Journal of Computer-Aided Molecular Design, **12:** 583–595, 1998.
KLUWER/ESCOM

Designing sedative/hypnotic compounds from a novel substructural graph-theoretical approach

Ernesto Estrada[a,b,*,‡], Alfredo Peña[a] & Ramón García-Domenech[b]
[a]*Department of Drug Design, Centro de Bioactivos Químicos, Universidad Central de Las Villas, Santa Clara 54830, Villa Clara, Cuba;* [b]*Research Unit on Molecular Connectivity and Drug Design, Department of Physical Chemistry, Faculty of Pharmacy, University of Valencia, Valencia, Spain.*

Received 20 November 1997; Accepted 20 March 1998

Key words: active fragments, bond matrix, discriminant analysis, drug design, graph theory, spectral moments

Summary

A novel approach to computer-aided molecular design is illustrated. This approach is based on the calculation of the spectral moments of the bond adjacency matrix of graphs representing molecular structures. Spectral moments are then expressed as linear combinations of the different sub-structures present in molecules. Two series of compounds, one containing sedative/hypnotic and the other containing different classes of drugs were used to find a discriminant function with the present approach. Several compounds from the Merck Index were identified by the model as sedative/hypnotic, five of them were found in the recent literature as possessing this activity. The critical fragments, actives and inactive ones, were detected.

Introduction

At present there are more than 15 million chemical compounds that have been discovered or synthesized in chemical laboratories. A great quantity of these compounds have not found pharmacological or agrochemical applications yet. This is a consequence of the differences that exist between the rate at which novel chemicals are discovered each year and the number of compounds that can be tested in chemical or pharmacological assays. These kinds of assays, specially pharmacological and toxicological ones, are in general expensive and time consuming. However, novel paradigms for drug discovery have been introduced recently, based on the availability of large chemical libraries and robotics systems for bioassays. These systems of high-throughput biochemical assays allow for the synthesis and testing of hundreds of compounds each day [1].

During last years, the pharmaceutical industries have reoriented their research strategies in order to give more attention to those methods that permit the 'rational' selection or design of novel compounds with the desired properties [2–4]. Several approaches to the computer-aided molecular design have been introduced in the literature [5–8]. All of them are based on the relationships between the chemical structure and the properties (physical, physico-chemical and biological) of molecules. The success of these methods is very dependent of the molecular descriptors that are selected to characterize the chemical structure and on the use of appropriate statistical methods. Several kinds of molecular descriptors and approaches have been used for drug design, such as molecular mechanics [9, 10], quantum chemical descriptors [11], similarity/dissimilarity approaches [12, 13], topological descriptors [14–17], physico-chemical descriptors [18, 19], 3D-QSAR [20–23], etc.

On the other hand, chemists have been interested for a long time in property representation as additive functions in which simple physical properties are expressed in terms of the different fragments of molecules with acceptable precision. The use of additivity functions to describe biological activities has also been studied in the literature. Some of the approaches used

*To whom correspondence should be addressed.
‡On leave from Centro de Bioactivos. E-mail address: farmaco@cbqvc.sld.cu.

in the generation of such models are the *de novo* schemes such as the Free–Wilson method [24].

Chemical graph theory can be applied to obtain additive models of properties, including biological ones, in a simple and efficient way. Smolenskii [25] was the first who developed a graph theoretical approach to express physical properties of organic molecules in terms of structural fragments. The method of cluster expansion also permits the generation of such models in a graph-theoretical way [26, 27].

Recently, one of the present authors [28–30] has developed a graph-theoretical approach that expresses physical [27–29], and biological properties [29] in terms of substructural features of molecules. This method is based on the calculation of the spectral moments of the bond matrix [31] of molecular graphs. The bond matrix has also been used by this author in order to generate graph-theoretical descriptors [31–33] which are very useful in quantitative structure-property (QSPR) and quantitative structure-activity (QSAR) relationships.

The method of spectral moments of the bond matrix [28] is based on the calculation of traces of the different powers of the bond matrix (spectral moments), then the property to be described is expressed in terms of the moments as in traditional QSPR methods. The main advantage of this novel approach is that the spectral moments can be expressed as linear combinations of structural fragments of molecules. As a consequence, we can substitute the spectral moments in the quantitative model by their expressions in terms of structural fragments of the molecules obtaining an equation that relates the property directly with the molecular structure. A second approach that permits the study of molecules containing any kind of heteroatoms in the structure was developed [29] by using edge weights in the molecular graphs. This method will be called henceforth as TOpological Sub-Structural MOlecular DEsign (TOSS-MODE).

The objective of the present work is to use the TOSS-MODE approach in the generation of a discriminant function that permits the classification of molecules as sedative/hypnotic. Then we will use this model to design novel compounds with the mentioned biological activity and we will also find the structural fragments that are responsible for this pharmacological action.

The TOSS-MODE approach

The present approach is based on the calculation of the spectral moments of the bond adjacency matrix of molecular graphs, defined as the traces of the different powers of such matrix. The spectral moments of this matrix have been expressed as linear combinations of the number of times that the different structural fragments appear as sub-graphs in the molecular graph. These linear combinations for simple molecular graph, i.e., not considering heteroatoms in the molecular structure, have been reported for acyclic [28] and cyclic [30] molecules. This approach was extended to consider molecules containing any kind of heteroatoms by the introduction of weights in the diagonal entries of the bond adjacency matrix [29].

The spectral moments $\mu_k^d = \mu_k^d(G, \mathsf{E})$ of the bond weighted adjacency matrix also contain information on the different fragments present in molecules. However, due to the combinatorial explosion in the number of structural fragments when heteroatoms are considered it is almost impossible to express spectral moments of the bond-weighted matrix as linear combination of the different fragments. For instance, the very simple expression which relates the fourth spectral moment of the bond matrix in non-weighted graphs [28], (see also Figure 1):

$$\mu_4 = 2 \cdot |F_2| + 12 \cdot |F_3| + 24 \cdot |F_4| + 4 \cdot |F_5| \quad (1)$$

is transformed into expression (2) which contains 4 additional terms [29]:

$$\mu_4^d = \mu_4 + \sum_i (d_i)^4 + 4 \sum_i (d_i)^2 + 8 \sum_i F_3^i d_i + 4 \sum_r (D_i d_j)_r . \quad (2)$$

These additional terms contain the information about the structural fragments which contribute to this spectral moment. However, the number of such fragments is very large even for the fourth spectral moment. For instance, 54 fragments of alkyl halides contribute to this spectral moment [29], and this number increases dramatically with the number of heteroatoms considered.

