Available online at www.sciencedirect.com
SCIENCE DIRECT®
Bioorganic & Medicinal Chemistry
Bioorganic & Medicinal Chemistry 12 (2004) 4467–4475

TOPS-MODE based QSARs derived from heterogeneous series of compounds. Applications to the design of new anti-inflammatory compounds

Maykel Pérez González,[a,c,*] Luiz Carlos Dias,[d] Aliuska Morales Helguera,[b,c] Yanisleidy Morales Rodríguez,[b] Luciana Gonzaga de Oliveira,[d] Luis Torres Gomez[e] and Humberto Gonzalez Diaz[c]

[a]*Unit of Service, Drug Design Department, Experimental Sugar Cane Station 'Villa Clara-Cienfuegos', Villa Clara, Ranchuelo 53100, Cuba*
[b]*Department of Chemistry, Faculty of Chemistry and Pharmacy, Central University of Las Villas, Villa Clara, Santa Clara 54830, Cuba*
[c]*Chemical Bioactive Center, Central University of Las Villas, Villa Clara, Santa Clara 54830, Cuba*
[d]*Universidade Estadual de Campinas, UNICAMP Instituto de Química CEP: 13084-971, C.P. 6154 Campinas, SP, Brazil*
[e]*Department of Pharmacy, Faculty of Chemistry and Pharmacy, Central University of Las Villas, Villa Clara, Santa Clara 54830, Cuba*

Received 1 April 2004; revised 22 May 2004; accepted 26 May 2004
Available online 26 June 2004

Abstract—A new application of *TOP*ological *S*ub-structural *MO*lecular *DE*sign (TOPS-MODE) was carried out in anti-inflammatory compounds using computer-aided molecular design. Two series of compounds, one containing anti-inflammatory and the other containing nonanti-inflammatory compounds were processed by a *k*-means cluster analysis in order to design the training and prediction sets. A linear classification function to discriminate the anti-inflammatory from the inactive compounds was developed. The model correctly and clearly classified 88% of active and 91% of inactive compounds in the training set. More specifically, the model showed a good global classification of 90%, that is, (399 cases out of 441). While in the prediction set, they showed an overall predictability of 88% and 84% for active and inactive compounds, being the global percentage of good classification of 85%. Furthermore this paper describes a fragment analysis in order to determine the contribution of several fragments towards anti-inflammatory property, also the present of halogens in the selected fragments were analyzed. It seems that the present TOPS-MODE based QSAR is the first alternate general '*in silico*' technique to experimentation in anti-inflammatory discovery.

1. Introduction

Anti-inflammatory drugs are widely used for the treatment of pain, inflammation, rheumatoid arthritis and osteoarthritis. The common dose limiting toxicity of anti-inflammatory compounds is the increased risk of gastrointestinal ulceration, perforation and hemorrhage.[1] The enzyme cyclooxygenase (COX) catalyses the biooxygenation of arachidonic acid to prostaglandin G_2, which serves as a precursor for the synthesis of prostaglandins, prostacyclins and thromboxanes, which are collectively termed as prostanoids.[2] The cyclooxygenase activity of the enzyme is the site of action of NSAIDs.[3,4] However, inhibition of prostanoid biosynthesis is associated with side effects such as ulceration and impairment of renal functions.[5] It has been well established that the cells express two isoforms of cyclooxygenases, namely cyclooxygenase-1 (COX-1) and cyclooxygenase-2 (COX-2).[6]

COX-1 is expressed in many normal tissues and is the major form present in platelets, kidneys, gastrointestinal

Keywords: TOPS-MODE; Linear discriminant analysis; Cluster analysis; Anti-inflammatory compounds.
* Corresponding author at present address: Estacion Territorial de Investigacion de la Caa de Azucar, Drug Design, Autopista Nacional Km 246, Ranchuelo 53100, Cuba. Tel.: +53-42281473; fax: +53-42-281130; e-mail: mpgonzalez76@yahoo.es

0968-0896/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmc.2004.05.035

tract and plays a key role in physiological processes,[7] whereas COX-2 is an inducible form by many pro-inflammatory cytokines and mitogens. The second isoform is generally not detectable in normal tissues, but is elevated in inflammatory conditions[8] and is also implicated in colon cancers,[9] and Alzheimer's disease.[10] COX-2 is also constitutively expressed in kidneys,[11] brain[12] spinal cord[13] and in mucosa of stomach.[14]

