On Application of Carhart Atom Pairs to Predict Anticonvulsant Activity

[1,2]Alan Talevi, [1]Julián J. Prieto, [1]Luis E. Bruno-Blanch and [2]Eduardo Castro
[1]Medicinal Chemistry, Department of Biological Sciences, Exact Sciences Collage, La Plata National
University (UNLP), B1900AVV, La Plata, Buenos Aires, Argentina
[2]Instituto de Fisicoquímica Teórica y Aplicada (INIFTA), Department of Chemistry
Exact Sciences College La Plata National University (UNLP) Suc.4, C.C. 16, B1900AVV
La Plata, Buenos Aires, Argentina

Abstract: About 50 million people in the world suffer from epilepsy, especially in childhood, adolescence and old age. Available treatment fails to control epilepsy in up to 30% of affected people. In developing countries, however, the amount of patients that do not receive adequate treatment climbs up to 75%. Moreover, the new generation of antiepileptic drugs (AEDs) causes important central and peripheral side effects, including ataxia, diplopia, dizziness, headache, nausea, allergies and sedation. A mathematical model previously developed by Bagchi and Maiti, involving Carhart atom pairs and similarity measures, is applied in the prediction of anticonvulsant activity of two sets of compounds which have shown to be active in the Maximal Electroshock Seizure (MES) test, meaning that their mechanism of action can be at least partially explained through sodium channels blockade. Nine structurally heterogeneous molecules define the first set of compounds, with Carhart similarities to carbamazepine ranging from 0.005 to 0.593. The second set is defined by four benzodiazepines derivatives with Carhart similarities to THIQ-10c ranging from 0.533 to 0.570. A new, more specific, model is constructed based on the one from Bagchi and Maiti and a pharmacophore previously identified in our laboratory through an active analog approach. Applied to both sets of compounds, our model shows smaller average percentage error and average absolute error in the prediction than the one form Bagchi and Maiti and smaller SD as well. Accuracy and precision in the prediction also increases compared to those obtained when using bare similarity coefficients as relative activity indicators.

Key words: MES test, antiepileptic, anticonvulsant, molecular topology, atom pairs

INTRODUCTION

Despite the continuous development of 3D and 4D QSAR methodologies, 2D-based descriptors still remain among the most widely used descriptors in pharmaceutical research today[1-5]. Intuitively one would think that since 3D descriptors capture more features of the molecular structure than 2D descriptors and therefore, they would probe themselves more successful at activity selection and prediction. However, many examples in literature show that this is not always the actual case[6-8]. Particularly interesting might be the approach of Schuffenhauer *et al.* who have suggested that the two types of descriptors are complementary in nature[8]. Although their physicochemical meaning is not always clear, topological descriptors have some advantages that should be taken into account; among these: they have a low computational cost and they can be easily calculated for all the existing, new and in-development chemical structures[9,6].

There are, essentially, two major types of molecular descriptors[10]. The first of them has been defined as "holistic". That is to say, this type of descriptor could be associated to an approach that does not consider a system as the mere sum of its parts. On the contrary, it assumes the individual parts interacting all together in the system they belong-to generate unique properties that may be unpredictable from a partial, fragment-based approach. Descriptors of this class are numbers that usually represent some important physical property of the whole molecule. These descriptors are either measured or calculated through algorithms. Examples that could be mentioned among many others are: octanol-water partition coefficient[11], Kier Shape Index[12], the shape indices of Hopfinger[13] and the Index of Electrotopological state[14].

The other descriptor category consists of the so called *chemical substructures*[1,10]. These are pieces of information strung together that may include atomic species, bond types and connectivity of two or more atoms. They could be seen as subgraphs of the graph that is associated to a molecular structure, labeled or colored in a way that reflects the information mentioned above. Some examples that fall in this category are: Atom Pairs (AP)[15,16,1], Topological Torsions (TT)[1,10]

Corresponding Author: Eduardo Castro: Instituto de Fisicoquímica Teórica y Aplicada (INIFTA), Department of Chemistry, Exact Sciences College La Plata National University (UNLP) Suc.4, C.C. 16, B1900AVV, La Plata, Buenos Aires, Argentin

and Atom Sequences (AS)[16]. Such descriptors are particularly useful in retrieval systems, data analysis and compound classification[17,18]. Our present work focuses on this latter category of descriptors.

Atom pair definition: In 1985 Carhart *et al.*[15] defined an atom pair as a substructure composed of two non-hydrogen atoms and an interatomic separation.

(atom 1 description)–(separation)–(atom 2 description)

Separation means how apart they are, measured as the number of atoms in the shortest bond-by-bond path that contains atoms 1 and 2. Description of each atom tells its chemical type, the number of non-hydrogen atom attached to it and the number of bonding π electrons it bears. Only the more common types of atoms are represented explicitly: C, O, N, S, F, Cl, Br, I, P, Si and B. Any other atom type is represented through the symbol "Y". An asterisk or dot following an atom symbol represents the presence of a bonding π electron. Xn following an atom symbol indicates the presence of n non – hydrogen neighboring atoms. For any given chemical structure denoted by *j*, the number of AP derived form the structure is calculated by:

$$d(j) = \frac{n(n-1)}{2} \qquad (1)$$

n being the number of non-hydrogen atoms present in the compound *j*.

Carhart *et al.* defined an Euclidean-distance-based similarity measure between generic structures *s* and *t* as:

$$S(s,t) = 1 - \frac{D(s,t)}{d(s)+d(t)}$$

$$= 1 - \frac{\sum_{\substack{different \\ types-of-pairs}} ABS\left[n(i,s)-n(i,t)\right]}{d(s)+d(t)} \qquad (2)$$

where $n(i,j)$ is the number of AP of type *i* derived from structure *j*. $D(s,t)$ is referred in literature both as the Hamming distance or the sem-theoretic form of Euclidean distance[16,18]. $S(s,t)$ can take values between 0 (completely non – similar structures) and 1 (identical structures). Latter, Carhart *et al.* applied this concept to structure activity relationships through similarity probe and trend vector analysis[10,15].