In expressions (1) and (2) we have used the following terminology. The term $|F_k|$ in Equation (1) is the number of fragments of type k in the molecular graph, d_i in Equation (2) is the weight used in the diagonal entries of the bond matrix, and F_k^i is the number of fragments F_k containing the bond i.

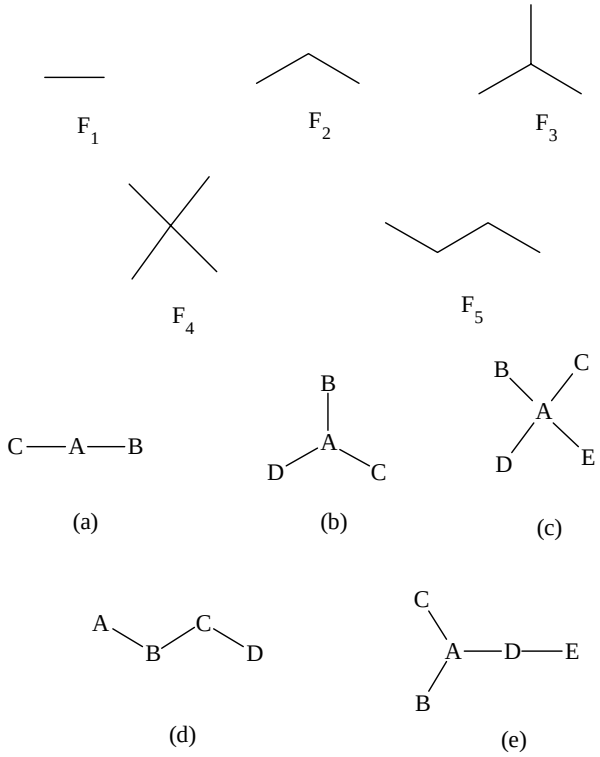

Figure 1. Structural fragments present in Equation (2) as A(BC) (a), A(BCD) (b), A(BCDE) (c), A-B-C-D-E (d) and A(BCD-E) (e).

Taking into account that the spectral moments of the bond matrix contain the sub-structural information of molecules, we have proposed their use in studies of structure-property and structure-activity relationships [28–30, 34]. The first step in this approach is to obtain a quantitative relationship between a property P and the spectral moments using standard statistical techniques. The equation obtained using such techniques and containing several spectral moments is expressed as the following general model:

$$P = a_0\mu_0 + a_1\mu_1 + a_2\mu_2 + a_3\mu_3 + \\ + \ldots + a_i\mu_i + \ldots + a_n\mu_n + b. \quad (3)$$

This equation can represent a QSAR model obtained from multivariate regression techniques, a discrimination function obtained from discriminant analysis, or any kind of quantitative model obtained from standard statistical methods. The selection of the spectral moments to be included into this equation is carried out with the use of the traditional statistical procedures, such as the stepwise selection.

After some algebraic manipulation, model (3) can be transformed into another expressing the property P directly in terms of structural fragments of molecules. This transformation is carried out by substituting the spectral moments by their corresponding expressions in terms of fragments, obtaining model (4):

$$\begin{aligned} P = {} & b + \sum_i (b_{AB}|AB|)_i \\ & + \sum_j (b_{A(BC)}|A(BC)|)_j \\ & + \sum_k (b_{A(BCD)}|A(BCD)|)_k \\ & + \sum_l (b_{A(BCDE)}|A(BCDE)|)_l \\ & + \sum_m (b_{A-B-C-D-E)}|A-B-C \\ & \qquad -D-E)|)_m \\ & + \sum_n (b_{A(BCD-E)}|A(BCD-E)|)_n \\ & + \ldots . \end{aligned} \quad (4)$$

In this expression, indexes i, j, k, l, m and n run over the different fragments in the molecules. The symbol *AB* represent a bond in which atoms are denoted by *A* and *B*. The fragments *A(BC)*, *A(BCD)*, *A(BCDE)*, *A-B-C-D* and*A(BCD-E)* represent different sub-structures in which *A*, *B*, *C*, *D* and *E* are atoms at different positions in the fragment as illustrated in Fig 1. The number of fragments included into this model is dependent on the maximum order of the spectral moment included in model (3), i.e., the value of n in μ_n.

Coefficients in Equation (4) represent the specific contribution of the corresponding fragment to property *P*. There are two approaches to obtain the values of these coefficients. The first approach uses the knowledge of the analytical expressions for the spectral moments in terms of molecular fragments, that is, expressions like Equation (2). In this case we only need to substitute the values of the spectral moments by their expressions in terms of fragments and we obtain model (4). The main limitation of this approach lies on the combinatorial explosion in the number of fragments contributing to higher order spectral moments. For this reason, at present we can apply this methodology when the order of spectral moments is not too high, for instance $n \leq 10$ [29–30, 34]. However, the future development of analytical expressions for higher order spectral moments will permit

the application of this method to a broad spectrum of structure-property problems.

The second method that we propose for the evaluation of these coefficients is based on a computational approach. 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 procedure is as follows. First, we select the substructure whose contribution to the moments we would like to determines. Then we generate all the fragments (sub-graphs) 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. After that, we only need to substitute these contributions into the quantitative model developed to describe the property studied, e.g., model (3), and we obtain the quantitative contribution of the different fragments to *P*, as in expression (4). This second approach will be used in the present study.

In order to carry out this labor efficiently we have developed a computer software named TOSS-MODE for Windows 95 [35]. This software allows the user to draw any series of structural fragments and then it calculates the contributions of these fragments to the different spectral moments of the bond matrix. A database of substructures, developed by accumulation of the different fragments which are generated in different works, can also be used in the calculation of contributions.

Development of the discrimination function

The use of discriminant analysis in drug design by using topological descriptors has been extensively applied by the Valencia group [16, 36–40]. The first step in the search of a discriminant function which permits the classification of drugs according to their biological activity is the selection of a training set composed of active and inactive compounds. This training set is made up, in the present study, of 187 compounds, 72 of them actives, i.e., sedative/hypnotic, and 115 compounds having other different biological activities, such as antibiotics, antivirals, antihistaminics, antiinflamatories, antifungals, analgesics, antineoplastics, cytotoxics, vasodilators, antiparasitics, central nervous system stimulants, relaxants, anorexics, antiseborrheics, and so forth. A test set composed of 77 compounds was chosen at random from these 187 compounds, 40 of them are actives and 37 are inactive. Compounds in the prediction set were never used in the development of the discriminant function during this work; they were only used to validate the model found.

