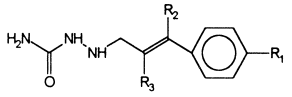
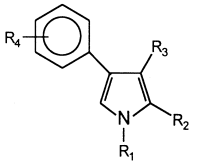
Model (3) explains almost 93% of the variance of the experimental effective dose in the maximum electroshock

Table 4. Structures of compounds selected from the literature for developing the quantitative structure–anticonvulsant potency model

Structure	No.	R_1	R_2	R_3	R_4	*n*	Ref
	**1**	$-CH_3$	$-CH_3$	—	—	—	37
	2	—	$-CH_2-CHO$	H	H	—	38
	3	—	$-CH_2-CO-CH_3$	H	4-Cl, 7-Me	—	38
	4	—	$-CH_2-CO-CH_3$	H	7-Me	—	38
	5	—		=O	H	—	38
	**6**	—	$-CH_2-CO-CH_3$	=O	H	—	39
	7	—	$-CH_2-CO-CH_3$	=O	6-Cl	—	39
	8	—	$-CH_2-CO-CH_3$	—	7-Cl	—	39
	9	—	$-CH_2-CO-CH_3$	—	4-Cl, 7-OMe	—	39
	10	—	$-CH_2-CO-CH_3$	—	4-Cl, 7-Me	—	39
	11	—	$-CH_2-CO-CH_3$	—	4,7-Cl_2	—	39
	**12**	-OH	-Pr	H	CH_3	—	40
	13	$-NH_2$	-*i*-Pr	CH_3	H	—	40
	**14**	—	—	—	—	0	41
	15	—	—	—	—	2	41
	**16**	CH_3	Ph	—	—	—	42
	**17**	—	—	—	—	—	43
	**18**	$-NH_2$	—	—	—	—	44
	**19**	$-NHCH_3$	H			—	44
	20	$-NH_2$	2-F			—	44
	21	$-NHCH_3$	3-F			—	44
	22	$-NHCH_3$	2,5-F_2			—	44
	23	$-NH_2$	2,6-F_2			—	44
	**24**	$-NHCH_3$	F			—	45
	25	$-N(CH_3)_2$	F			—	45
	**26**	$-NHCH_3$	H			—	45
	27	$-NHCH_3$	F			—	45
	28	NH_2	F			—	45
	**29**	$-NHCH_3$	F			—	45
	**30**	$-NH_2$	H			—	46
	31	$-NHCH_3$	H			—	46
	32	$-NH_2$	2-F			—	46
	33	$-NHCH_3$	2-F			—	46
	34	$-NHCH_2CH_3$	2-F			—	46
	35		2-F			—	46
	36	$-N(CH_3)_2$	2-F			—	46

(continued on next page)

Table 4 (*continued*)

Structure	No.	R_1	R_2	R_3	R_4	*n*	Ref
	37	$-NH_2$	2,6-F_2			—	46
	38	$-NHCH_3$	2,6-F_2			—	46
	39	$-NHCH_3$	2,5-F_2			—	46
	**40**	2-F	3-F			—	47
	41	2-F	6-F			—	47
	42	3-F	4-F			—	47
	43	2-CH_3	H			—	47
	44	3-CH_3	H			—	47
	45	4-CH_3	H			—	47
	46	4-C_2H_5	H			—	47
	47	4-*n*-C_3H_7	H			—	47
	48	4-*n*-C_4H_9	H			—	47
	49	4-*t*-C_4H_9				—	47
	**50**	H	CH_3			—	47
	51	F	CH_3			—	47
	52	F	C_2H_5			—	47
	53	Cl	CH_3			—	47
	54	Cl	C_2H_5			—	47
	55	Br	CH_3			—	47
	**56**	H	H			—	47
	57	F	H			—	47
	58	CH_3	H			—	47
	**59**	Br	H	H		—	48
	**60**	H	$COOCH_3$	$N(CH_2CH_2)_2O$	4-Cl	—	49
	61	H	$COOCH_2CH_3$	$N(CH_2CH_2)_2O$	4-Cl	—	49
	62	H	$COOCH_3$	$N(CH_2CH_2)_2O$	4-Br	—	49

test of a data set of anticonvulsant compounds with great structural variability. The standard deviation of the regression model represents less than 15% (14.4%) of the mean response. This is an acceptable value for a biological measurement such as the effective dose. The predictive ability of model (3) is evidenced by the value of the root of the mean squared error in the cross-validation test, which is only 2.9% higher than that of the regression model. Considering all the previously stated features of the QSAR model (3), it can be used as a criterion for evaluating the potency of novel compounds predicted by model (2) as anticonvulsant.

Virtual search of anticonvulsant compounds

One of the main features that any theoretical approach to drug discovery needs to have is the identification of active compounds from databases of chemicals. This search can be understood as an alternative to the screening approaches to drug discovery. In this approach, instead of essaying a large number of chemicals in a series of biological tests (through mass screening, high throughput screening or any other protocol) we 'virtually essay' these compounds by evaluating their activities by the models developed to this effect; this process is known today as virtual screening.[51,52]

In order to prove the possibilities of the TOSS-MODE approach for the virtual screening of anticonvulsant compounds, we have selected a series of 71 compounds whose structures are given in Tables 4 and 6.[45–49,53–59] They have been selected from the current medicinal chemistry literature reporting them as promising anticonvulsant compounds and proving that they are active in one or several anticonvulsant tests. From the 115 compounds reported in Tables 4 and 6 we select only these 71 compounds by using the following criterion. Compounds with numbers **1**–**17** were reported in the literature before 1990 and they were not included in this data set. The reason for doing so is that their structures are similar to several other anticonvulsant drugs included in the training set used to derive the classifier model. Compounds **18**–**29**, even when reported in recent literature, have structures very close to each other. Consequently we only select two of these compounds as representative of this class. The rest of the compounds were used as a data set to be evaluated with the model (2). By this means, the present study is conducted to test the possibilities of the classification model developed here in detecting anticonvulsant[45–49,53–59] compounds of diverse chemical structures. The verification of the predictions carried out by model (2) comes from the recent reports in the literature from where these compounds were selected.