For this reason, novel anti-inflammatory compounds with more selectivity and less toxicity are needed for the future.[15,16] Many researchers worldwide have been worked in the synthesis and evaluation of novel compounds.[17–22]

On the other hand, Graph–Theoretical methods have shown to be very useful in QSAR problems in order to perform a rational analysis of different pharmacological, toxicological and other activities.[23,24] In the context of the Graph–Theoretical and Topological methods for modelling physicochemical and biological properties of chemical there has been introduced the *TOP*ological *S*ub-structural *MO*lecular *DE*sign (TOPS-MODE) approach. The TOPS-MODE has been applied to the description of physicochemical properties of organic compounds. Several applications for the design of biologically active compounds have been described.[25–32] Thereby, the aim of this work is to find rationality in the search of novel anti-inflammatory compounds using TOPS-MODE approach. Secondly, to continue the validation of the methods for describing biological activity of heterogeneous series of compounds.

2. Linear discriminant analysis and TOPS-MODE approach

Here, we use the TOPS-MODE approach to obtain molecular descriptors through which we developed the QSAR function. The mathematical details of the method have been largely reported,[24–30] thus we will outline only the fundamental remarks.

Briefly, this method codifies the molecular structure by means of the edge adjacency matrix **E** (likewise called bond adjacency matrix **B**). The **E** or **B** matrix is a square table of order m (the number of chemical bonds in the molecule).[37] The elements of such a matrix (e_{ij}) are equal to 1 if the bonds i and j are adjacent (it means that an atom exist, which participates either in the bond i or in the bond j) or 0 otherwise. In order to codify information related to heteroatom, the TOPS-MODE approach use $\mathbf{B}(\mathbf{w_{ij}})$ weighted matrices instead of **B**. The weights ($\mathbf{w_{ij}}$) are chemically meaningful numbers such as bond distances, bond dipole, bond polarizability or even mathematical expressions involving atomic weights such as polarizability.[31–36] These weights are introduced in the main-diagonal of the matrix $\mathbf{B}(\mathbf{w_{ij}})$. Afterwards, the spectral moments of this matrix may be used as molecular fingerprints in QSAR studies in order to codify molecular structure. By definition, the expression 'spectral moments' must be understood as the sum of the elements in the natural powers of $\mathbf{B}(\mathbf{w_{ij}})$. It means that the spectral moment of order k (μ_k) is the sum of the main diagonal elements (e_{ii}) of the matrix $\mathbf{B}(\mathbf{w_{ij}})^{\mathbf{k}}$. In the present work the $\mathbf{B}(\mathbf{w_{ij}})$ matrix was weighted in the main diagonal with the parameter $w_{ij} = h_i/\delta_i + h_j/\delta_j$, which characterizes the bond polarizability. In this expression h_i is the standard polarizability of the atom I bonded with j and δ_i is the vertex or atom degree.[34,35] Such a parameter μ_1 equals to sum of atom polarizability in the molecule. The calculation of the μ_k was carried out by means of the software package MODES LAB 1.0 b®.[38]

Linear discriminant analysis (LDA) has been the election statistical technique in most of the QSAR studies carried out using TOPS-MODE.[28,31,32] In the present work, a similar expression for the QSAR is derived.

$$A. = b + b_0\mu_0 + b_1\mu_1 + b_2\mu_2 + \cdots + b_k\mu_k, \qquad (1)$$

where, $A.$ (acronym of anti-inflammatory activity) is an indicator variable. This variable reach the values $A. = 1$ for anti-inflammatory compounds or $A. = -1$ for the nonactive ones. Deciding whether or not a compound may be classified as anti-inflammatory, is based on the information extracted from the literature.[40]

In Eq. 1 the b_k are the coefficients of the classification function determined by least squared as implemented on the LDA modulus of STATISTICA 6.0.[41] Forward stepwise was fixed as the strategy for variable selection.[42] In order to develop the QSAR for anti-inflammatory/nonanti-inflammatory compound discrimination, we use the first 15 μ_k as molecular descriptors. Examining Wilk's U-statistic, Mahalanobis distance, the percentage of good classification and the proportion between the cases and variables in the equation determined the quality of the model. Additionally, calculating the percentages of good classification in the external prediction series carried out the validation of the model. Compounds in the external prediction series were never used to develop the classification function.