The first 15 spectral moments of the edge-weighted adjacency matrix for all graphs representing molecules of the training and prediction sets were calculated by using standard bond distances and standard bond dipole moments. Both kinds of spectral moments were used in order to find the best discrimination function for the training set of compounds. The best result was obtained by using the bond dipole moments as edge weights in the graphs. In Table 1 we illustrate the values of the bond dipole moments which are used in the present work.

The discrimination function was obtained by using the stepwise linear discriminant analysis as implemented in STATISTICA version 4.13 [41]. The default parameters of this program were used in the development of the model. The best discrimination function obtained for the training set is given below:

$$\begin{aligned} Act \;=\; & 1.52\mu_0 - 0.43\mu_2 - 0.19\mu_3 - 0.21\mu_4 \\ & + 6.37 \cdot 10^{-2}\mu_6 - 4.72 \cdot 10^{-3}\mu_8 \\ & + 1.19 \cdot 10^{-4}\mu_{10} - 4.48 \cdot 10^{-9}\mu_{15} \\ & - 3.56. \end{aligned} \tag{5}$$

The value of the Wilk's λ parameter for this model is 0.661 and the Fisher ratio F (8, 178) = 11.43. There was no one compound detected as statistical outlier in the training set. In Table 2 we illustrate the results obtained in the classification of compounds of the training set. Here we illustrate only 72 of the 115 inactive compounds which were used in the development of the discrimination function. Compounds were considered as unclassified or undetermined by the model, denoted by U in Table 2, when the differences in the percentage of classification as actives and inactives do not differ in more than 5%. For these compounds the probability was too close in the change for a decision to be made as to whether they were active or inactive. The percentage of undetermined compounds in the training set is 5.35% (10/187). In this set the 8.02% (15/187) of compounds are classified as false active and the 10.16% (19/187) as false inactive. False active and false inactive compounds are those that the model predicts as active or inactive and they are in-

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 a positive charge on the hydrogen atom.

active and actives, respectively. The overall accuracy of the model is 76.47% (143/187) for the training set, while the adjusted overall accuracy, i.e., that obtained when undetermined compounds are excluded of the calculation, is 80.79% (143/177).

One of the most important aspects of any quantitative structure-property model is its ability to predict the property studied for compounds not included in the training set. When the discrimination function (5) is applied to the prediction set of 77 compounds we obtain the following results. The percentage of undetermined compounds is only 1.30% (1/77), while the false actives and false inactive are 2.60% (2/77) and 14.28% (11/77), respectively. The overall accuracy in the prediction set is 81.82% (63/77) and the adjusted overall accuracy is 82.89% (63/76). If we take into account only the number of compounds used in the prediction set we see that 72.5% (29/40) of the actives and 94.4% (34/36) of the inactive are well classified by the quantitative model found by using the spectral moments of the bond matrix. In Tables 2 and 3 we give the classification of compounds in the training set and in the prediction set as active (+), inactive (−) or undetermined (U).

It is interesting to analyze the false actives which are found in the training and the prediction sets. These are compounds having different biological activities which were predicted by the model as possessing sedative/hypnotic properties. We look for some of these compounds in an intensive literature search in order to find some reports concerning their sedative/hypnotic properties. Two of them were reported to have sedative or hypnotic properties in the literature, they are the Eucalyptol and the Baclofen. Ortiz de Urbina et al. [42] reported in 1989 that the essential oil of *Calamintha sylvatica subsp. ascendens* exerts significant sedating and antipyretic activities in rats. Among the responsible components for these activities appears the Eucalyptol (1,8-cineol). On the other hand, the (–)-Baclofen, a muscle relaxant drug, has been observed to has sedative and antinociceptive effects [43]. This is a very interesting result because it proves that the discriminant function found is able to recognize compounds whose structures are unrelated to those of known sedative/hypnotic activity, as in the case of Eucalyptol.

It is evident from the results of the statistical analysis of the active compounds that there are some structural patterns that are not correctly described by the discriminant function obtained here. This is reflected by the fact that there exists 10% and 14% of false inactives in the training and prediction sets, respectively. These compounds that have sedative/hypnotic activity are recognized by the model as inactive possibly because there are some structural features of these molecules not well-accounted for in the first 15 spectral moments of the bond matrix or in the bond weights used in the present study. The inclusion of a greater number of spectral moments, not only up to μ_{15}, and/or the use of other bond weights can produce a decrease of the number of false inactives. However, there is also the possibility that the biological activity of these compounds dependent on structural features which are not accounted for in the present approach based on spectral moments. In this case, the combination of the present method to other well-established approach to drug design can produce the improvement desired.

Design of sedative/hypnotic compounds

There are two different approaches to finding novel sedative/hypnotic compounds: (i) to design novel compounds to be synthesized in the chemical laboratory and then tested for the biological activity or (ii) to find known compounds, with other activities or uses, in databases and to test them for the sedative/hypnotic action. Both approaches have strong and weak points, the first permits the design of novel lead compounds having the desired activity but they first need to be synthesized, then tested for the pharmacological activity and finally they need to pass through the toxicological, pharmacodynamical and pharmaceutical tests. However, compounds selected from the second way can have well-established methods of synthesis and in many cases their toxicological, pharmacodynamical and pharmaceutical properties are well-known. This is, for instance, the case of compounds having other

Table 2. Classification of compounds in the training series by the discriminant function obtained with the TOSS-MODE approach