A hierarchical cluster analysis was first realized to observe the molecular diversity of this data set. In the dendrogram illustrating the results of this cluster analysis, we can observe that a few subsets of compounds can be differentiated (see Fig. 4). Even at cut-off points of

Table 5. Experimental and calculated values of the anticonvulsant potency of compounds given in Table 4

No.[a]	Log ED_{50}[b] obsd	Log ED_{50}[c] calcd	Residual[d]
1	1.881	1.814	0.066
2	2.009	1.937	0.072
3	1.748	1.939	−0.191
4	2.061	1.942	0.119
5	2.233	1.977	0.256
6	2.301	2.152	0.149
7	2.428	2.152	0.276
8	2.377	2.152	0.225
9	1.940	2.150	−0.210
10	2.097	2.046	0.051
11	1.748	2.050	−0.302
12	1.763	2.050	−0.286
13	1.957	2.009	−0.052
14	1.809	1.963	−0.155
15	1.955	1.966	−0.010
16	2.049	2.197	−0.148
17	2.262	2.121	0.142
18	1.079	0.966	0.113
19	1.204	1.201	0.003
20	0.903	1.004	−0.101
21	1.301	1.044	0.257
22	1.000	0.917	0.083
23	0.845	0.905	−0.060
24	1.079	1.054	0.025
25	1.079	1.144	−0.065
26	1.176	1.059	0.117
27	0.903	0.917	−0.014
28	1.146	0.857	0.289
29	0.903	0.860	0.043
30	1.301	1.010	0.291
31	1.000	1.072	−0.072
32	1.061	0.871	0.190
33	0.663	0.930	−0.268
34	1.000	0.861	0.139
35	0.799	0.581	0.218
36	0.820	1.017	−0.197
37	0.568	0.748	−0.179
38	0.724	0.805	−0.081
39	0.633	0.743	−0.109
40	0.789	0.631	0.158
41	0.743	0.618	0.125
42	0.375	0.541	−0.166
43	0.752	0.721	0.031
44	0.487	0.698	−0.210
45	0.535	0.695	−0.160
46	0.812	0.621	0.191
47	0.420	0.542	−0.122
48	0.507	0.462	0.045
49	0.225	0.425	−0.200
50	0.988	0.725	0.263
51	0.528	0.582	−0.054
52	0.628	0.521	0.107
53	0.465	0.578	−0.113
54	0.461	0.518	−0.057
55	0.642	0.520	0.123
56	0.632	0.836	−0.204
57	0.697	0.650	0.047
58	0.562	0.697	−0.134
59	1.097	1.302	−0.205
60	1.057	1.069	−0.012
61	0.898	1.008	−0.110
62	1.025	0.975	0.050

[a]Number of compounds given in Table 4.
[b]Experimental value of the effective dose in the maximum electroshock seizure test in mg/kg taken from references given in Table 4.
[c]Values calculated by eq (3).
[d]Observed minus calculated values.

67–70% of similarity between structures, i.e. 30–33% of dissimilarity between structures, seven different subsets can be observed. This cut-off point has been recommended previously in analysis of molecular diversity in pharmaceutical lead discovery.

In Table 7 we give the classification of these 71 compounds as active/inactive for anticonvulsant action. Two compounds were not classified by model (3), representing less than 3% of the data set, and another eight compounds (11.3%) were classified as inactive (false inactive). In closing, more than 88% of the compounds screened were detected by the model (3) as anticonvulsant.

A pictorial representation of the good classification of these compounds is given in Fig. 3, where the compounds in the training and prediction sets (active and inactives) are also shown. These compounds taken from the latest literature can be included in the training series in order to derive more robust classification models. By this means, the derivation of the classifier model is considered as an iterative process in which novel compounds with novel structural features are incorporated into the training set for improving the quality of the models so developed.

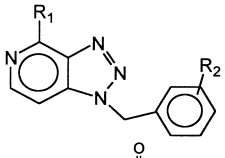
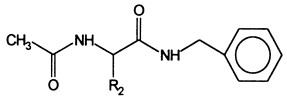
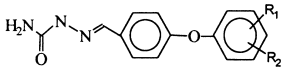
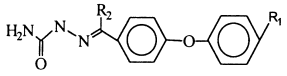
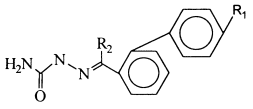
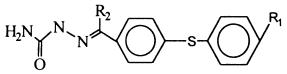
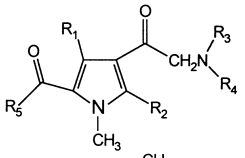
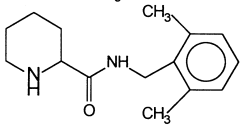
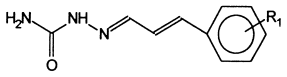
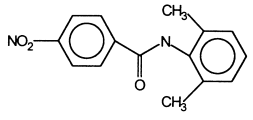
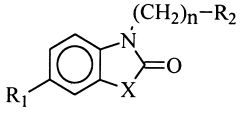
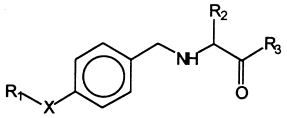
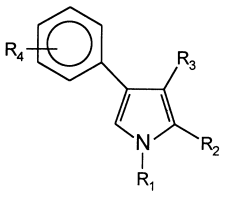
Active/inactive anticonvulsant fragments

As we previously explained, the TOSS-MODE approach is able to compute the contribution of any structural fragment (real or hypothetical) to the biological property or activity studied.[27–32,36] In the present case, we can find the positive and negative contributions of such fragments to the development of the anticonvulsant activity. These fragments will be named here as active and inactive ones, respectively. The presence of active fragments does not presuppose the development of the anticonvulsant activity per se, because it is well known that the activity is the consequence of the sum of the contributions of all fragments in the molecule. However, the identification of such fragments can be of interest for the development of bi-dimensional (2-D) pharmacophores as well as for orientating the synthesis of novel compounds in which some of these active fragments can appear combined in one common chemical structure. On the other hand, excluding inactive fragments from the structures of novel compounds expected to be anticonvulsant should be done with care. It is clear that these fragments decrease the probabilities for the development of the biological activity but the possible contribution of these fragments for other necessary processes in the development of a drug, such as distribution, metabolism, excretion, etc., should be taken into account.

In Figure 5, we show the structures of a series of fragments selected from our fragment database. The contributions to the anticonvulsant activity of these fragments were computed by using the model (2). These quantitative contributions are given in Table 8.