One of the most important step in computer-aided search of novel anti-inflammatory is to design a representative, randomized training and predicting series. With this aim we select a large data set of 530 compounds having great structural variability; 154 of them are active (anti-inflammatory) and the others are non-anti-inflammatory.[39] Later, two k-means cluster analysis (k-MCA) were performed for active and inactive series of compounds.[43]

3. Computation of fragment contributions

Each of the μ_k spectral moments given in Eq. 1 contains structural information on the molecules that can be directly obtained by the following computational approach.[31] In this approach we calculated the spectral moment for all the fragments contained in a given substructure and by difference of these moments obtained the contribution of the substructure.

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

Having the contributions of the different structural fragments in which we are interested, we only need to substitute these contributions into the quantitative model developed to describe the property studied, for example, model[1] in which we obtain the quantitative contribution of the different fragments to P.

4. *k*-Means cluster analysis

The *k*-MCA may be used in training and predicting series design.[43,44] The idea consists of carrying out a partition of either active or nonactive series of compound in several statistically representative classes of chemicals. Thence, one may select from the member of all these classes of training and predicting series. This procedure ensures that any chemical classes (as determined by the clusters derived from *k*-MCA) will be represented in both compounds series (training and predicting). It permits to design both training and predicting series, which are representative of the entire 'experimental universe'. Figure 1 graphically illustrates the above-described procedure where two independent cluster analyses (one for the active compounds and other for the inactive compounds) were carried out to select a representative sample for the prediction and training sets.

A first *k*-MCA (*k*-MCA1) splits anti-inflammatory in four clusters with 28, 45, 24, 31 members and standard deviations of 0.32, 0.41, 0.23 and 0.36, respectively. On other hand the series of nonactive compounds was partitioned into four clusters (*k*-MCA2) with 127, 72, 56, 58 members and standard deviations of 0.63, 0.56, 0.43 and 0.48, respectively. Selection of the training and prediction set was carried out by taking, in a random way, compounds belongs to each cluster.

To ensure a statistically acceptable data partition into several clusters, we took into account the number of members in each cluster and the standard deviation of the variables in the cluster (as low as possible). We also made an inspection of the standard deviation between and within clusters, the respective Fisher ratio and their p-level of significance considered to be lower than 0.05.[44,45] All spectral moments (from μ_0 to μ_{15}) were used in both analysis; all variables show p-levels < 0.05 for Fisher test, the results are depicted in Table 1.

Table 1. Main results of the *k*-means cluster analysis for active and inactive compounds

	Variance analysis			
Spectral moments	Between SS[a]	Within SS[b]	Fisher ratio (F)	p-Level[c]
Statistics for active compound clusters (k-MCA 1)				
μ_2	30.11	24.26	75.93	**0.00**
μ_5	42.34	30.16	83.48	**0.00**
μ_1	33.35	11.35	55.29	**0.00**
μ_8	32.71	27.22	44.31	**0.00**
μ_6	91.71	31.18	99.80	**0.00**
Statistics for inactive compound clusters (k-MCA 2)				
μ_2	114.37	55.12	99.17	**0.00**
μ_5	62.76	41.61	51.39	**0.00**
μ_1	86.38	43.12	86.75	**0.00**
μ_8	74.13	53.71	49.33	**0.00**
μ_6	29.39	25.41	39.27	**0.00**

[a] Variability between groups.
[b] Variability within groups.
[c] Level of significance.