The Bagchi and Maiti model: In 2003, Bagchi and Maiti[19] used Carhart AP and Carhart similarity in the development of a mathematical model that was successfully applied to the prediction of the ordination of activities -relative to a reference drug- of antituberculosis drugs. This model can be described as follows:

$$A = S\left[1 - \sum_{i=1}^{m} \frac{(\lambda i)}{\sum \lambda i} \cdot \frac{\Delta ni}{ni}\right] \qquad (3)$$

A denotes activity relative to the standard drug; *S* denotes similarity between the standard drug and its analog (as defined by Carhart *et al.*). The factor between brackets takes into account, only, heteroatom APs common to the standard drug and its analog; the influence of all APs is expressed only through the similarity factor. Bagchi and Maiti gave more importance to the APs containing heteroatoms on the basis that heterocyclic compounds are usually biologically active compared to carbocyclic compounds and that heterocyclic compounds are composed of heteroatom APs. They also assume that heteroatom APs with the highest path lengths have greater influence over activity; as they cover a major portion of the molecule, coincidence regarding this type of APs implies greater probability of similar activity. Following this idea, λ_i denotes the path length of the ith heteroatom AP, Δn_i denotes the difference in quantity of ith heteroatom AP between the standard drug and the analog that is being tested (note that if the ith heteroatom AP appears more times in the tested structure, Δn_i becomes negative and the second term between brackets turns positive, adding to the activity). n_i denotes the number of ith heteroatom APs present in the standard drug. The factor between brackets could be barely related to the pharmacophore concept (see below: *Pharmacophore definition*).

Bagchi and Maiti validated their model not only with a set of known antituberculosis drugs, but with a little set of diuretic drugs as well. Both set of compounds were formed by structurally similar drugs. On this basis we decided to verify is Bagchi and Maiti model could be applied to predict the activities (or at least, the ordination of activities relative to a lead compound) of anticonvulsant drugs with mechanisms of action involving sodium channels blockade. Our objectives can be summarized as follows:

* Verifying if Bagchi and Maiti model can be applied to the prediction of anticonvulsant activity mediated by sodium channel blockade.
* Corroborating, through a structurally homogeneous set of compounds, the usefulness of Bagchi and Maiti model as an optimization tool.
* Testing, through a structurally heterogeneous set of compounds, if Bagchi and Maiti model could be used in the search for new lead molecules.

Motivation: Epilepsy is one of the most common neurological disorders that affects human condition. The word epilepsy describes disorders characterized by recurrent seizure attacks that range form a brief lapse on attention to sever, frequent convulsions due to synchronous neuronal firing. About 50 million people in the world suffer from epilepsy, especially in

childhood, adolescence and old age[20,21]. Available treatment fails to control epilepsy in up to 30% of affected people. However, in developing countries about three-fourths of patients don´t receive the treatment they need[21]. Moreover, even the new generation of antiepileptic drugs (AEDs) causes important side effects, including ataxia, diplopia, dizziness, headache, nausea, sedation, allergies, blood dyscrasias and hepatotoxicity[22]. Thus, new AEDs with better safety, less toxicity and higher efficacy in difficult- to-control patients are urgently needed.

All over the world the social consequences of epilepsy are often more difficult to overcome that seizures themselves. Fear, misunderstanding and even superstition are just some of the issues faced by epilepsy patients. In developing countries like Cameroon, India and Indonesia, at present, epilepsy is considered as related to evil, demoniac forces. At present, in both China and India, common opinion is that epilepsy is a valid reason for annulling marriage[23].

Pharmacophore definition: Medicinal Chemistry usually faces the problem of how to generate libraries of promissory compounds starting from fragmentary but relevant information. The word pharmacophore, derived from Greek, means "medicine carrier". It refers to a substructure that embodies essential elements that must be present in a compound to elicit a specific physiological response, that is to say, to have a desired activity. Typically, a pharmacophore consists of a list of atom types that have been determined to be significant to the activity, along with the geometrical relation between them (including inter-atomic distances). Significance to the activity includes chemical functions such as charge, hydrogen bonding propensity, hydrophobicity or aromaticity[24,25].

At present and in the past, our laboratory has identified pharmacophores related to anticonvulsant activity[26,27] through fitting in SYBYL 6.6[28] on Silicon Graphics Octane. A model of the identified pharmacophore can be seen in Fig. 1.

In the present work we employed information from this identified pharmacophores to generate a new mathematical model on the basis of the one from Bagchi and Maiti (which already incorporated, indirectly, the pharmachophore notion).

On similarity determination: In general, there are two classes of similarity coefficients: association and distance coefficients. The essential difference between them is that the latter considers the common absence of certain structural features as the evidence of similarity between two chemical structures, whereas the former does not. Examples of association coefficients that can be mentioned are Tanimoto, Dice and Cosine
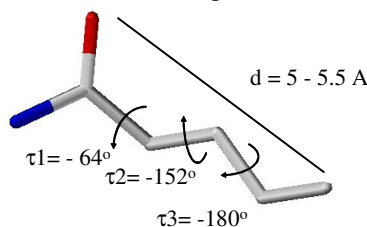
coefficients. Examples of distance coefficients are

Fig. 1: Identified pharmacophore for anticonvulsant drugs with sodium channel blocking mechanism

Hamming and Euclidean distances[16,18]. Carhart *et al*. and therefore, Bagchi and Maiti -as can be seen in expression (3)- used for similarity determination the radio between the Hamming distance and the total number of features (APs) derived from the two chemical structures being compared.