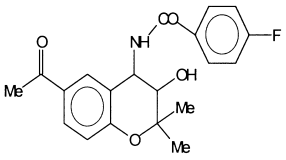
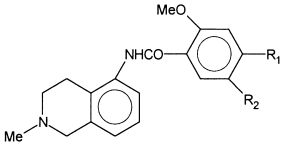
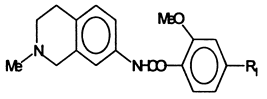
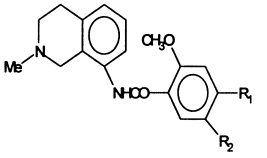
As can be seen in Fig. 5 and Table 8, there are a few active fragments that commonly appear in anticonvulsant

Table 6. Structures of compounds selected from the literature used in the virtual screening of anticonvulsants

Structure	No.	R_1	R_2	R_3	R_4	R_5	X	Ref
	**63**	-$NH(CH_2)_2CH3$	2-F	—	—	—	—	46
	**64**	—	-CH_2OCH_3	—	—	—	—	53
	65	—	-$NHOCH_3$	—	—	—	—	53
	66	—	-$N(CH_3)OCH3$	—	—	—	—	53
	**67**	2-F	5-F	—	—	—	—	47
	68	2-F	4-Cl	—	—	—	—	47
	69	4-*O*-*n*-C_4H_9	H	—	—	—	—	47
	**70**	H	C_2H_5	—	—	—	—	47
	71	Br	C_2H_5	—	—	—	—	47
	**72**	F	CH_3	—	—	—	—	47
	**73**	H	C_2H_5	—	—	—	—	47
	74	H	C_2H_5	—	—	—	—	47
	**75**	H	H	-$(CH_2)_4$-		*p*-ClPh	—	54
	76	H	H	-C_2H_5	-C_2H_5	*p*-ClPh	—	54
	77	CH_3	CH_3	-$(CH_2)_4$-		*p*-ClPh	—	54
	78	H	H	-$(CH_2)_6$-		*p*-ClPh	—	54
	79	CH_3	CH_3	-$(CH_2)_4$-		2,4-Cl_2Ph	—	54
	80	CH_3	CH_3	-C_2H_5	-C_2H_5	*p*-ClPh	—	54
	**81**	—	—	—	—	—	—	55
	**82**	OC_6 $H_6$4-F	—	—	—	—	—	48
	**83**	—	—	—	—	—	—	56
	**84**	H	CH_3 $n=0$	—	—	—	S	57
	85	C_2H_5CO	C_2H_5CO $n=0$	—	—	—	S	57
	86	C_6H_5CO	C_4H_8N $n=2$	—	—	—	S	57
	87	C_2H_5CO	C_4H_8N $n=2$	—	—	—	S	57
	88	C_2H_5CO	$C_5H_{10}N$ $n=2$	—	—	—	S	57
	89	C_3H_7	$C_5H_{10}N$ $n=2$	—	—	—	S	57
	90	C_6H_5CO	$(CH_3)_2N$ $n=2$	—	—	—	O	57
	91	$CH_2C_6H_5$	C_4H_8NO $n=2$	—	—	—	O	57
	92	C_2H_5CO	$C_5H_{10}N$ $n=2$	—	—	—	O	57
	**93**	-$CH_2C_6H_4$-3-Cl	H	NH_2	—	—	O	58
	94	C_6H_5	CH_3	NH_2	—	—	-CH_2O-	58
	95	C_6H_5	CH_3	$NHCH_3$	—	—	-CH_2O-	58
	96	C_6H_5	CH_2OH	$NHCH_3$	—	—	-CH_2O-	58
	97	C_6H_5	C_2H_5	NH_2	—	—	-CH_2O-	58
	98	C_6H_5	CH_3	NH_2	—	—	CH_2CH_2-	58
	99	C_6H_5	CH_3	NH_2	—	—	-CH_2S-	58
	100	3-F-C_6H_5	CH_3	NH_2	—	—	-OCH_2-	58
	101	C_6H_5	CH_3	NH_2	—	—	-$(CH_2)_3O$	58
	102	3-$CF_3C_6H_4$	CH_2OH	$NHCH_3$	—	—	-CH_2O-	58
	103	3-F-C_6H_4-	CH_2OH	$NHCH_3$	—	—	-CH_2O-	58
	**104**	H	$COOCH_3$	$N(CH_2CH_2)_2HC_6H_5$	4-Cl	—	—	49
	105	$COCH_3$	$COOCH_3$	$N(CH_2CH_2)_2O$	4-Cl	—	—	49
	106	COC_6H_5	$COOCH_3$	$N(CH_2CH_2)_2O$	4-Cl	—	—	49

(continued on next page)

Table 6 (*continued*)

Structure	No.	R_1	R_2	R_3	R_4	R_5	X	Ref
	**107**	—	—	—	—	—	—	59
	**108**	Et	H	—	—	—	—	59
	109	*t*-Bu	H	—	—	—	—	59
	110	H	Cl	—	—	—	—	59
	111	Me	Cl	—	—	—	—	59
	112	Et	Cl	—	—	—	—	59
	**113**	*t*-Bu	—	—	—	—	—	59
	114	H	Cl	—	—	—	—	59
	**115**	NH_2	Cl	—	—	—	—	59

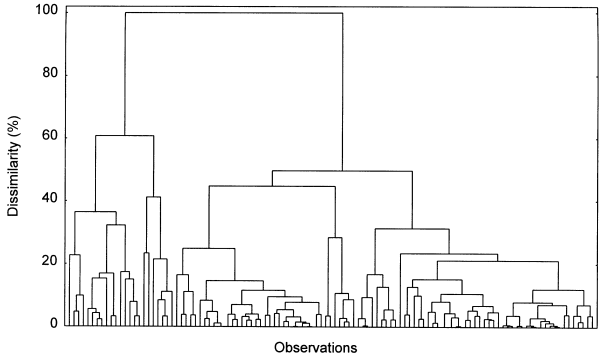


Figure 4. A dendrogram illustrating the results of the hierarchical cluster analysis of the set of compounds used for evaluating the predictive capacity of the model for virtual screening.