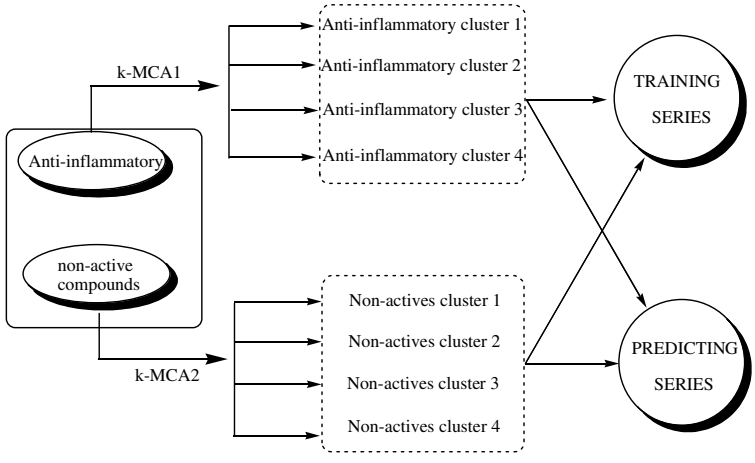

Figure 1. Training and Predicting series design throughout *k*-MCA.

The main conclusion should be achieved from k-MCA: the structural diversity of several up-to-date known anti-inflammatory (as codified by TOPS-MODE descriptors) may be described at least by four statistically homogeneous clusters of chemicals.

5. Development of the discriminant function

Once we perform a representative selection of training series it could be used to fit the discriminant function. The model selection was subjected to the principle of parsimony. Then we chose a function with high statistical significance but having few parameters (b_k) as possible.

In order to derive a discriminant function that permit the classification of chemicals as active (anti-inflammatory) or inactive (nonanti-inflammatory) we use the linear discriminant analysis in which spectral moments are used as independent variables. The classification model obtained is given below together with the statistical parameters of the LDA:

$$\begin{aligned} A. = {} & 1.92 \cdot \mu_0 - 2.47 \cdot \mu_1^{\mathrm{P}} - 7.32 \cdot \mu_2^{\mathrm{P}} + 3.63 \cdot \mu_3^{\mathrm{P}} \\ & + 2.66 \cdot \mu_4^{\mathrm{P}} - 1.94 \cdot \mu_5^{\mathrm{P}} + 0.36 \cdot \mu_6^{\mathrm{P}} \\ & - 0.01 \cdot \mu_8^{\mathrm{P}} + 4.41 \cdot 10^{-4} \mu_9^{\mathrm{P}} - 5.02 \\ N = 441 \quad U = 0.531 \quad F = 42.169 \quad & D^2 = 4.12 \end{aligned} \tag{2}$$

In this model the coefficient U is the Wilk's statistics (statistic for the overall discrimination is computed as the ratio of the determinant of the within-groups variance/covariance matrix over the determinant of the total variance covariance matrix), D^2 is the squared Mahalanobis distance where the D^2 in the square distance among the centroids of the populations and the centroid of the population is the centre of gravity of this population based in a set of variables and F is the Fisher ratio. The Wilk's U-statistics is the standard statistic that is used to denote the statistical significance of the discriminatory power of the current model.[46,47] Its value will range from 1.0 (no discriminatory power) to 0.0 (perfect discriminatory power). For the discrimination of active/inactive compounds studied here, the model classified correctly 88.28% of active and 91.05% of inactive compounds in the training series; for a global good classification of 90.24%. The percentages of false actives and false inactive compounds in the training series were 8.94% and 11.72% respectively and statistical outliers were not detected. The previous statement basis on two facts, all misclassified chemicals (accordingly to posterior probabilities and Mahalanobis's distance) does not rise to model improvement after leaving-out from it. Additionally, k-MCA demonstrates that any group of chemicals did not exists (possible outliers) that differentiate appreciable from the remnant ones. False actives are those inactive compounds that model classifies as actives, and the false inactive are those actives classified as inactive by the model. In Tables 2 and 3, the compounds classification using the above-given model is depicted.