However, work from Chen *et al*.[16] demonstrates that Tanimoto coefficient showed better performance than Euclidean distance in 2D fragment-based similarity searching, whereas James *et al*.[19] have suggested that Hamming and Euclidean distances are useful in "relative" distance comparisons (i.e., the distance of two molecules to the same target) but not for "absolute" comparisons (between two independent pairs of molecules). This means Hamming distance might not be the best choice for the similarity factor if one wants to use Bagchi and Maiti model as a tool in the search of new leads, since in that case one would use the model to compare, in an independent manner, each of the tested analogs with the drug chosen as reference. Hamming distance, then, would be the most appropriate choice in optimization (which is the original use proposed by Bagchi and Maiti), since for this purpose obtaining an ordination of relatives activities between the possible analogs would be good enough.

At the moment of constructing our model on the basis of the Bagchi and Maiti one, we evaluated how the performance of our model varies when utilizing three different similarity coefficients: the one proposed by Carhart *et al*.; the binary form of the Tanimoto coefficient (in which features are confined to dichotomous values: 0 indicating the absence of a particular type of atom pair, 1 indicating its presence) and the algebraic form of the Tanimoto coefficient. The descriptions of these coefficients are given, respectively, in expressions (2), (4) and (5).

$$S_{A,B} = \frac{c}{a+b-c} \qquad (4)$$

a is the number of unique fragments in compound A; b, the number of unique fragments in compound B; c, the number of unique fragments shared by compounds A and B.

$$S_{A,B} = \frac{\sum_{i=1}^{m} n_{A,i} n_{B,i}}{\sum_{i=1}^{m} n_{A,i}^2 + \sum_{i=1}^{m} n_{B,i}^2 - \sum_{i=1}^{m} n_{A,i} n_{B,i}} \qquad (5)$$

where $n_{A,i}$ is the number of fragment i in compound A; $n_{B,i}$ is the number of fragment i in the compound B. (5)

MATERIALS AND METHODS

Construction of the Set of Compounds: Two sets of compounds which have shown activity in the MES test were constructed.

Set A (Fig. 2) is defined by nine anticonvulsant drugs previously used in our laboratory to identify a pharmacophore which seems to be responsible of the anticonvulsant activity through sodium channels blockade[26]. Carbamazepine was chosen as the reference drug of this set, provided that its measured activity at the MES test was the greatest among the group ($ED_{50} = 37$ μmol kg^{-1} ip). Set A can be referred as a heterogeneous set (and therefore, it can be utilized to test the efficacy of Bagchi and Maiti model as a predictive tool not only in the optimization of a compound also in the rational search of new lead compounds). Calculated ranges of Carhart Similarity Coefficient, the binary form of the Tanimoto Coefficient and the algebraic form of the Tanimoto Coefficient for set A were, in that order, 0.005 to 0.760, 0.009 to 0.586 and 0.001 to 0.804.

Set B (Fig. 2) is defined by four benzodiazepine derivatives tested in different experimental seizure models by De Sarro *et al.*[29,30]. THIQ – 10c was chosen as reference drug, with a measured activity of 5.17 μmol kg^{-1} in the MES test. The similarity coefficients determined by (2), (4) and (5) were calculated for compounds of this set, obtaining, in that order, similarity ranges of 0.51 to 0.54, 0.29 to 0.34 and 0.57 – 0.60. Therefore, set B can be referred as a homogeneous set and can be used to verify the efficacy of the model in the optimization of compounds of known activity.

Construction of the new model: It can be seen in the Bagchi and Maiti model (3) that the second term between brackets

$$-\sum_{i=1}^{m} \frac{(\lambda i)}{\sum \lambda i} \cdot \frac{\Delta ni}{ni} \qquad (6)$$

defines if the relative activity estimated by the model for each structure is greater or smaller than the Carhart
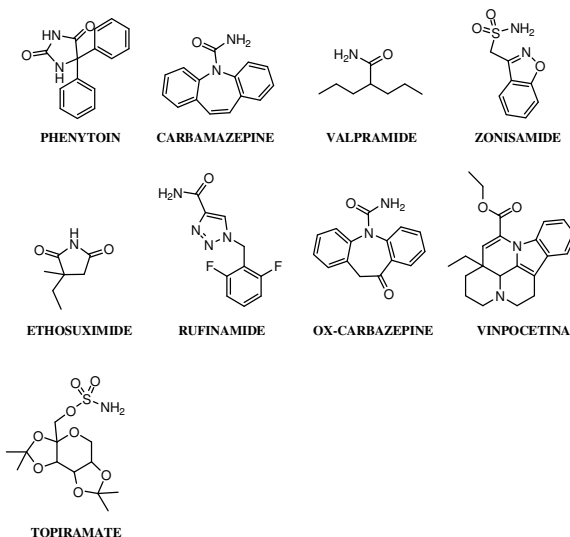

Fig. 2: Schematic representation of compounds that define set A. Structural diversity can be appreciated, as confirmed by the wide range of similarity coefficients

similarity calculated from the AP. Each one of the terms composing the summatory can be interpreted as the contribution of the ith type of AP to the activity of the whole structure. The sign of each of the this terms is given by the Δni factor, which is the only one among all the factors involved in the model that can take values either greater or smaller than zero. Since the reference drug of each set is chosen as the most active structure in it, *the types of APs present in the reference drugs are thought as desired features* involved in the interaction of the molecule and its site of action. *This may as well be the fundamental hypothesis of our model.* Therefore, if the number of APs of the ith type is greater in the reference drug than in the tested compound, then the tested molecule is lacking a desired feature and the fraction of the activity that can be explained through that particular element will be smaller in the tested compound than in the reference one. In the other hand, if the number of ith type elements is greater in the tested compound than in the reference drug, then the tested compound has more of a desired characteristic and a greater probability of this feature being expressed at the time of interacting with the site of action. On this basis we believe the presence and the place that Bagchi and Maiti have given to Δni in their model is accurate.

However; have all the types of atom pairs equal importance in determining the activity of a particular structure? Introducing λ_i, the length of the ith element, in the numerator of (6) Bagchi and Maiti gave more importance to long atom pairs, while ni, being in the denominator gave low importance to those types of AP that have a high frequence of occurrence in the reference drug.