Training active group			Training inactive group[a]		
Compound	Prob.	Class.	Compound	Prob.	Class.
Bromobutanol	41.5	−	Depreton	92.1	−
Ectylurea	64.7	+	Metirosine	67.4	−
Carbromal	85.1	+	Regutensin	81.2	−
Cenestil	54.5	+	Methyldopa	68.3	−
Somnamid	53.4	+	Hydralazine	77.3	−
Isopral	27.4	−	MJ10459−2	80.5	−
Lamotane−X	79.5	+	Dihydralazine	21.5	+
Fenobam	50.4	U	Pildralazine	71.1	−
Trichloroisobutylsalicylate	91.4	+	Oxdralazinehydrochloride	88.4	−
Porpionylphenetidin	63.1	+	Clonidinehydrochloride	61.5	−
Thalidomide	94.3	+	Lofemizolehydrochloride	90.4	−
Bromoglycine	5.1	−	Chlorthenoxazine	88.3	−
Carbochloral	61.9	+	Tianafac	62.5	−
Hedonal	34.2	−	GOBAB	84.8	−
Etchlorovynol	53.6	+	Ag307	88.3	−
Penthrichloral	79.1	+	Abbott−29590	93.5	−
Bromizoval	22.7	−	Tizoprolicacid	58.6	−
Acecarbromal	94.3	+	Triclacetamol	46.6	−
Brallobarbital	50.3	U	Tetridamine	68.5	−
Capuride	71.4	+	Geraniol	89.1	−
Fepiron	56.1	+	VUFB−7904	87.8	−
Femerazo	45.3	−	Diethylcarbamazine	72.2	−
Nimustine	54.4	+	Tetramizolehydrochloride	67.6	−
Centazolone	19.8	−	Ciclobendazole	37.8	+
Glutethimide	94.4	+	Bitoscanate	99.1	−
Alonimid	95.1	+	Certuna	27.1	−
Paracetaldehide	4.6	−	Wormin	70.2	−
Brocalcin	50.8	U	Tiabendazole	90.5	−
Etomidate	53.1	+	Iodothymol	88.8	−
Zapizolam	59.2	+	Bromthymol	88.1	−
Lormetazepam	57.6	+	Noxitiolin	82.2	−
Chinoin−1045	80.3	+	Selectan−neutral	90.5	−
Baldrianol	32.7	−	Antibrucellin	77.9	−
Amylurea	50.7	U	Spinulosin	91.8	−
Anansiol	92.2	+	Thienamicyn	81.3	−
Petrichloral	100	+	Isopropicillin	10.8	−
Methaqualone	71.1	+	Cefamandole	16.1	+
Roletamide	76.9	+	Patulin	83.8	−
Quisqualamine	26.5	−	Fervenulin	94.7	−
Tricetamide	81.3	+	Tromacaps	94.4	−
Valperinol	67.4	+	Fumigatin	89.0	−
Triazolam	66.5	+	Flucytosine	89.8	−
Butoctamide	51.8	U	Fuberidazole	62.6	−
Fenadiazole	14.5	−	Azaconazole	88.7	−
Haloxazolam	81.9	+	Buclosamide	58.3	−
Carburazepam	91.1	+	Protiofate	64.7	−
Etaqualone	85.1	+	Desderman	96.9	−
Zopiclone	6.5	−	Mercaptosuccinic acid	74.0	−

Table 2. (continued)

Training active group			Training inactive group[a]		
Compound	Prob.	Class.	Compound	Prob.	Class.
Apomorphine	70.0	+	Ribavirin	93.8	–
Metomidate	50.9	U	BVDU	84.7	–
Neobonyval	77.0	+	Cutison	76.2	–
Cinolazepam	66.2	+	Acedoben	44.1	+
Anetamin	26.6	–	Carbodine	78.9	–
EthylLoflazepate	80.7	+	Dimepranol	97.4	–
Perlapine	86.2	+	Riodoxol	94.8	–
Valerylphenetidine	77.7	+	Ketoxal	95.1	–
Amylenehydrate	45.3	–	Citenazon	93.7	–
Azoperona	83.6	+	Aciclovir	95.6	–
Suriclone	50.6	U	Metoprine	79.3	–
Canadine	60.1	+	Pathocidin	76.9	–
Valeric acid	42.1	–	Fluoxidin	98.1	–
Propionazine	94.6	+	Camfazolinum	87.8	–
Febarbamate	97.1	+	BA1	83.7	–
Bason	28.7	–	Bromebricacid	52.6	U
Alozafone	86.6	+	Lysepsina	94.1	–
Cloperidone	93.7	+	Novembitol	93.2	–
Benzoclidinehydrchlorid	66.5	+	Formycin	93.0	–
L–Tryptophan	54.8	+	Thioguanosine	88.8	–
Cetohexazine	16.5	–	Dabutaminehydrochloride	58.8	–
Valbornine	60.0	+	Bromotheaminum	82.8	–
Menthoval	61.3	+	Milrinone	80.9	–
Methamphidon	72.4	+	Quazodine	81.3	–

[a]Only 72 of the 115 compounds are shown.

biological activities which are marketed as drugs. For these reasons we have selected the second kind of search for the novel sedative/hypnotic compounds.

We have carried out an exhaustive search in the Merck index [44] looking for compounds which are predicted to be active by the model. A few compounds were identified by the discrimination function as possible sedative/hypnotic, among them we can find known drugs with other pharmacological properties and several natural products with different uses. We also looked for these compounds in the literature in order to determine if they have been reported as possessing the sedative/hypnotic activity. In Table 4 we show five of them identified by the discrimination function as actives which are reported in the literature as sedatives or hypnotic. There is great variability in the functions of these compounds, two anesthetic, one analgesic, one antidyskinetic and one hormone, and also there is great variability in their molecular structures. Other compounds identified as actives but not reported in the literature as sedative/hypnotic are now in analysis in order to demonstrate their pharmacological activity. Among them there are several natural products used in foods and perfumery.

This result is the most important validation for the model developed here because it has been able to detect a series of compounds as active from a database composed of thousands of chemical compounds and some of these compounds have shown the predicted property.

Active and inactive sub-structures

One of the most interesting features of the TOSS-MODE approach to molecular design is the possibility to obtain the quantitative contribution of any kind of sub-structure to the property studied. The number of structural fragments which can be evaluated to determine their contribution to the sedative/hypnotic activity is, of course, very large.