compounds, such as the cyclic fragments F6 and F7 present in phenytoin and derivatives. The active fragment F19 is an interesting one. This substructure appears as the only heteroatom-containing fragment in the molecule of milacemide (2-(*n*-pentylamino)acetamide), which possesses a weak anticonvulsant activity acting as an antagonist of iv-bicuculline (iv-BIC)-induced seizures with a novel mechanism of action.[60–62] It was further shown that the very similar molecule of α-methylmilacemide was also active in the iv-BIC test.[63] By a systematic modification of this structure, Pavarello et al.[58] found a new lead compound having this fragment in its molecular structure: 2-[[4-(3-chlorobenzoxy)benzyl]amino]-acetamide. This lead compound permitted the synthesis of many potent anticonvulsant compounds (compounds **93**–**103** in Table 6), such as PNU-151774E (compound **100** in Table 6), that have a potent activity and outstanding therapeutic indices in different animal tests.[58] Several other anticonvulsant compounds have this fragment in their structures, such as rufinamide (CGP 33101)[64] and levitiracetam (ucb LO59).[65]

As a final example of the possibilities that the present method of finding active/inactive fragments brings to the medicinal chemist, we select to study the active fragment F38. This fragment appears in the molecule of Flupirtine, a drug first developed as an analgesic but showing broad anticonvulsant properties in animal models of epileptic seizures, which acts not only via known but also new modes of action.[66] This compound is now under development as an anticonvulsant drug with the name D-23129.[67–69] It was detected as an active compound in the external prediction set used in the present work by means of model (2) (see Table 3). The distance between the nitrogen atom of the free amino group and the carbonyl oxygen in this compound is 2.91 Å as determined by X-ray diffraction.[70] This distance falls within the range varying from 2.76 to 4.51 Å postulated as important for the anticonvulsant activity.[71] For instance, this distance is 2.76 Å in Trihexylphenidyl, 3.35 Å in Diazepam, 4.13 Å in Phenacemide and 4.51 Å in Phenobarbital (see Fig. 1 in ref 66)

In order to study the 3-D features of the fragment F38 in anticonvulsant molecules, we selected five compounds having this substructure, which have been predicted as active by model (2): D-23129 (Table 4), compound **64** (Table 6),[53] SB-204269 (compound **107** in Table 6).[59] Clobazam (Table 4) and compound **88** (Table 6).[57] The molecular structures of these compounds are shown in Figure 6. Compound SB-204269 has an oxygen atom instead of nitrogen at the end of the fragment. However, we decided to include it taking into account the

Table 7. Results of the virtual screening of anticonvulsant compounds selected from the recent literature (see text for selection criterion)

No.[a]	Obsd.[b]	Predicted[c]	Probability[d]
25	+	—	3.4
28	+	—	0.7
30	+	U	52.4
31	+	+	76.9
32	+	—	28.8
33	+	+	55.2
34	+	U	52.4
43	+	+	59.8
46	+	+	85.2
47	+	+	92.1
48	+	+	96.6
50	+	+	56.7
52	+	+	94.1
54	+	+	94.2
56	+	+	82
60	+	+	64.9
61	+	+	60.4
62	+	+	64.7
63	+	+	72.5
64	+	+	53.0
65	+	+	58.1
66	+	+	58.9
67	+	—	25.1
68	+	—	26
69	+	+	91.2
70	+	+	88.6
71	+	+	94.2
72	+	+	79.1
73	+	+	89.4
74	+	+	98.3
75	+	+	65.7
76	+	+	57.2
77	+	+	66.0
78	+	+	79.2
79	+	—	45.5
80	+	+	57.3
81	+	—	32.3
82	+	+	88.7
83	+	+	95.7
84	+	—	28.6
85	+	+	84.5
86	+	+	99.6
87	+	+	92.9
88	+	+	95.0
89	+	+	96.0
90	+	+	90.7
91	+	+	96.4
92	+	+	94.2
93	+	+	74.4
94	+	+	84.4
95	+	+	93.4
96	+	+	98.0
97	+	+	95.6
98	+	+	97.3
99	+	+	82.7
100	+	+	84.5
101	+	+	96.5
102	+	+	69.6
103	+	+	90.2
104	+	+	99.2
105	+	+	64.6
106	+	+	99.4
107	+	+	96.5
108	+	+	84.2
109	+	+	91.8
110	+	+	92.7
111	+	+	84.6
112	+	+	97.0
113	+	+	84.3
114	+	+	85.9
115	+	+	84.2

[a]Number corresponds to those given in Tables 4 and 6.
[b]Observed anticonvulsant activity as reported in references given in Tables 4 and 6.
[c]Predicted anticonvulsant activity by model (2).
[d]A posteriori probability as anticonvulsant.

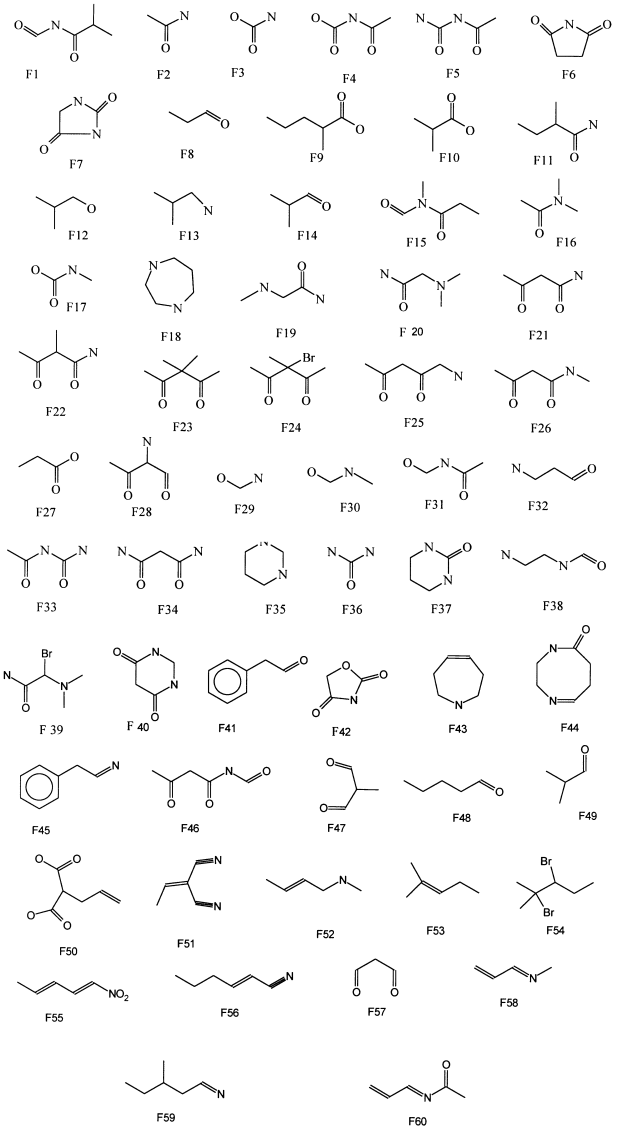


Figure 5. Structures of selected fragments for which their contributions to the anticonvulsant activity are evaluated.

possible bioisosterism between these two fragments. The contribution of this 'bioisosteric' fragment to the anticonvulsant activity is 0.062, which is exactly the same as that of the original fragment.