Table 1:   AP sets derived from carbamazepine and THIQ-10c structures

Carbamazepine		THIQ-10c			
Count	Atom pair	Count	Atom pair	Count	Atom pair
7	C*X2-2-C*X2	2	C*X2-2-C*X2	2	CX1-5-C*X2
2	C*X3-2-C*X3	8	C*X2-2-C*X3	4	CX1-5-C*X3
6	C*X2-2-C*X3	2	C*X2-3-BrX1	1	CX1-5-CX2
1	C*X3-2-NX1	2	C*X2-3-C*X2	2	CX1-5-OX2
3	C*X3-2-NX3	8	C*X2-3-C*X3	3	CX1-6-C*X2
1	C*X3-2-O*X1	2	C*X2-3-OX2	3	CX1-6-C*X3
6	C*X2-3-C*X2	2	C*X2-4-BrX1	1	CX1-6-CX1
11	C*X2-3-C*X3	3	C*X2-4-C*X2	1	CX1-6-CX2
2	C*X2-3-NX3	3	C*X2-4-C*X3	1	CX1-6-CX3
8	C*X2-4-C*X2	3	C*X2-4-NX3	3	CX1-7-C*X2
10	C*X2-4-C*X3	2	C*X2-4-OX2	1	CX1-7-C*X3
6	C*X2-4-NX3	8	C*X2-5-C*X2	2	CX1-7-CX2
7	C*X2-5-C*X2	3	C*X2-5-C*X3	1	CX1-7-CX3
14	C*X2-5-C*X3	3	C*X2-5-NX3	1	CX1-7-NX3
2	C*X2-5-NX1	4	C*X2-6-C*X2	2	CX1-8-C*X2
2	C*X2-5-NX3	7	C*X2-6-C*X3	5	CX1-8-C*X3
2	C*X2-5-O*X1	3	C*X2-6-O*X1	1	CX1-8-CX2
7	C*X2-6-C*X2	2	C*X2-7-C*X2	1	CX1-8-NX3
8	C*X2-6-C*X3	4	C*X2-7-C*X3	1	CX1-8-OX2
6	C*X2-6-NX3	3	C*X2-7-O*X1	1	CX1-9-BrX1
6	C*X2-6-O*X1	2	C*X2-7-OX2	4	CX1-9-C*X2
7	C*X2-7-C*X2	1	C*X2-8-BrX1	1	CX1-9-C*X3
2	C*X2-7-NX1	3	C*X2-8-C*X3	1	CX1-9-CX1
2	C*X2-7-O*X1	4	C*X2-8-OX2	1	CX1-9-O*X1
3	C*X2-8-C*X2	1	C*X2-9-BrX1	1	CX1-9-OX2
4	C*X3-3-C*X3	2	C*X2-9-OX2	1	CX2-2-C*X3
2	C*X3-3-NX3	1	C*X3-10-BrX1	1	CX2-2-CX2
5	C*X3-4-C*X3	1	C*X3-10-OX2	1	CX2-2-NX3
2	C*X3-4-NX1	1	C*X3-2-BrX1	1	CX2-3-C*X2
2	C*X3-4-OX1	2	C*X3-2-C*X3	3	CX2-3-C*X3
2	C*X3-5-NX1	1	C*X3-2-NX3	1	CX2-3-CX3
2	C*X3-5-O*X1	1	C*X3-2-O*X1	1	CX2-3-NX3
1	NX1-3-O*X1	2	C*X3-2-OX2	2	CX2-4-C*X2
1	NX3-3-NX1	3	C*X3-3-C*X3	4	CX2-4-C*X3
1	NX3-3-O*X1	2	C*X3-3-NX3	1	CX2-4-CX3
7	C*X2-2-C*X2	2	C*X3-3-OX2	1	CX2-4-O*X1
2	C*X3-2-C*X3	6	C*X3-4-C*X3	3	CX2-5-C*X2
6	C*X2-2-C*X3	1	C*X3-4-NX3	3	CX2-5-C*X3
1	C*X3-2-NX1	2	C*X3-4-OX2	1	CX2-5-O*X1
3	C*X3-2-NX3	1	C*X3-5-BrX1	1	CX2-5-OX2
1	C*X3-2-O*X1	2	C*X3-5-C*X3	4	CX2-6-C*X2
6	C*X2-3-C*X2	1	C*X3-5-NX3	1	CX2-6-C*X3
11	C*X2-3-C*X3	2	C*X3-5-O*X1	2	CX2-6-OX2
2	C*X2-3-NX3	2	C*X3-5-OX2	2	CX2-7-C*X2
8	C*X2-4-C*X2	3	C*X3-6-C*X3	1	CX2-7-C*X3
10	C*X2-4-C*X3	2	C*X3-6-NX3	1	CX2-7-OX2
6	C*X2-4-NX3	1	C*X3-6-O*X1	1	CX2-8-BrX1
7	C*X2-5-C*X2	1	C*X3-6-OX2	1	CX2-9-BrX1
14	C*X2-5-C*X3	1	C*X3-7-BrX1	2	CX3-2-C*X3
2	C*X2-5-NX1	1	C*X3-7-C*X2	1	CX3-2-NX3
2	C*X2-5-NX3	2	C*X3-7-C*X3	3	CX3-3-C*X2
2	C*X2-5-O*X1	1	C*X3-7-O*X1	2	CX3-3-C*X3
7	C*X2-6-C*X2	2	C*X3-7-OX2	3	CX3-4-C*X2
8	C*X2-6-C*X3	2	C*X3-8-BrX1	1	CX3-4-C*X3
6	C*X2-6-NX3	1	C*X3-8-C*X3	1	CX3-4-O*X1
6	C*X2-6-O*X1	2	C*X3-8-O*X1	2	CX3-5-C*X3
7	C*X2-7-C*X2	1	C*X3-8-OX2	1	CX3-5-OX2
2	C*X2-7-NX1	1	C*X3-9-BrX1	1	CX3-6-BrX1
2	C*X2-7-O*X1	1	C*X3-9-C*X3	1	CX3-6-OX2
3	C*X2-8-C*X2	1	C*X3-9-OX2	1	CX3-7-C*X3
4	C*X3-3-C*X3	2	CX1-10-C*X2	1	CX3-8-C*X3
2	C*X3-3-NX3	1	CX1-10-C*X3	1	NX3-3-O*X1
5	C*X3-4-C*X3	1	CX1-10-CX1	1	NX3-6-OX2
2	C*X3-4-NX1	1	CX1-10-O*X1	1	NX3-7-BrX1
2	C*X3-4-OX1	1	CX1-11-BrX1	1	NX3-7-OX2
2	C*X3-5-NX1	1	CX1-11-C*X3	1	O*X1-9-BrX1
2	C*X3-5-O*X1	1	CX1-12-BrX1	1	OX2-10-BrX1
1	NX1-3-O*X1	1	CX1-2-C*X3	1	OX2-11-BrX1
1	NX3-3-NX1	2	CX1-2-OX2	1	OX2-4-OX2
1	NX3-3-O*X1	2	CX1-3-C*X3	1	OX2-8-O*X1
7	C*X2-2-C*X2	1	CX1-3-NX3	1	OX2-9-O*X1
2	C*X3-2-C*X3	1	CX1-3-O*X1	1	CX1-4-CX2
6	C*X2-2-C*X3	2	CX1-4-C*X2	1	CX1-4-CX3
1	C*X3-2-NX1	2	CX1-4-C*X3		