Table 2. Classification and name of the active compounds in the training series

Name	P(H)	Name	P(H)
Active compounds			
21-Acetoxy-pregnenolone[a]	0.257	Halobetasol propionate	1.000
Aceclofenac	0.925	Halopredone acetate	0.999
Acemetacin	0.954	Ibufenac	0.754
acetamidocaproic acid	0.585	Ibuprofen	0.754
Acetaminosalol	0.902	Imidazole Salicylate	0.929
Alclometasone dipropionate	0.881	Indometacin	0.923
Algestone	0.938	Isofezolac	0.999
Amfenac	0.982	Isoflupredone	0.912
Ampiroxicam	0.984	Isonixin[a]	0.330
Amtolmetin Guacil	0.991	Isoxepac	0.747
Balsalazide	0.999	Isoxicam[a]	0.345
Beclometasone	1.000	Ketorolac	0.941
Bendazac	0.892	Lexipafant	0.995
Benorilate	0.943	Lonazolac	0.993
Benoxaprofen	0.847	Lornoxicam	0.940
Benzydamine	0.628	Loteprednol Etabonate	0.969
Bermoprofen	0.656	Mazipredone	0.969
Bromofenac	0.996	Meclofenamic acid	0.983
Bucloxic acid	0.890	Mefenamic acid	0.962
Budesonide	0.988	Meloxicam	0.941
Calcium Acetylsalicylate	1.000	Mesalamine	0.634
Carprofen	0.914	Mofezolac	0.967
Celecoxib	1.000	Mometasone furoate	1.000
Cinmetacin	0.970	Morazone	0.627
Clidanac[a]	0.337	Nabumetone	0.568
Clobetasone	0.999	Naproxen	0.829
Clopirac	0.760	Nimesulide[a]	0.376
Cortisone[a]	0.395	Olsalazine	0.997
Deflazacort	0.994	Oxaprozin	0.967
Desonide	0.990	Oxyphenbutazone[a]	0.214
Dexamethasone	0.980	Paranyline	0.996
Diclofenac sodium	0.916	Phenyl Acetylsalicylate	0.704
Diclofenac	0.916	Phenyl Salicylate	0.568
Difenamizole	0.954	Piketoprofen	0.883
Difenpiramide	0.978	Pipebuzone[a]	0.219
Diflucortolone	0.744	Pirazolac	0.999
Diflunisal	0.998	Pranoprofen	0.772
Difluprednate	0.999	Prednisone	0.539
Ditazol[a]	0.312	Proglumetacin	0.839
Droxicam	0.969	Propyphenazone[a]	0.350
Enoxolone	0.902	Proquazone[a]	0.351
Etersalate	0.910	Ramifenazone[a]	0.326
Etofenamate	0.851	Rimexolone	0.983
Felbinac	0.985	Salacetamide	0.697
Fenbufen	0.996	Salicylamide	0.878
Fendosal	1.000	Salicylamide	0.665
Fenoprofen	0.650	Salicylsulfuric acid[a]	0.396
Fentiazac	0.999	Salsalate	0.971
Fluazacort	0.999	Sulfasalazine	1.000
Flufenamic	0.992	Sulindac	0.991

Table 2 (*continued*)

Name	P(H)	Name	P(H)
Active compounds			
Flunixin	0.995	Suprofen	0.999
Flunoxaprofen	0.979	Talniflumate	0.999
Fluocinolone acetonide	0.999	Tenidap	0.753
Fluocinonide	1.000	Tenoxicam	0.995
Fluocortin butyl	0.658	Terofenamate	0.777
Fluperolone acetate	0.996	Thiazolinobutazone[a]	0.461
Fluprednidene acetate	0.998	Tixocortol	0.936
Fluprednisolone	0.798	Tolfenamic acid	0.963
Flurandrenolide	0.991	Tolmetin	0.996
Flurbiprofen	0.987	Triamcinolone hexacetonide	1.000
Gentisic acid	0.803	Tropesin	0.957
Glucametacin	0.996	Ximoprofen[a]	0.443
Guaiazulene	0.984	Zaltoprofen	0.983
Halcinonide	1.000	Zomepirac	0.936

P(H): The posterior probability that a case belongs to an anti-inflammatory group. It is basically proportional to the Mahalanobis distance from that group centroid.
[a] Misclassified compounds.