The semiempirical quantum chemical method PM3[72] was used in order to optimize the chemical structures of these molecules by using the program MOPAC 6.0.[73] After the full geometry optimization of the structures, we proceeded to superpose their structures with a maximum overlap for the six atoms of the fragment F38. As can be seen in Figure 7, the atoms of the fragment studied are positioned about the same spatial region with deviations not greater than 0.6 Å. The average distance between the oxygen of the carbonyl group and the terminal nitrogen atom (or oxygen for compound SB-204269) is 3.718 ± 0.338 Å while the distance between both nitrogen atoms (or nitrogen–oxygen for compound SB-204269) is 3.177 ± 0.296 Å (see also Table 9). The

Table 8. Contribution of some selected fragments to the anticonvulsant activity

Fragment[a]	Contribution
F1	−0.014
F2	3.21
F3	−0.037
F4	−0.037
F5	−0.013
F6	0.201
F7	0.174
F8	1.174
F9	−0.005
F10	−0.806
F11	0.196
F12	0.791
Fl3	0.991
F14	−0.504
F15	−0.02
F16	−0.976
F17	−1.304
F18	−0.323
F19	0.071
F20	0.172
F21	0.106
F22	−0.17
F23	−0.256
F24	−0.261
F25	−0.014
F26	−0.185
F27	−0.056
F28	0.097
F29	−4.591
F30	1.895
F31	0.143
F32	−0.256
F33	−0.013
F34	0.082
F35	−0.185
F36	3.057
F37	0.057
F38	0.062
F39	−0.194
F40	0.017
F41	0.000
F42	0.127
F43	−0.557
F44	0.102
F45	0.000
F46	−0.001
F47	0.189
F48	0.063
F49	−0.504
F50	0.000
F51	−0.022
F52	0.064
F53	−0.417
F54	−0.162
F55	−0.001
F56	0.000
F57	−0.281
F58	−0.369
F59	−0.001
F60	−0.010

[a]Fragment structures are given in Figure 5.

standard deviations of the distances measured are not greater than 10% of their average values, which indicates a clear superposition of the selected structural features.

It appears to be interesting to investigate if these fragment similarities, and others that can be found with the use of the TOSS-MODE approach, are responsible for the anticonvulsant activity of these compounds even when they act by different mechanisms. The following observations taken from experimental results published in the literature point in the same direction as our results. Concerning 2(3*H*)-benzoxazolone and 2(3*H*)-benzo-thiazolone derivatives, Ucar et al. have stated that "the amino substituents on the 3-position of 6-acyl derivatives enhanced significantly the anticonvulsant activity of these compounds".[57] These amino substituents on 3-position introduce a fragment F38 when $n=2$ (see structures **86**–**92** in Table 6). These authors also observed that "the lengthening of the alkyl chain, situated between the nitrogen of the heterocycle and the amino moiety, from two to three methylene groups increased significantly the neurotoxicity, and no separation was observed between anti-MES activity and neurotoxicity for these compounds".[57] On the other hand, concerning compound SB-204269 Chan et al. have stated that "the 3-hydroxyl group appears essential for anticonvulsant activity".[59] A final example is provided by a recent work of Jimonet et al.[74] reporting the synthesis and anticonvulsant activity of Riluzole derivatives. They studied a series of sulfur-containing 3-substituted benzothiazolines, finding that the more potent analogues have two carbon atoms separating a nitrogen in the heterocycle from a sulfur in the exocyclic alkyl chain. This fragment corresponds to another bioisoster of the fragment F38 with a structure HNC=N–C–C–S instead of O=C–N–C–C–N or O=C–N–C–C–O previously studied here. These authors concluded that lengthening or shortening the alkyl chain by one carbon atom resulted in a loss of anticonvulsant effect.[74] All these experimental observations are in complete agreement with our findings concerning the possible role of the studied fragment in the development of the anticonvulsant activity of such compounds. These results show that the present approach is useful in identifying 2-D active fragments responsible for a specific biological activity, and more importantly that these fragments are directly related to 3-D substructures that can correspond with important pharmacophores.

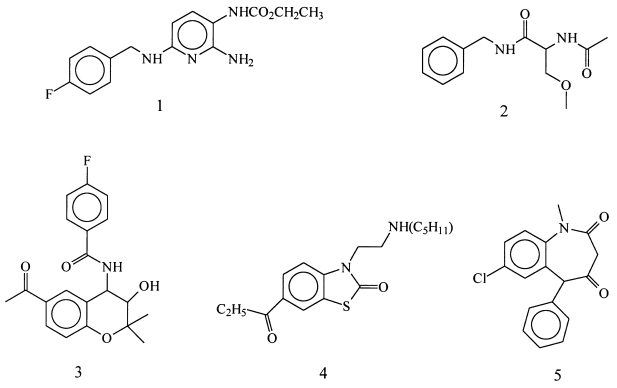


Figure 6. Chemical structures of five anticonvulsant compounds (D-23129 (1), compound **64** in Table 6 (2), SB-204269 (3), compound **88** in Table 6 (4) and Clobazam (5) with a common structural fragment detected by model (2) as an active fragment.