Corresponding Author: Eduardo Castro: Instituto de Fisicoquímica Teórica y Aplicada (INIFTA), Department of Chemistry, Exact Sciences College La Plata National University (UNLP) Suc.4, C.C. 16, B1900AVV, La Plata, Buenos Aires, Argentin

Table 2: Heteroatom APs common to carbamazepine and phenytoin. Calculations derived from each type of AP for the two models that are tested

Atom pair	ount in carbamazepine	Count in phenytoin	$\dfrac{(\lambda i)}{\sum \lambda i} \cdot \dfrac{\Delta ni}{ni}$	$\dfrac{\left(\dfrac{ni \cdot \Delta n_i}{1 + ABS(\lambda i - 6)}\right)}{c}$
C*X3-2-O*X1	1	2	-0.069	-0.200
C*X2-5-O*X1	2	4	-0.172	-2.000
C*X2-6-O*X1	6	7	-0.193	-6.000
C*X2-7-O*X1	2	6	-0.482	-4.000
C*X3-4-OX1	2	4	-0.138	-1.333
C*X3-5-O*X1	2	2	0.000	0.000

Table 3: Similarity coefficients for every compound tested. Compounds composing set A were compared to carbamazepine. Compounds composing set B were compared to THIQ-10c

	Compound	Carhart similarity	Binary form of Tanimoto coefficient	Algebraic form of Tanimoto coefficient
Set A	phenytoin	0,59	0,35	0,64
	valpromide	0,03	0,05	0,002
	zonisamide	0,24	0,21	0,20
	ethosuximide	0,04	0,04	0,009
	rufinamide	0,17	0,11	0,15
	ox-carbazepine	0,76	0,59	0,80
	vinpocetine	0,23	0.08	0,30
	topiramate	0,005	0,009	0,001
Set B				
	CFM2S	0,53	0,29	0,57
	CFM11	0,57	0,35	0,58
	CFM2	0,57	0,34	0,60
	carbamazepine	0,32	0,14	0,40

Since the reference is the drug with maximum activity, we believe that the more times a particular type of AP appears in the reference structure, the more important is the contribution of that type of AP to the activity. Thus, our model includes the factor *ni* not in the denominator but in the numerator of the second term between brackets. We also believe that, being in knowledge of an identified pharmacophore, we should give more attention not to the longest types of AP but to the types of AP that accomplish the characteristics of the pharmacophore. These would be those AP that include heteroatoms and a separation of six, since six is the maximum number of atoms comprised in the possible AP derived from the identified pharmacophore. On this consideration we propose the following model:

$$A = S^b \left[1 - \frac{\displaystyle\sum_i \frac{ni \cdot \Delta n_i}{1 + ABS(\lambda i - 6)}}{c} \right] \quad (7)$$

which is an specific form of the more of the general expression:

$$A = S^b \left[1 - \frac{\displaystyle\sum_i \frac{ni \cdot \Delta n_i}{1 + ABS(\lambda i - \alpha)}}{c} \right] \quad (8)$$

that could be applied not only to anticonvulsants but also to compounds of any given activity, *provided that they interact with the same site of action*. A is the activity relative to a reference drug, which is chosen as the most active structure of the working set. S is a coefficient of similarity.

In this work we tested our model with both distance and association similarity coefficients: (2), (4) and (5). *ni* represents the number of ith AP in the reference drug; Δni stands for the difference of ith AP between the reference drug and the tested drug; λi is the number of atoms included in the ith AP and α is the number of atoms involved in the longest AP of the identified pharmacophore. *b* and *c* are constants. Depending on the value assigned to it, *b* may moderate the influence of the similarity on the estimated activity. Although *c* main purpose is to normalize the second term between brackets, keeping its value between zero and one, eventually it may as well serve to compare the relative influence of different aspects of the pharmacophore. Lets say, for instance, the identified pharmacophore -essential for the occurrence of a given

activity- has both an electrostatic and an hydrophobic component. The electrostatic component may be represented by the coincident heteroatom AP (or by certain kinds of common heteroatom AP) between the reference and the tested compounds. The hydrophobic component may require the introduction of another term in the model, which took into account the coincident non-heteroatom AP or, if aromaticity were an important feature, the non-heteroatom AP that included π electrons. In that case the magnitude taken by c in each one of the terms would be an indicator of the importance of the electrostatic component relative to the hydrophobic one. E.g., if the constant c for the electrostatic term were 100 and the constant c for the hydrophobic term were 50, then the hydrophobic contribution can be though as twice important in activity determination as the electrostatic. This idea is widely applied in QSAR models, where the normalized coefficients associated to each independent variable (or descriptor) are compared in order to quantify the relative importance of the descriptors included in the model.