Table 3. Classification of compounds in the external prediction series by the discriminant function obtained with the TOSS-MODE

Test active group			Test inactive group		
Compound	Prob.	Class.	Compound	Prob.	Class.
Bromazepam	64.3	+	Guanfacine	61.1	−
Fepitrizol	33.0	−	Debrisoquin	56.4	−
Chlordiazepoxide	78.3	+	Alarnine	76.2	−
Ethyldirazepate	80.9	+	Praxadine	94.2	−
Flurazepam	77.2	+	Nimazone	89.5	−
Lorazepam	56.0	+	Strinoline	94.2	−
Chlometiazole	0.7	−	R8231	82.7	−
Aponal	79.5	+	Antienite	85.2	−
Pentizidone	64.3	+	Diamide	72.0	−
Subdamine	23.1	−	Azaserine	77.4	−
Diazepam	80.3	+	Cefatiam	23.8	+
Carbromide	59.5	+	Furanomycin	74.7	−
Hoechst264	75.3	+	Mycosid	73.6	−
Ethanion	35.1	−	Trichofytocid	97.0	−
Isoladol	66.3	+	Acluracil	52.5	U
Aleudrin	2.0	−	IMPY	95.5	−
Butoctamidesuccinate	94.5	+	Moroxydine	90.7	−
Profexalone	75.2	+	Mercaptopurine	98.6	−
Damotepine	75.5	+	Guanazole	86.3	−
GYKI21622	32.6	−	Tenuazonicacid	67.4	−
Tuso−Blandin	93.3	+	Tioguanine	94.7−	
Niaprazine	69.0	+	Enoximone	97.2	−
Thiourethane	84.4	−	Heptaminol	97.2	−
Phenodoxone	98.7	+	Protheobromine	96.7	−
Valyfen	78.4	+	Nicotinicacid	90.8	−
Chlorazepate	84.5	+	Fenamole	29.2	+
Nordazepam	78.2	+	Manozodil	85.3	−
Alcabrol	7.7	−	Salsolidine	76.7	−
Oxazepam	64.6	+	Glycabylamide	89.8	−
Prazepam	85.4	+	Aminopicoline	90.9	−
Temazepam	66.5	+	Bijosal	88.1	−
Chloralhydrate	22.1	−	APPA	73.3	−
Meprobamate	95.5	+	Progallin−P	72.7	−
Brotizolam	55.6	+	Ethylnoradrenaline	66.9	−
Ibrotamide	57.1	−	Isoverin	80.3	−
Midazolam	70.4	+	Piperazine	95.1	−
Ketazepam	93.8	+	Aminothiazole	87.8	−
Clobazam	85.8	+			
Bromophenazone	67.7	−			
Alprazolam	74.1	+			

Table 4. Compounds predicted by the TOSS-MODE approach as sedative/hypnotic after a search in the Merck index and then found in the literature as possessing this activity

Compound	Reported activity[a]	% classification	Ref.[b]
Melatonine	hormone of the pineal gland	60.8	45
Midazolam	anesthetic	70.4	46
Haloperidol	antidyskinetic	99.4	47
Fentanyl	analgesic, narcotic	98.87	48
Ketamine	anesthetic	84.4	49

[a]The biological activity for which the compound is known and/or marketed.
[b]Reference of the experimental determination of the sedative/hypnotic properties of each compound.

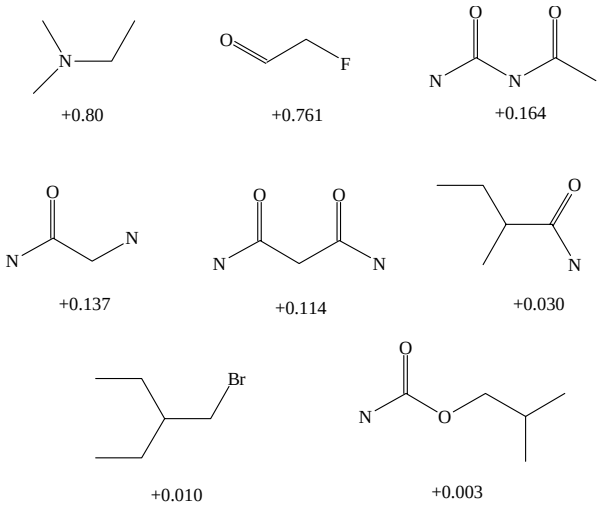

Figure 2. Some active (sedative/hypnotic) sub-structures found by using the TOSS-MODE approach.

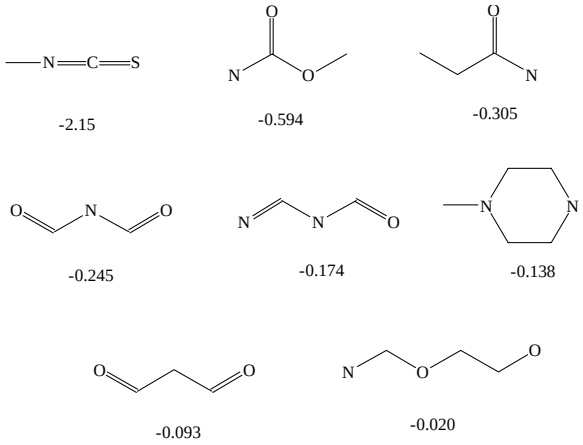

Figure 3. Some inactive (non-sedative/hypnotic) sub-structures found by using the TOSS-MODE approach.

Using the approach previously described, we have generated a number of structural fragments of interest and determined if they contribute positively or negatively to the sedative/hypnotic activity. In Figure 2 we illustrate some of these fragments having a positive contribution to the sedative/hypnotic activity. Some of these fragments appear in very well known sedative/hypnotic compounds, such as Meprobamate, Acecarbromal, barbiturics, and so forth. In Figure 2 we have also shown some structural fragments that appear in a wide range of chemical compounds such as substituted amides, bromine compounds, etc.

From the practical point of view, one of the most interesting things that the present approach allows is the detection of those fragments which contribute negatively to the property studied. In Figure 3 we show some of these structural fragments which have been detected by the TOSS-MODE approach. There are some of these fragments which appear in active molecules, such as two fragments which always appear in barbituric compounds. Of course, the presence or not of the biological activity in one compound depends on the sum of the contribution of all fragments present in the molecule. In this case the sum of contributions coming from all fragments in barbiturics determine that they are classified as sedative/hypnotic compounds. However, maybe those fragments having negative contribution to the biological activity are not directly related to the interaction of such compounds with the biological receptor and they are not determinant for the development of such activity.

The inclusion or not of fragments having negative contribution in the design of new molecules expected to have the desired biological property should be decided with care. In the first place, these fragments can be effectively inhibitors of the desired biological activity, and they should be eliminated from the structures of new compounds. However, in other cases

these fragments can contain other substructures which are really responsible for the biological activity. There is a third possibility that should be taken into account, this is concerned with the fact that even when a fragment has negative contribution to the biological activity, it can contribute in a decisive way to the distribution, metabolism, pharmacokinetics, etc., of such compounds.