Table 3. Classification and names of the non-antiinflammatory compounds in the training series

Name	P(H)	Name	P(H)
Inactive compounds			
Acyclovir	0.106	Fenchlorazole	0.041
Adapin	0.288	Fenclorim	0.004
Ademetionine	0.245	Fenilpropanolamine	0.124
Adipex	0.042	Fenipentol	0.173
Adiphenine	0.025	Fenitrothion	0.116
Adrafinil	0.198	Fenobarbital	0.106
Adrenalone	0.458	Fenobucarb	0.108
Afloqualone	0.076	Fenoterol[a]	0.562
Afrazine	0.088	Fenoverine	0.220
Afungil	0.128	Fenoxaprop-P	0.347
Ag-307	0.261	Fenoxedil	0.003
Airbron	0.330	Fenozolone	0.032
Alacepril	0.188	Ganciclovir	0.005
Alachlor	0.161	Ganciclodivir	0.378
Alarmine	0.149	Gardenal	0.106
Albendazole	0.226	Gemfibrozil	0.171
Albuterol	0.089	Gentamicin	0.290
Alchloquin	0.077	Geraniol	0.018
Alcobon	0.204	Glafenine	0.399
Aldicarb sulfone	0.343	Glaziovine	0.079
Aldicarb	0.214	Glutethimide	0.052
BCPC	0.035	Glycarbylamide	0.075
Beclobrate	0.054	Glycidyl acrylate	0.202
Beflubutamid[a]	0.743	Glycidyl methacrylate	0.031
Befunolol[a]	0.692	Glycine	0.125
Benexate	0.354	Hexachlorophene	0.647
Benfotiamine[a]	0.832	Hexamethonium chloride	0.012
Benoxacor	0.004	Hexazinone	0.089
Benproperine	0.439	Hexestrol	0.177
Benserazide	0.152	Hexetidine	0.005
Bensulfuron	0.242	Hexobendine	0.001
BENT	0.020	Hexocyclium	0.371
Benzalkonium	0.044	Hexylresorcinol	0.265
Benzamide[a]	0.580	Hidroxiamfetamina	0.335
Benzamyl	0.129	Histamine	0.008
Benzene	0.057	Homatropine methylbromide	0.121
Benzethonium chloride	0.095	Homatropine	0.115
Benziodarone	0.172	Homofenazine	0.388
Benzipram	0.008	Imazalil	0.458
Benzocaine	0.136	Imidapril	0.153
Benzoctamine	0.425	Imipenem	0.155
Benzofenap	0.215	Imipramine	0.183
Benzoyl peroxide[a]	0.651	Imolamine	0.001
Benzoylprop	0.311	Improsulfan	0.000
Benzphetamine	0.008	IMPY	0.051
Benzthiazide	0.109	Indanazoline	0.013
Butropium bromide	0.008	Indanofan	0.021
Buturon	0.054	Indanorex	0.046
Butyl acetate	0.034	Indobufen[a]	0.530
Butylate	0.107	Kanamycin	0.133
BVDU	0.018	Kavain	0.194
Cacodylic acid	0.056	Kebuzone	0.267
Cafaminol	0.137	Ketamine	0.013
Cafedrine[a]	0.702	Ketobemidone	0.068
Cafenstrole	0.098	Lacidipine	0.047
Caffeine	0.378	Lamivudine	0.036
Camazepam	0.282	Lamotane-x	0.350
Camostat	0.008	Lamotrigine	0.313
Camylofin	0.002	Leptophos	0.485
Candesartan	0.458	Lethane	0.005
Carbamazepine	0.439	Mabuterol[a]	0.638
Carbaryl[a]	0.542	Malathion	0.052
Carbazochrome	0.058	MAMA	0.111
Carbenoxolone	0.024	Manidipinex	0.415
carbetamide	0.100	Mannitol	0.000
Carbimazole	0.223	Maprotiline	0.073
Carbinoxamine	0.007	Naftidrofuryl	0.003
Carbocromen	0.018	Nalbuphine	0.004
Carbocrundid	0.005	Naled	0.000
Carbodine	0.010	Nalidixic acid[a]	0.571
carbofuran	0.010	Nalmefene[a]	0.553
Carbophenothion[a]	0.577	Nalorphine	0.137
Carboquone	0.144	Oleandomycin	0.216
Carboxazole	0.037	Omnastine	0.054
Carbromal	0.161	Omoconazole	0.053
Carbromide	0.084	Opromacine	0.129
Carburazepam	0.030	Orazamide[a]	0.574
Cardiazol	0.002	Paroxetine	0.061
Carfentrazone[a]	0.590	Pathocidin	0.117
Carisoprodol	0.016	Patulin	0.006
Carmofur	0.077	Pecazine	0.020
Carmustine	0.003	Pefloxacin	0.052
Caroxazone	0.306	Pemoline	0.017
Carphenazine	0.366	Penbutolol	0.043
Carpipramine	0.338	Penciclovir	0.006
Carteolol	0.102	Penfluridol[a]	0.550
Carticaine[a]	0.552	Penicillin	0.090
CDEA	0.013	Pentachlorophenol	0.150
Cefacetrile[a]	0.559	Quinmerac	0.236
Cefamandole[a]	0.772	Quinoclamine	0.407
Dalactine	0.002	Quinupramine	0.139
Dalapon	0.047	Quizalofop	0.016
Daraprim	0.111	R8231[a]	0.549