Application of both models: The AP set derived of each of the thirteen compounds was constructed. Table 1 shows the complete sets of AP for carbamazepine and THIQ-10c. Once in knowledge of all the AP sets the similarity coefficients of Carhart and Tanimoto (this latter in its binary and algebraic form) were calculated as specified in expressions (2), (4) and (5); the values of these coefficients for each structure are shown in Table 3. In order to determine the common heteroatom APs, the APs of each one of the structures was compared with the APs of the correspondent reference drug. Table 2 shows the types of AP common to carbamazepine and phenytoin and the occurrence of each type in each one of these structures. It is also shown the calculated value for the negative term between brackets for the two models; c was chosen as equal to 100 in (7). Different values for b were evaluated for each set of compounds: 0.5, 0.6, 0.75 and 1. c was given value of 100 in set A and values of 50 and 100 in set B. Average percentage error and average absolute error of the predicted value of relative activity were calculated in all cases. We present the results for the model of Bagchi and Maiti and for the forms of our model that proved more accurate. We also calculated carbamazepine similarity to THIQ-10c for further analysis in set B, as explained in *Results and discussion*. Finally, bare similarity coefficients were compared to the relative activity in order to verify if the mathematical models improved the prediction that could be inferred using just the similarity coefficients as activity indicators.

RESULTS AND DISCUSSION

When applied to set B (structurally homogeneous) the Bagchi and Maiti model confirmed its usefulness as an optimization tool, since it predicted the correct ordination of relative activities in the same manner Bagchi and Maiti predicted the ordination of relative activities for diuretics and antituberculosis drugs. The actual ordination of relative activities of THIQ-10c, CFM2S, CFM11, CFM2 and carbamazepine is: THIQ-10c > CFM2S > CFM2 > CFM11 > carbamazepine, while the order predicted by Bagchi and Maiti model is THIQ-10c = CFM2S > CFM2 > CFM11 > carbamazepine. However, as can be seen in Table 4, the values of relative activity predicted are quite different from the actual values (average percentage error and average absolute error -not including carbamazepine- in set B equal to 69.9% and 0.20, in that order; when including carbamazepine this parameters rise, respectively, to 95.3% and 0.21).

Three forms of our model showed good in the prediction of *absolute relative activities* as well, as can be appreciated in Table 4. These forms of the model are:

$$A = S^{1.2}\left[1 - \frac{\sum_i \dfrac{ni \cdot \Delta n_i}{1 + ABS(\lambda i - 6)}}{50}\right] \qquad (9)$$

S being the algebraic form of the Tanimoto coefficient;

$$A = S\left[1 - \frac{\sum_i \dfrac{ni \cdot \Delta n_i}{1 + ABS(\lambda i - 6)}}{50}\right] \qquad (10)$$

S being the binary form of the Tanimoto coefficient; and

$$A = S\left[1 - \frac{\sum_i \dfrac{ni \cdot \Delta n_i}{1 + ABS(\lambda i - 6)}}{100}\right] \qquad (11)$$

S being in this case the binary form of the Tanimoto coefficient.

The average percentage errors in the prediction were, for this models and not including carbamazepine, 16.7% (9) 29.2% (10) and 31.5% (11). The average absolute error calculated were 0.09, 0.15 and 0.16, in that order. When considering carbamazepine, the average errors were, respectively, 44.7, 25.5 and 25.4%, while the average absolute errors were 0.11, 0.12 and 0.12. These values may be very good as a rough prediction, since a percentage error of 50% in a relative activity of 0.10 implies a predicted relative activity of 0.05 or 0.15,

When comparing bare similarity measures to relative activities, percentage and average errors were considerable higher, meaning the models improved the

Table 4: Predicted activities for set B (with the incorporation of carbamazepine to the set). When applying the (9) and (10) forms of our model, the average percentage error is reduced, compared with the predictions obtained from bare similarity coefficients and Bagchi and Maiti model.

Compound	Measured relative activity	Relative activity predicted by Bagchi and Maiti model	Relative activity predicted by (9, S= binary form of Tanimoto coefficient)	Relative activity predicted by (10)
CFM2S	0,67	0,61	0,30	0,29
CFM11	0,23	0,49	0,28	0,31
CFM2	0,325	0,61	0,29	0,32
Carbamazepine	0,14	0.38	0,16	0,15

Table 5: Comparison of some widely used descriptors calculated for dissimilar molecules topiramate and carbamazepine

Descriptor	Topiramate	Carbamazepine
Topological charge index of order 1	10.5	2.5
Moriguchi octanol water partition coefficient	0.146	2.695
Ghose Crippen molecular refractivty	64.046	67.737
Number of atoms acceptor of H-bonds	9	3
Mean atomic Van der Waals volume	0.560	0.680
Connectivty index chi 0	16.389	12.535
Randic connectivity index	9.917	8.771
3D Wiener index	2769.08	1168.68

Table 6: Comparison of measured activity and predicted activity by Bagchi and Maiti model and form (11) of our model

Compound	Measured relative activity	Relative activity predicted by Bagchi and Maiti model	Relative activity predicted by (11)
Phenytoin	0,97	1,21	0,83
Valpromide	0,10	0,03	0,12
Zonisamide	0,40	0,25	0,32
Rufinamide	0,52	0,17	0,35
Ox-carbazepine	0,47	0,95	0,92
Vinpocetina	0,48	0,16	0,16

Table 7: Comparison between the percentage error obtained in the prediction with our models to those obtained when applying Bagchi and Maiti or when using bare similarity coefficients as indicators of relative activity