As an example we illustrate the case of the fragment of piperazine bonded to a molecule for one of the nitrogen atoms. The contribution of such fragment is negative, having a value of -0.138. However, there are several compounds having a piperazine in their structures and possessing the sedative/hypnotic activity. Five of these compounds were included in the present study: Zopiclone, Perlapine, Suriclone, Cloperidone and Niaprazine, the last one as a member of the prediction set. Two of them, Zopiclone and Suriclone, were not well-classified or non-classified by the model, respectively. In order to increase the number of compounds having a piperazine fragment in the prediction set, we select five more compounds which were evaluated by model (5). The results obtained for these compounds are as follows: Etodroxizine (+, 84.7%); Meclozine (+, 95.2%); Peraptene (+, 91.5%); Toprilidine (+, 75.9%) and Laprazolam (+, 89.1%). As can be seen all compounds in the prediction set having a piperazine group in their structures were correctly predicted to be sedative/hypnotics. From the structural point of view, in all sedative/hypnotics having piperazine group, this fragment appears substituted in both nitrogen atoms. The contribution of this fragment, piperazine disubstituted in both nitrogen atoms, is positive (+ 0.036), as seen in Figure 4A. Now, the following question arises: Is the disubstituted piperazine fragment the responsible for this biological activity or is it a fragment contained in it? The result obtained by using TOSS-MODE is the following. Even when the disubstituted piperazine fragment has a positive contribution to the sedative/hypnotic activity, there are several other fragments contained in it which also have positive contributions. These fragments, in general, involve both nitrogen atoms with substitutions. Some of these fragments are illustrated in Figure 4A. There are two interesting features in these fragments. First, that the positive contribution of the disubstituted piperazine appears to be related to the second fragment in this figure, which has almost the same contribution than it. The second feature is that these fragments contained in the disubstituted piperazine are also present in many sedative/hypnotic drugs. In Figure 4B, we

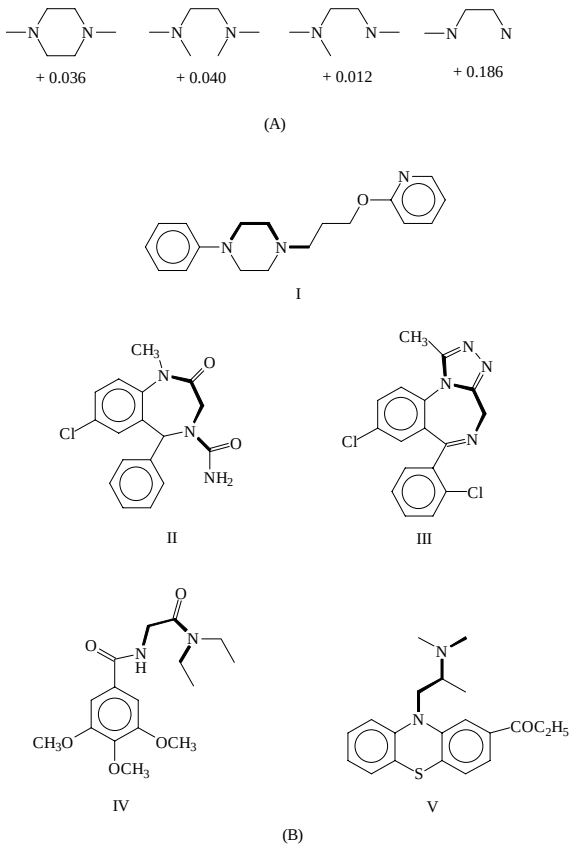

Figure 4. Contribution to the sedative/hypnotic activity of disubstituted piperazine and different fragments contained in it (A), and some sedative/hypnotic drugs which have a common fragment included into disubstituted piperazine (B). The names of compounds are as follows: Toprilidine (I), Carburazepam (II), Triazolam (III), Valperimol (IV) and Febarbamate (V).

have shown some sedative/hypnotic compounds having the same common substructure, which correspond to the fragment given in Figure 4A which has the maximal contribution to the activity.

It is evident that these fragments can be considered as a necessary, but not sufficient, condition for production of a desired biological activity. The main limitation of this kind of substructures is given by their topological nature, which obviously does not contain explicitly three-dimensional information about the pharmacophore. However, we show in Figure 5 that this common fragment has conformational similarity in at least two compounds containing it. The conformational similarity was evaluated by calculating the root mean square (rms) deviation between heavy atoms after superimposition of this fragment in two molecules. Previously, we optimized the geometries of both molecules by using the molecular mechanics

Table 5. Classification of barbituric compounds in a prediction set as sedative/hypnotic compounds

Compound	Prob.	Class.	Compound	Prob.	Class.
Pentobarbital	67.5	+	Methohexital	78.5	+
Fenobarbital	84.7	+	Amobarbital	50.7	U
Mefobarbital	78.5	+	Thiopental	56.3	+
Barbital	52.5	U	Allobarbital	78.0	+
Ethallobarbital	66.5	+	Crotarbital	74.7	+
Butobarbital	74.7	+	Butalbital	50.8	U
Butabarbital	55.0	+	Propylbarbital	78.0	+
Aprobarbital	51.4	U	Butallylonal	37.7	−
Cyclobarbital	84.8	+	Norhexobarbital	71.0	+
Secobarbital	78.8	+	Pentenal	84.8	+
Metabarbital	41.7	−	Bromoaprobarbital	33.7	−
Talbutal	68.4	+	Hexobarbital	60.9	+
Heptobarbital	71.0	+			

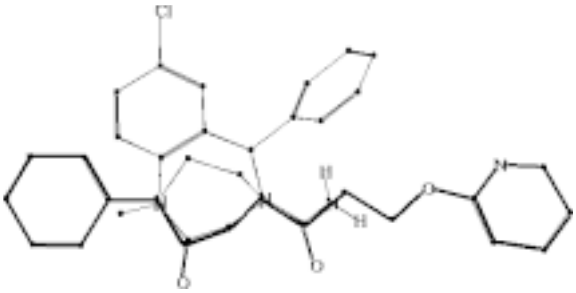

Figure 5. Superimposition of a common fragment contained in disubstituted piperazine and present in Toprilidine (bold) and Carburazepam.

force field mmX as implemented in PCMODEL software package [50]. The rms obtained for the five atoms in this fragment is only 0.108 Å for the molecules compared in Figure 5.

The final conclusion in relation to the disubstituted piperazine fragment is that maybe it is just a bioisoster of a smaller fragment which is contained in it. However, this hypothesis needs an experimental verification which is not provided here. Similar kind of bioisosterism is given in many other drugs, such as Hexetal and Ciclobarbital, or Promazine and Metdilazine, for which an acyclic chain is transformed into a cyclic one maintaining the biological activity of the bioisoster.