(*continued on next page*)

Table 3 (*continued*)

Name	P(H)	Name	P(H)
Inactive compounds			
Daunomycin	0.459	Racefemine	0.219
Daunorubicin	0.429	Raclopride	0.209
Dde	0.038	Raloxifene	0.001
Ddt	0.003	Ramosetron	0.055
Delachlor	0.203	Ranimustine	0.556
Demeclocycline	0.137	Ranitidine	0.000
Demeton-*S*-methyl	0.003	Razoxane	0.026
Demoxepam	0.105	Regutensin	0.045
Deseril	0.069	remacemide	0.099
Desipramine	0.100	Remifentanil	0.359
Desmedipham	0.201	Remikiren	0.002
Desmethylmethalibur	0.105	Repirinast	0.002
Desmetryn	0.014	Rescinnamine	0.224
Dexambutol	0.000	Reserpine	0.006
D-Glucose	0.006	Resistomycin	0.006
Di-allate	0.062	Sertraline[a]	0.712
Diamide	0.086	Setastine	0.055
Diazepam	0.097	Sevoflurane	0.004
Diazinon	0.016	Sibutramine	0.324
Dibenamina	0.138	Siduron	0.001
Dibenyline	0.095	Sildenafil	0.278
Dibenzyline	0.075	Simazine	0.035
Dibromo-chloromethane	0.092	Simetryn	0.003
Dibromodichloromethane	0.000	Simfibrate	0.008
Dicamba	0.150	Simvastatin	0.104
Dicapthon	0.228	Sisomicin	0.002
Dicaptol	0.005	Sitafloxacin	0.170
Dichlobenil	0.094	SKF101468	0.038
Dichlormate	0.029	SKF89124	0.086
Dichloromethane	0.002	SL75102	0.177
Dichlorprop	0.139	Sobuzoxane	0.197
Dichlorvos	0.000	Sofalcone	0.108
Diclofop	0.371	Sorivudine	0.000
Diclosulam[a]	0.565	Sotalol	0.002
Didanosine	0.015	Sparfloxacin	0.077
Didrex	0.311	Spinulosin[a]	0.568
Diethamquat	0.156	Spiperone	0.001
Eflornithine	0.097	Tenuazonic acid[a]	0.820
Efudex	0.111	Tepraloxydim	0.026
Elmustine	0.001	Terazosin	0.020
Eltroxin	0.004	Terbinafine	0.000
Emylcamate	0.007	Terbuchlor	0.362
Enalapril	0.025	Terbumeton[a]	0.511
Enalaprilat	0.174	Terbutaline	0.002
Endosulfan	0.097	Terbuthylazine	0.036
Endothal	0.174	Terbutryn	0.001
Endoxan	0.012	Terconazole	0.005
Enidrel	0.100	Terodiline	0.031
Enoxacin	0.100	Tertatolol	0.236
Epinephrine	0.301	Testolactone	0.283
Epinephridine	0.345	Testosterone[a]	0.550
Epirubicin	0.055	Tetracaine	0.094
Epronaz	0.092	Tetrachloroethylene	0.005
Eptc	0.004	Tetrachloromethane	0.005
Erbon	0.005	Tetracycline	0.000
Famotidine	0.268	Valethamate	0.238
Febuprol	0.056	Valproic acid	0.001
Felodipine	0.010	Vernolate	0.142
Femerazo	0.050	Vidarabine	0.026
Fenalcomine	0.337	Vidarabidine	0.082
Fenamiphos	0.039	Vigabatrin[a]	0.544
Fenamole	0.024	Vindesine	0.288
Fenasulam[a]	0.754	VUFB-7904	0.063
Fenbutrazate	0.081		

P(H): The posterior probability that a case belongs to an anti-inflammatory group. It is basically proportional to the Mahalanobis distance from that group centroid.
[a] Misclassified compounds.