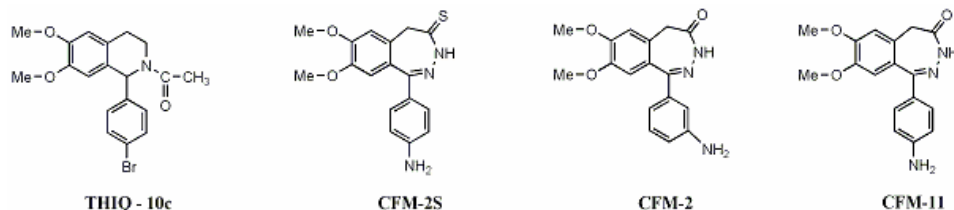| B	Bare Carhart Similarity	Bare Tanimoto algebraic form	Bare Tanimoto binary form	Bagchi and Maiti model	(9)	(10)	(11)	
CFM2S	20,4	15,0	56,7	8,95	31,3	55,0	56,7	
CFM11	147,8	152,2	52,2	113,0	17,4	21,8	34,8	
CFM2	75,4	84,6	4,4	87,7	1,6	10,8	3,1	
Carbamazepine	132,9	185,7	0	171,4	128,6	14,3	7,1	
Average percentage error	94,0	109,4	28,3	95,3	44,7	25,5	25,4	

Fig. 3: Schematic representation of compounds that define set B. It can be appreciated that the four structures are much more similar between them that the ones composing set A. This is confirmed by the narrow range of values taken by the similarity coefficients

prediction that could be made based only in similarity coefficients. The average percentage error in this case were –not including carbamazepine and utilizing, in that order, Carhart similarity coefficient, algebraic form of Tanimoto coefficient and binary form of Tanimoto coefficient- 81.2, 83.8 and 37.8%

Table 6 shows the results obtained from applying Bagchi and Maiti model to Set A. Based on their low similarities to the reference drug, we discarded ethosuximide and topiramate. Being similarity a factor of much weight in both Bagchi and Maiti and our

model, we considered the two models would fail to predict activity of molecules with high relative activity but very poor similarity, as is the case of topiramate. We believe this responds to a logical criteria: molecules with very dissimilar structures, even if they act through the same mechanism, probably interact with different receptors or at least with different zones of their molecular receptor or site of action. Not only this type of molecules have great probability of not being retrieved by a similarity based method but, if actives, they could be used to propose and alternative

interaction site or mechanism of action. Table 5 shows the quite different values for several molecular descriptors widely used in QSAR analysis, when calculated for topiramate and carbamazepine; this may be a further explanation of topiramate and ethosuximide exclusion from set A. The series of descriptors used for comparison was calculated with Dragon Academic version[31].

Obtained average percentage error and average absolute error in the prediction when applying Bagchi and Maiti model were, respectively, 61.4% and 0.27. When we applied the following form of our model:

$$A = S^{0.6} \left[1 - \frac{\sum_i \dfrac{ni \cdot \Delta n_i}{1 + ABS(\lambda i - 6)}}{100} \right] \qquad (12)$$

S being the Carhart similarity coefficient, we reduce the average percentage error to 41% and the absolute error to 0.24. Bare similarity, considering Carhart similarity as indicative of activity, predicted relative activity with a percentage error of 55.1% and an absolute error of 0.25.

CONCLUSION

We confirmed the applicability of Bagchi and Maiti model in the determination of relative activities ordination of a series of analogues referred to a lead drug. Our model seems to be adequate for the rough prediction of relative activity, therefore it may be useful as an approximation tool in the search of new leads and in drug optimization. Almost all tested forms of our model reduce the percentage error and the absolute error that are obtained when utilizing both bare similarity and Bagchi and Maiti model for the prediction. In every case, the predicted value is closer to the measured value when comparing the prediction by a particular form of our model to the bare similarity coefficient used in that particular form of our model. For instance, if one considers the bare Tanimoto binary coefficient as indicative of relative activity for the set B, predicted values have an average percentage error of 28.3%, which could be thought as quite a good rough prediction. However, if we apply either the forms (10) or (11) of our model, both including Tanimoto coefficient in its binary form as similarity coefficient, this average percentage error reduces, in that order, to 25.5 and 25.4%, meaning our model always improves the accuracy on the prediction; the standard deviation (SD) also decreases in both cases, meaning the proposed models increase the precision in the predictive capability. The same behavior appears when comparing average percentage error using bare algebraic form of the Tanimoto coefficient as indicator of relative activity (average percentage error: 109.4%) with the predicted values applying form (9) of our model, which includes

that same form of the Tanimoto coefficient as similarity coefficient (average percentage error: 44.7%) The optimal form of our model still needs to be determined, as well as the most efficient similarity coefficient to be used in it, although it appears that, as suggested by James *et al.*[19], association coefficients like Tanimoto coefficient may have a slight advantage over distance coefficients since three of the four best working forms of our model employ Tanimoto similarity coefficients and only one employs the one proposed by Carhart *et al.*

We believe optimizing the model would imply the next steps to be followed:
* testing a wider set of molecules.
* inclusion of a term that takes into account the non-heteroatom AP that may be present in the pharmacophore.
* improving the selection of types of common heteroatom AP on the basis of a better study of the identified pharmacophores or new identified pharmacophores.
* generation of a new type of AP that considers not atoms but bioisosters (meaning that the type of AP C*X3-4-O*X1 will be associated, for instance, to the AP C*X3-4-N*X2, since O*X1 and N*X2 are functional groups that work as bioisosters).

ACKNOWLEDGMENTS

REFERENCES

1. Merlot, C., D. Domine, C. Cleva and D.J. Church, 2003. Chemical substructures in drug discovery. Drug Discov. Today, 8: 594-602.

2. Merlot, C., D. Domine and D.J. Church, 2002. Fragment análisis in small molecule discovery. Curr. Opin. Drug. Discov. Devel., 5: 391-9.