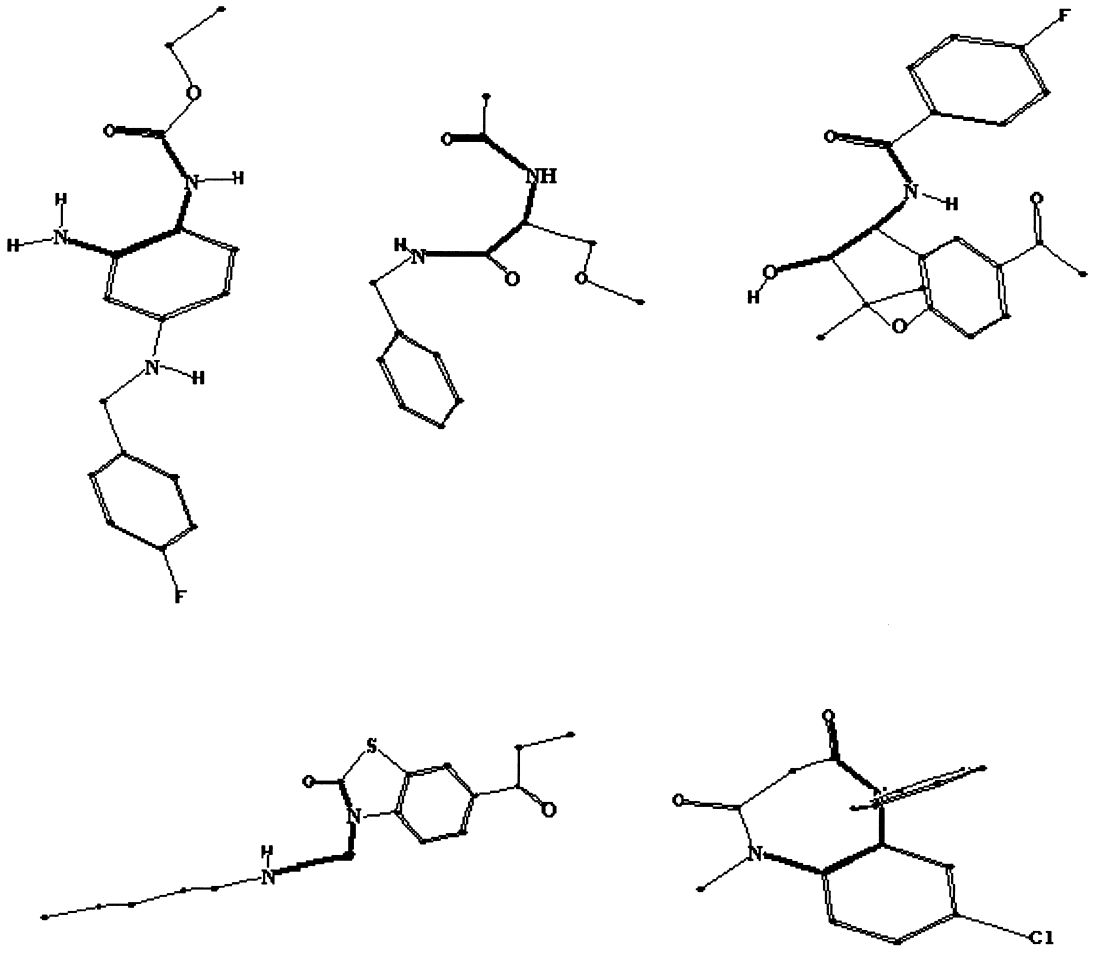


Figure 7. Atom-by-atom superposition of the fragment F38 in five anticonvulsant compounds.

Table 9. Inter-atomic distances among heteroatoms in the fragment F38 for different anticonvulsant compounds as computed from PM3 full geometry optimization

Compound[a]	=O···N-[b]	-N···N-[c]
D-23129	3.244	2.924
64 (Table 6)	3.479	3.275
SB-204269	3.930	3.643[d]
88 (Table 6)	4.012	3.100
Clobazam	3.925	2.945
Average	3.718	3.177
sd[e]	0.338	0.296

[a]For structures see Figure 6.
[b]Distance between the oxygen atom of the carbonyl and the terminal nitrogen atom of fragment F38 (see Fig. 6).
[c]Distance between both nitrogens in F38.
[d]Distance -N···O-.
[e]Standard deviation.

Conclusions

The strategies for virtual screening have the objective of selecting a diverse subset of compounds from a much larger population by using an in silico method. In this virtual method, the structures of many compounds can be stored in the form of molecular descriptors. Mathematical models that discriminate between active and inactive compounds can be generated by using the quantitative structure–activity relationships, permitting the identification of novel subsets of active compounds. This selection procedure manages to avoid the time and resources expenses of either synthesizing or screening inactive compounds. In this sense, the virtual screening should be considered as a starting point in the drug development process more than as a final stage. The active compounds selected in this first stage should be further optimized by using the traditional or even new techniques of molecular modeling, 3-D QSAR, and so forth.

In virtual screening studies as well as in molecular diversity analysis of large data set of compounds, many of the traditional QSAR descriptors are not applicable as they apply to congeneric series. In this sense, the global 2-D descriptors offer an important alternative for the molecular description in such kinds of studies. They are not only easy-to-compute descriptors but also contain important structural information that is useful for discriminating pharmacologically active compounds from those being inactive. However, in spite of maintaining the principle of maximal simplicity that these descriptors offer, it is desirable that they bring the greater structural information for the next steps of the drug development process. Consequently, those descriptors bringing information on the structural fragments or

groups responsible for the activity/inactivity of the compounds selected are preferable to those only permitting the correct classification of compounds as active/inactive. In the first group of descriptors, we can include the 2-D fingerprint and the present ones, while the second group is composed of the most traditional topological indices as well as many other local or global QSAR descriptors.

As we have shown in the present work, the TOSS-MODE approach is not only able to discriminate between active and inactive anticonvulsant compounds, but also to select active compounds from a pool of novel structures and to identify the active/inactive fragments responsible for the presence or not of the pharmacological activity. This identification of active/inactive fragments has been recognized as one of the most important features of the present approach for the virtual screening and rational design of novel compounds. We have proved here that these 2-D active fragments can be related in many cases with 3-D pharmacophores previously identified in the literature. By this means the present approach can be useful in identifying 2-D pharmacophores that can be further studied by 3-D molecular modeling techniques.

The present work reaffirms our previous findings that the TOSS-MODE approach is a useful alternative to the virtual screening/rational design of novel pharmacologically active compounds. Here we have developed some useful tools for the discovery of novel anticonvulsant compounds from large databases of chemical structures.

The current models were applied for the identification of active molecules in a database of 684 synthesized compounds. Some of the compounds so identified were used for the design of novel anticonvulsant compounds that were further synthesized and characterized. These compounds are now under study for the pharmacological screening as anticonvulsants and the results will be published elsewhere.