As may be seen, the TOSS-MODE approach not only permits the correct classification of a structurally diverse set of compounds according to a specific biological activity, but it also permits the determination of the sub-structures which are critical for the development of this pharmacological activity. In this sense the present approach to molecular design represents a step forward in the development of a theory which permits not only the prediction of biological activities but also the easy interpretation of the results in terms of structural concepts.

An experiment of lead generation

One of the main objectives of the approach developed here is the selection of *sub-systems* from a large group of chemical compounds. A sub-system is in general understood as a number of chemical compounds formed by a significant variations in a given parent structure, which is referred to as the *lead compound*. The model found by using the TOSS-MODE approach recognize some structural patterns which are non-related to the common patterns that appear in sedative/hypnotic compounds, such as those of eucalyptol and melatonine.

It is believed that the chance of discovering compounds with the biological properties desired is greater in those sub-systems than in any other. Franke [51] has stated that unfortunately even the first step in the search for new drugs, which is usually called ‘lead generation’, involves a high degree of uncertainty. In the present approach, this uncertainty depends in a great part of the correct selection of the training set used in the development of the model. An adequate selection of this training set should permit the se-

lection of novel compounds whose structures are not contained in the training set.

In the present section we simulate an experiment of lead generation. This hypothetical experiment is based on the assumption that barbituric compounds are unknown in the literature. Consequently, the training set used in the development of the model should contain not one of such compounds. The training set that was used in the development of the model (5) contains only one barbituric compound: Brallobarbital. By eliminating this compound from the original training set we have now a new series of compounds containing no one barbituric. The discrimination function found using the modified training set is given below:

$$\begin{aligned} Act \;=\; & 1.53\mu_0 - 0.40\mu_2 - 0.19\mu_3 - 0.22\mu_4 \\ & + 6.40 \cdot 10^{-2}\mu_6 - 4.68 \cdot 10^{-3}\mu_8 \\ & + 1.14 \cdot 10^{-4}\mu_{10} - 4.04 \cdot 10^{-9}\mu_{15} \\ & - 3.53. \end{aligned} \qquad (6)$$

Now, we suppose that our organic synthesis laboratory has obtained a new series of compounds (a sub-system) which will be evaluated by the model (6) in order to predict if they have or not sedative/hypnotic activity. As this series of 'new compounds' we select 25 barbiturics. The predictions carried out by the model (6) for the new sub-system are illustrated in Table 5. The percentage of undetermined compounds is 16% (4/25) and the false inactive are the 12% (3/25). The overall accuracy of this prediction is 72% (18/25) and the adjusted overall accuracy is 85.7% (18/21). This percentage of good classification for the novel set of compounds permits to identify them as a new subsystem composed of a new lead structural pattern: the barbituric pattern.

In closing the TOSS-MODE approach has been able to recognize 18 of the 25 barbiturics as sedative/hypnotic even when the training set contains no barbituric compounds. This result is a simulation of the lead generation procedure. It is similar to the design of a new class of lead compounds (barbituric compounds) with the desired biological activity from a data set which does not contain a similar structural pattern. The question concerned with the presence or not in the training set of some fragments contained in the barbituric pattern is not important here. Spectral moments are 'global' descriptors which contain information on the molecule as a whole, even when they can be expressed in terms of structural fragments. The recognition of Eucaliptol as an active molecule can be offered as an illustration of this fact.

The next step in this approach would be the inclusion of these 'novel' compounds into the training set and the developing of a new discrimination model. This new model can be significantly different from the previous one, due to the inclusion of a new structural pattern, but it should be able to recognize a greater number of such compounds as active. However, this point is out of the general scope of the present work.

Concluding remarks

Computer-aided molecular design has become a very important tool in the development of novel chemical compounds to be used in different areas of human life. Another aspect of molecular design which influences quality of life is its use in the search of new indications for known compounds. The present study has illustrated how theoretical methods can help this diversification of uses for known drugs and natural products.

The use of graph theoretical descriptors, such as connectivity indices, and the introduction of new ones have proved to be a useful tool in the search of novel drugs. The easy calculation of graph theoretical invariants permits their calculation for large databases of chemical compounds in relatively short time and using personal computers.

The present approach is very flexible in many aspects. Firstly, it can be applied to large sets of chemical structures in short time, it can also be modified to obtain several series of descriptors by changing the weights used in the diagonal entries of the bond matrix. For instance, we used here standard bond distances and standard bond dipole moments as edge weights, but we can also use bond polarizabilities, bond charges, or the average of properties calculated for the atoms forming the bond, e.g., atomic charges, electrotopological state indices, etc. The main step forward that the TOSS-MODE approach represents to the use of graph-theoretical methods in molecular design is that it permits the identification and quantification of the structural fragments which are responsible for the property studied. These results point out the importance of the continued search for novel graph theoretical invariants and descriptors to be used in molecular design.

Acknowledgements

We sincerely thank the referees for their critical opinions of the manuscript which contribute to a significant improvement of the presentation, and to the Editor for her patience and help. One of the authors (E.E) thanks the Program 'Una Nau de Solidaritat' for a fellowship to work at Valencia University. R. G-D. thanks the support by CICYT through the projects SAF96–0158-C02–02 (Ministerio Español de Educación y Ciencia) and GV-3300/95 (Generalitat Valenciana).