3. Bajaj, S., S.S. Sambia and A.K. Madanb, 2005. Topological Models for prediction of anti-HIV activity of acylthiocarbamates. Bioorg. Med. Chem., 13: 3263–3268.

4. Duart, M.J., G.M. Antón-Fos, J.V. de Julian-Ortiz, R. Gozalbes, J. Gálvez and R. García-Domenech, 2002. Use of molecular topology for the prediction of physicochemical, pharmacokinetic and toxicological properties of a group of antihistaminic drugs. Int. J. Pharm., 246: 111-119.
5. García-García, A., J. Gálvez, J.V. de Julián-Ortiz, R. García-Domenech, C. Muñoz, R. Guna1 and R. Borrás, 2004. New agents active against Mycobacterium avium complex selected by molecular topology: a virtual screening method. J. Antimicrob. Chemotherap., 53: 65-73.
6. Sheridan, R.P. and S.K. Kearsley, 2002. Why do we need so many chemical similarity search methods?. Drug Discov. Today, 7: 903-911.
7. Matter, H. and T. Pötter, 1999. Comparing 3D pharmacophore Triplets and 2D fingerprints for selecting diverse compound subsets. J. Chem. Inf. Comput. Sci., 39: 1211-1225.
8. Schuffenhauer, A., V.J. Gillet and P. Willett, 2000. Similarity searching in files of three-dimensional chemical structures: Analysis of the BIOSTER database using two-dimensional fingerprints and molecular field descriptors. J. Chem. Inf. Comput. Sci., 40: 295-307.
9. Devillers, J. and A.T. Balaban, 1999. Topological Indices and Related Descriptors in QSAR and QSPR. Gordon and Breach Science Publishers, pp: 1-17.
10. Nilakantan, R., N. Bauman, S. Dixon and R. Venkataraghavan, 1987. Topological Torsión: A new molecular descriptor for SAR applications. Comparison with other descriptors. J. Chem. Inf. Comput. Sci., 27: 82-85.
11. Leo, A., C. Hansch and D. Elkins, 1971. Partition coefficients and their uses. Chem. Rev., 71: 525-616.
12. Kier, L., 1985. A shape index from molecular graphs. Quant. Struct.-Act. Relat. Pharmacol., Chem. Biol., 4: 109-116.
13. Walters, D.E. and A.J. Hopfinger, 1986. Case studies of the application of molecular shape analyis to elucidate drug action. J. Mol. Struct. (Theochem), 134: 317-323.
14. Kier, L.B. and L.H. Hall, 1999. Molecular Structure Description. The Electrotopological State. Academic: New York.
15. Carhart, R.E., D.H. Smith and R. Venkataraghavan, 1985. Atom pairs as molecular features in structure-activity studies: Definitions and applications. J. Chem. Inf. Comput. Sci., 25: 64-73.
16. Chen, X. and C.H. Reynolds, 2002. Performance of similarity measures in 2D fragment – based similarity searching: Comparison of structural descriptors and similarity coefficients. J. Chem. Inf. Comput. Sci., 42: 1407-1414.
17. Brown, R.D. and Y.C. Martin, 1996. Use of structure-activity data to compare structure-based clustering methods and descriptors for use in compound selection. J. Chem. Inf. Comput. Sci., 36: 572–584.
18. Wilett, P., 1998. Chemical Similarity Searching. J. Chem. Inf. Comput. Sci., 38: 983-996.
19. Bagchi. M.C. and B.C. Maiti, 2003. On application of atom pairs in drug design. J. Mol. Struct. (Theochem), 623: 31 – 37.
20. http://www.who.int/. World Health Organization Fact sheet nº 265: "Mental and neurological disorders". December 2001.
21. http:// www.who.int/. World Health Organization Fact sheet nº 165: "Epilepsy: aetiogy, epidemiology and prognosis". Revised on February 2001.
22. Bialer, M., S.I. Johannessen, H.J. Kupferberg, R.H. Levy, P. Loiseau and E. Perucca, 2002. Progress reporto on new antiepileptic drugs: A summary of the Sixth Eilat Conference (EILAT VI). Epilepsy Res., 51: 31-71.
23. http:// www.who.int/. World Health Organization Fact sheet nº 166: "Epilepsy: social consequences and economic aspects". Revised on February 2001.
24. Loew, G. H., H.O. Villar and I. Alkorta, 1993. Strategies for indirect computer aided drug design. Pharm. Res., 10: 475-486.
25. Bywater, R.P., T.A. Poulsen, P. Røgen and P.G. Hjorth, 2004. De novo generation of molecular structures using optimization to select graphs on a given lattice. J. Chem. Inf. Comput. Sci., 44: 856-861.
26. Tasso, S.M., S.Ch. Moon, L.E. Bruno–Blanch and G.L. Estiú, 2004. Characterization of the anticonvulsant profile of valpromide derivatives. Bioorg. Med. Chem., 12: 3857-3869.
27. Gavernet L.; Dominguez-Cabrera M.J.; Bruno-Blanch L. E.; Estiú "3D- QSAR Design of Novel Antiepileptic Sulfamides". Bioorg. Med Chem. In press. 2006.
28. SYBYL is a product by Tripos Inc., San Diego CA.
29. De Sarro, G., A. De Sarro, P. Gareri, R. Gitto, E. Russo, A. Chimirri and G. Ferreri, 2004. Comparative anticonvulsant activity of N-acetyl – 1 – aryl – 6,7 – dimethoxy – 1,2,3,4 – tetrahydroisoquinoline derivatives in rodents. Pharmacol., Biochem. Behav., 77: 85-94.
30. De Sarro, G., G. Ferreri, P. Gareri, E. Russo, A. De Sarro, R. Gitto and A. Chimirri, 2003. Comparative anticonvulsant activity of some 2,3 – benzodiazepine derivatives in rodents. Pharmacol., Biochem. Behav., 74: 595-602.
31. Dragon Academic version is a product by Milano Chemometrics and QSAR research group, Milano, Italy.