Acknowledgements

E. E. thanks the Ministerio de Educación y Cultura, Spain for a fellowship as Foreign Young Researcher in Spain.

References and Notes

1. Hann, M.; Green, R. *Curr. Opin. Chem. Biol.* **1999**, *3*, 379.
2. McNamara, J. O. In *Goodman & Gilman's The Pharmacological Basis of Therapeutics*, 9th ed. Hardman, J. G.; Limbird, L. E.; Molinoff, P. B.; Ruddon, R. W., Eds.; McGraw-Hill: New York, 1996; p 461.
3. Shorron, S. D. *Epilepsia* **1996**, *37* (Suppl. 2), 1.
4. Perucca, E. *Br. J. Clin. Pharmacol.* **1996**, *42*, 531.
5. Mattson, R. H. *Epilepsia* **1996**, *37* (Suppl. 6), S1.
6. Patsalos, P. N. *Emerging Drugs* **1999**, *4*, 87.
7. Krall, R. L.; Penry, J. K.; White, B. G.; Kupferberg, H. J.; Swinyard, E. A. *Epilepsia* **1978**, *19*, 393.
8. Eadie, M. J. *Drugs* **1984**, *27*, 328.
9. Leppik, I. E. *Epilepsia* **1994**, *35*, S29.
10. Brodie, J. M. *Lancet* **1992**, *339*, 1397.
11. Wagner, M. L. *Am. J Hosp. Pharm.* **1994**, *51*, 1657.
12. Brodie, M. J.; Richens, A.; Yuen, A. W. *Lancet* **1995**, *345*, 476.
13. Walker, M. C.; Sander, J. W. *Expert Opin. Invest. Drugs* **1999**, *8*, 1497.
14. Cosford, N. D. P.; McDonald, I. A.; Schweiger, E. J. *Annu. Rep. Med. Chem.* **1998**, *33*, 61.
15. Patsalos, P. N. *Ann. Clin. Biochem.* **1999**, *36*, 10.
16. Bialer, M.; Johannsen, S. I.; Kupferberg, H. J.; Levy, R. H.; Loiseau, P.; Perucca, E. *Epilepsy Res.* **1999**, *34*, 1.
17. Elger, C. E.; Bauer, J. *Neuropsychobiology* **1998**, *38*, 145.
18. White, H. S. *Epilepsia* **1999**, *40* (Suppl.), S2.
19. Rogawski, M. A.; Porter, R. J. *Pharmacol. Rev.* **1990**, *42*, 223.
20. Bialer, M.; Johannessen, S. I.; Kupferberg, H. J.; Levy, R. H.; Loiseau, P.; Perucca, E. *Epilepsy Res.* **1996**, *25*, 299.
21. Su, T.-Z.; Lunney, E.; Campbell, G.; Oxender, D. L. *J. Neurochem.* **1995**, *64*, 2125.
22. Loscher, W.; Schmidt, D. *Epilepsy Res.* **1994**, *17*, 95.
23. Somberg, J. C. In *Drug Development Process: Increasing Efficiency and Cost-effectiveness*; Welling, P. G.; Lasagna, L.; Banakar, U. V., Eds.; Marcel Dekker: New York, 1996; pp 1–37.
24. Lunney, E. A. *Med. Chem. Res.* **1998**, *8*, 352.
25. Gordon, E. M.; Kerwin, J. F. Jr. *Combinatorial Chemistry and Molecular Diversity in Drug Discovery*. Wiley-Liss: New York, 1998.
26. Kenny, B. A.; Bushfield, M.; Parry-Smith, D. J.; Fogarty, S.; Treherme, J. M. *Prog. Drug Res.* **1998**, *51*, 245.
27. Estrada, E. *SAR QSAR Environ. Res.* **2000**, *11*, 55. This reference also includes a description of the TOSS-MODE software.
28. Estrada, E.; Uriarte, E.; Montero, A.; Teijeira, M.; Santana, L.; De Clercq, E. *J. Med. Chem.* **2000**, *43*, 1975.
29. Estrada, E.; Peña, A.; Garcia-Domenech, R. *J. Comput.-Aided Mol. Des.* **1998**, *12*, 583.
30. Estrada, E. *J. Chem. Inf Comput. Sci.* **1996**, *36*, 844.
31. Estrada, E. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 320.
32. Estrada, E. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 23.
33. Estrada, E. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 31.
34. Estrada, E. In *Topological Indices and Related Descriptors in QSPR and QSAR;* Devillers, J.; Balaban, A. T., Eds.; Gordon and Breach: Amsterdam, 1999; p 505.
35. Negwer, M. *Organic-Chemical Drugs and their Synonyms*. Akademie-Verlag: Berlin, 1987.
36. Estrada, E.; Gutierrez, Y. *J. Chromatogr. A* **1999**, *858*, 187.
37. Kornet, M. J.; Chu, J. Y.-R. *J. Pharm. Sci.* **1983**, *72*, 1213.
38. Pajouhesh, H.; Parson, R.; Dopp, F. D. *J. Pharm. Sci.* **1983**, *72*, 318.
39. Popp, F. D.; Pajouhesh, H. *J. Pharm. Sci.* **1982**, *71*, 1052.
40. Haj-Yehia, A.; Bialer, M. *J. Pharm. Sci.* **1990**, *79*, 719.
41. Hernandez-Gallegos, Z.; Lebmann, P. A. *J. Pharm. Sci.* **1990**, *79*, 1032.
42. Crider, A. M.; Kolczynski, T. M.; Miskell, D. L. *J. Pharm. Sci.* **1981**, *70*, 192.
43. Darling, C. M.; Pryor, P. *J. Pharm. Sci.* **1979**, *68*, 108.
44. Kelley, J. L.; Linn, J. A.; Bankston, D. D.; Burchall, C. J.; Soroko, F. E.; Cooper, B. R. *J. Med. Chem.* **1995**, *38*, 3676.
45. Kelley, J. L.; Davis, R. G.; McLean, E. W.; Glen, R. C.; Soroko, F. E.; Cooper, B. R. *J. Med. Chem.* **1995**, *38*, 3884.
46. Kelley, J. L.; Koble, C. S.; Davis, R. G.; McLean, E. W.; Soroko, F. E.; Cooper, B. R. *J. Med. Chem.* **1995**, *38*, 4131.
47. Dimmock, J. R.; Puthucode, R. N.; Smith, J. M.; Hetherington, M.; Quail, J. W.; Pugazhenthi, U.; Lechier, T.; Stables, J. P. *J. Med. Chem.* **1996**, *39*, 3984.
48. Puthucode, R. N.; Pugazhenthi, U.; Quail, J. W.; Stables, J. P.; Dimmock, J. R. *Eur. J. Med. Chem.* **1998**, *33*, 595.