References

1. Lutz, M.W., Menius, J.A., Laskody, R.G., Domanico, P.L., Goetz, A.G., Saussy, D.L. and Rimele, T., Network Science, Vol. 2, Issue 9, September 1996. http:\\www.netsci.org\Science\Screening\ feature05.html.
2. Loew, G.H., Villar, H.O. and Alkorta, Y., Pharm. Res., 10 (1993) 475.
3. Briggs, J.M., Marrone, T.J. and McCammon, J.A., Trends Cardiovasc. Med., 6 (1996) 198.
4. Wess, G., Drug Discovery Today, 1 (1996) 529.
5. Veerapandian, B., In Wolff, M.E. (Ed.) Burger's Medicinal Chemistry and Drug Discovery, 5th ed., John Wiley & Sons, New York, NY, 1995, pp. 303–348.
6. Sun, E. and Cohen, F.E., Gene, 137 (1993) 127.
7. Kuntz, J, D., Meng, E.C. and Shoichet , B.K., Acc. Chem. Res., 27 (1994) 117.
8. Jackson, R.C., Curr. Opin. Biotech., 6 (1995) 651.
9. Siebel, G.L. and Kollman, P.A., In Hansch, C., Sammes, P.G., Taylor, J.B. and Ramdsen, C.A. (Eds.) Comprehensive Medicinal Chemistry. Vol. 4. Quantitative Drug Design, Pergamon, 1990, pp. 125–138.
10. Bowen, J.P. and Allinger, N.L., In Lipkowitz, K.B. and Boyd, D.B. (Eds.) Reviews in Computational Chemistry, Vol. 2, VCH, 1991, pp. 81–97.
11. Weinstein, H., Osman, R and Green, J.P., In Olson, E.C. and Christoffersen, R.E. (Eds.) Computer-Assisted Drug Design, ACS Symp. Ser., 112, 1979, pp. 161–170.
12. Johnson, M.A. and Maggiora, G.M. (Eds.) Concepts and Applications of Molecular Similarity, John Wiley Intersciences, New York, NY, 1990.
13. Heller, S.R., J. Chem. Inf. Comput. Sci., 32 (1992) 578.
14. Hall, L.H. and Kier L.B., In Lipkowitz, K.B. and Boyd, D.B. (Eds.) Reviews in Computational Chemistry. Vol. 1, VCH, 1991, pp. 367–422.
15. Bersuker, Y.B. and Dimoglio, A.S., In Lipkowitz, K.B. and Boyd, D.B. (Eds.) Reviews in Computational Chemistry, Vol 1, VCH, 1991, pp. 423–479.
16. Basak, S.C., Grunwald, G.D. and Niemi, G.I. In Balaban, A.T. (Ed.) From Chemical Topology to three-Dimensional Geometry, Plenum Press, New York, NY, 1997, pp. 73–116.
17. Gálvez, J. and García-Domenech, R., In Mosqueira, A. (Ed.) Diseño de Medicamentos, Farmaindustria, Madrid, 1994, pp. 357–392.
18. Hansch, C. and Klein, T.E., Acc. Chem. Res., 19 (1986) 392.
19. Fujita, T., In Hansch, C., Sammes, P.G., Taylor, J.B. and Ramdsen, C.A. (Eds.) Comprehensive Medicinal Chemistry. Vol. 4. Quantitative Drug Design, Pergamon, 1990, pp. 497–560.
20. Martin, Y.C., Bures, M.G., Danaher, E.A., Delazzer, J., Lico, I. and Pavlik, P.A., J. Comput.-Aided Mol. Design, 7 (1993) 83.
21. Kaminski, J.J., Adv. Drug Delivery Rev., 14 (1994) 331.
22. Cramer III, R.D., Patterson, D.E. and Bunce, J.D., J. Am. Chem. Soc., 110 (1988) 5959.
23. Kellogg, G., Kier, L.B., Gaillard, P. and Hall, L., J. Comput.-Aided Mol. Design, 10 (1997) 513.
24. Kubinyi, H., In Hansch, C., Sammes, P.G., Taylor, J.B. and Ramdsen, C.A. (Eds.) Comprehensive Medicinal Chemistry. Vol. 4. Quantitative Drug Design, Pergamon, 1990, pp. 589–643.
25. Smolenskii, G.A., Zh. Fiz. Khim., 38 (1964) 1288.
26. Klein, D.J., Int. J. Quantum Chem.: Quantum Chem. Symp., 20 (1986) 153.
27. McHughes, M.C. and Poshusta, M.C., J. Math. Chem., 4 (1990) 227.
28. Estrada, E., J. Chem. Inf. Comput. Sci., 36 (1996) 844.
29. Estrada, E., J. Chem. Inf. Comput. Sci., 37 (1997) 320.
30. Estrada, E., J. Chem. Inf. Comput. Sci., 38 (1998) 27.
31. Estrada, E., J. Chem. Inf. Comput. Sci., 35 (1995) 31.
32. Estrada, E., J. Chem. Inf. Comput. Sci., 35 (1995) 701.
33. Estrada, E. and Ramírez, A., J. Chem. Inf. Comput. Sci., 36 (1996) 837.
34. Estrada, E., J. Chem. Soc. Faraday Trans., 94 (1998) 1407.
35. Gutierrez, Y. and Estrada, E., TOSS-MODE for Windows '95, version 3.0. Universidad Central de Las Villas, Santa Clara, Cuba, 1997.
36. Gálvez, J., García-Domenech, R., De Julian-Ortiz, J.V. and Soler, R., J. Chem. Inf. Comput. Sci., 35 (1995) 272.
37. Gálvez, J., García-Domenech, R., De Julian-Ortiz, J.V. and Soler, R., J. Chem. Inf. Comput. Sci., 34 (1994) 1198.
38. García Domenech, R., García-Mareh, F. J., Soler, R., Gálvez, J., Antón-Fos, G.M. and De Julián-Ortiz, J. V., Quant. Struct.-Act. Relat., 15 (1996) 201.
39. Gálvez, J., García-Domenech, R., De Gregorio-Alapont, C., De Julián-Ortiz, J.V. and Popa, L., J. Mol. Graphics, 14 (1996) 272.
40. García-Domenech, R., De Gregorio-Alapont, C., De Julián-Ortiz, V., Gálvez, J. and Popa, L., Bioorg. Med. Chem. Lett., 7 (1997) 567.
41. STATISTICA version 4.13, Statsoft, Inc. (1993).
42. Ortiz-de-Urbina, A.V., Martin, M.L., Montero M. J., Moran, A. and San-Roman, L., J. Ethnopharmacol., 25 (1989) 165.
43. De Luca, C. and Massotti, M., Prog. Neuropsychopharmacol. Biol. Psychiatry, 14 (1990) 597.
44. The Merck Index 11th Edition, Merck & Co., Inc., New York, NY, 1989.
45. Tzischinsky, O., Sleep, 17 (1994) 638.
46. Klotz, W., Anaesthesiol. Reanim., 14 (1989) 347.
47. Levine, R.L., Crit. Care Clin., 10 (1994) 709.
48. Gerwels, J.W., Bezzant, J.L., Le Mariel, L., Pauley, L.F. and Streisand, J.B., J. Dermatol. Surg. Oncol., 20 (1994) 823.
49. Freye, G., Latasch, L., Schmidhommey, H. and Portoghese, P., Anaesthesist., 43 (1994) Suppl. 2, 552.
50. PCMODEL. Serena Software: Bloomington, USA.
51. Franke, R. Theoretical Drug Design Methods, Elsevier, Amsterdam, 1984.