49. Unverferth, K.; Engel, J.; Hofgen, N.; Rostock, A.; Günter, R.; Lankau, H.-J.; Menzer, M.; Rolfs, A.; Liebscher, J.; Müller, B.; Hofmannn, H.-J. *J. Med. Chem.* **1998**, *41*, 63.
50. Basak, S. C.; Grunwald, G. D.; Niemi, G. J. In *From Chemical Topology to Three-Dimensional Geometry;* Balaban, A. T., Ed.; Plenum: New York, 1997; p 73.
51. Walters, W. P.; Stahl, M. T.; Murcko, M. A. *Drug Disc. Today* **1998**, *3*, 160.
52. Venkatesh, S.; Lipper, R. *J. Pharm. Sci.* **2000**, *89*, 145.
53. Choi, D.; Stables, J. P.; Kohn, H. *J. Med. Chem.* **1996**, *39*, 1907.
54. Carson, J. R.; Carmosin, R. J.; Pitis, P. M.; Vaught, J. L.; Almond, H. R.; Stables, J. P.; Wolf, H. H.; Swinyard, E. A.; White, H. S. *J. Med. Chem.* **1997**, *40*, 1578.
55. Ho, B.; Venkataranghan, P. M.; Cruce, S. F.; Hinko, C. N.; Andersen, P. H.; Crider, A. M.; Adloo, A. A.; Roane, D. S.; Stables, J. P. *Eur. J. Med. Chem.* **1998**, *33*, 23.
56. Vamecq, J.; Lambert, D.; Puopaert, J. H.; Masereel, B.; Stables, J. P. *J. Med Chem.* **1998**, *41*, 3307.
57. Ucar, H.; van Derpoorten, K.; Cacciaguerra, S.; Spampinato, S.; Stables, J. P.; Depovere, P.; Isa, M.; Masereel, B.; Delarge, J.; Poupaert, J. H. *J. Med. Chem.* **1998**, *41*, 1138.
58. Pavarello, P.; Bonsignori, A.; Doster, P.; Heidempergher, F.; Pinciriui, V.; Colombo, M.; McArthur, R. A.; Salvati, P.; Post, C.; Fariello, R. G.; Varasi, M. *J. Med. Chem.* **1998**, *41*, 579.
59. Chan, W. N.; Hadley, M. S.; Harling, J. D.; Herdon, H. J.; Jerman, J. C.; Orlek, B. S.; Stean, T. O.; Thompson, M.; Upton, N.; Ward, R. W. *Bioorg. Med. Chem. Lett.* **1998**, *8*, 2903.
60. Van Dorsser, W.; Barns, D.; Cordi, A.; Roba, J. *Arch. Int. Pharmacodyn.* **1983**, *266*, 239.
61. Janssens de Varebeke, P.; Cavalier, R.; David-Remade, M.; Youdim, M. B. H. *J. Neurochem.* **1988**, *50*, 1011.
62. Janssens de Varebeke, P.; Niebes, P.; Pauwels, G.; Roba, J.; Korf, J. *Biochem. Pharmacol.* **1983**, *32*, 2751.
63. O'Brien, E. M.; Dostert, P.; Pevarello, P.; Tipton, K. F. *Biochem. Pharmacol.* **1994**, *48*, 905.
64. Cheung, W. K.; Kianifard, F.; Wong, A.; Mathieu, J.; Cook, T.; John, V.; Redalieu, E.; Chan, K. *Pharm. Res.* **1995**, *12*, 1878.
65. Gower, A. J.; Noyer, M.; Verloes, R.; Gobert, J.; Wülfert, E. *Eur. J. Pharmacol.* **1992**, *222*, 193.
66. Seydel, J. K.; Schaper, K.-J.; Coats, E. A.; Cordes, H. P.; Eming, P.; Engel, J.; Kutscher, B.; Polymeropoulos, E. E. *J. Med. Chem.* **1994,** *37*, 3016, and references cited therein.
67. Tober, C.; Rundfeldt, C.; Rostock, A.; Bartsch, R.; Nickel, B.; Szeleny, I.; Herbst, M.; Dieter, R.; Olbrich, A.; Kutscher, B.; Sauer, W.; Jainta, H.; Jahn, W.; Engel, J. *Drugs of the Future* **1995**, *20*, 1112.
68. Rostock, A.; Tober, C.; Rundfeldt, C.; Bartschm, R.; Engel, J.; Polymeropoulos, E.; Kutscher, B.; Loscher, W.; Honack, D.; White, H. S.; Wolf, H. H. *Epilepsy Res.* **1996**, *23*, 211.
69. Tober, C.; Rostock, A.; Rundfeldt, C.; Bartsch, R. *Eur. J. Pharmacol.* **1996**, *303*, 163.
70. von Bebenburg, W.; Thiele, K.; Engel, J.; Sheldnick, W. S. *Chem.-Ztg.* **1981**, *105*, 217.
71. Camerman, A.; Camerman, N. *Antiepileptic Drugs*. Raven: New York, 1982, p. 87.
72. Stewart, J. J. P. *J. Comput. Chem.* **1989**, *10*, 209.
73. MOPAC 6.0; Frank J. Seiler Research Lab.: US Air Force Academy, CO, 1990.
74. Jimonet, P.; Audiau, F.; Barreau, M.; Blanchard, J.-C.; Boireau, A.; Bour, Y.; Coléno, M. A.; Doble, A.; Doerflinger, G.; Do Huu, C.; Donat, M.-H.; Duchesne, J. M.; Ganil, P.; Guérémy, C.; Honoré, E.; Just, B.; Kerphirique, R.; Gontier, S.; Hubert, P.; Laduron, P. M.; Le Blevec, J.; Meunier, M.; Miquet; J.-M.; Nemecek, C.; Pasquet, M.; Piot, O.; Pratt, J.; Rataud, J.; Reibau, M.; Stutzmann, J.-M.; Mignani, S. *J. Med. Chem.* **1999**, *42*, 2828.