One of the most important criteria for the acceptance or not for a discriminant model, such as model,[2] is based on the statistics for the external prediction series.[42,48,49] Model 2 classified correctly 88.46% and 84.12% of active and inactive compounds in the prediction series respectively, which represents an overall predictability of 85.39%. In Table 4 we give the classification of compounds in the prediction series together with their difference between the Posteriori Probability Percentage of Classification in Active or Inactive Group.

Table 4. Classification and names of the compounds in external prediction series

Name	P(H)	Name	P(H)
Active compounds			
1-Naphthyl salicylate	0.948	Halometasone	0.961
Alclofenac[a]	0.457	Indoprofen	0.927
Aspirin	0.843	Ketoprofen	0.955
Benzpiperylon[a]	0.359	Lysine Acetylsalicylate	0.991
Bumadizone[a]	0.294	Metiazinic acid	0.963
Clobetasol	0.999	Niflumic Acid	0.991
Desoximetasone	0.565	Perisoxal	0.586
Diflorasone	0.991	Piroxicam	0.941
Enfenamic Acid	0.893	Protizinic acid	0.917
Fenclozic Acid	0.989	Salicylic acid	0.747
Flumetasone	0.991	Suxibuzone	0.552
Fluorometholone	0.995	Tiaprofenic acid	0.998
Fluticasone propionate	1.000	Xenbucin	0.855
Inactive compounds			
Adefovir	0.445	Fenobam	0.220
Aethoform	0.136	Fenoxazoline	0.020
Akineton[a]	0.582	Gefarnate	0.078
Albutoin	0.013	Glufosinate	0.310
Beclamide	0.150	Glycopyrronium bromide	0.018
Benomyl	0.083	Hexobarbital	0.388
Bentazone	0.027	Histapyrrodine	0.016
Benzilonium bromide	0.041	Iminodibenzyl	0.174
Benzofluor	0.006	Indalpine	0.352
Butroxydim	0.188	Karbutilate	0.096
Cadralazine	0.137	Lactophenin	0.218
Cambendichlor	0.000	Letosteine	0.388
Carbasulam	0.115	Manozodil	0.025
Carbocisteine	0.018	Nalline[a]	0.587
Carboplatin	0.004	Ondansetron[a]	0.656
Carbuterol[a]	0.579	Pebulate	0.055
Carnitine	0.369	Pendimethalin	0.006
Catechin[a]	0.597	Quinonamid[a]	0.594
Dalmane	0.009	Raffinose	0.001
Debrisoquine	0.068	Redimyl	0.076

Table 4 (*continued*)

Name	P(H)	Name	P(H)
Desipramine	0.100	Rescimetol[a]	0.656
D-Glucosamine	0.010	Sethoxydim	0.250
Diazoxide	0.144	Simeton	0.001
Dibucaine	0.035	Sitosterol[a]	0.552
Dichlormid	0.001	Sorbitol	0.066
Dicrotophos	0.000	Spiramycin[a]	0.619
Eglinazine	0.011	Terbufos	0.318
Encainide[a]	0.742	Terfenadine	0.110
Ephedrine	0.048	Tetracene	0.430
Equilin	0.071	Vecuronium	0.024
Fenadiazone	0.117	Vinyl chloride	0.001
Fencamfamin	0.019		

P(H): The posterior probability that a case belongs to an anti-inflammatory group. It is basically proportional to the Mahalanobis distance from that group centroid.
[a] Misclassified compounds.

6. Fragment contributions

As we previously explain, the TOPS-MODE approach is able to compute the contribution of any structural fragment (real or hypothetical) to the biological property or activity studied.[24–26] In the present case, we can find the positive and negative contributions of such fragments to the development of the anti-inflammatory activity. These fragments will be named here as active and inactive, respectively. The presence of active fragments does not presuppose the development of the anti-inflammatory activity per se, because it is well know that the activity is the consequence of the sum of contributions of all fragments in the molecule.[51–54]

In the Figure 2, we show the structure of a series of fragments selected from our database. The contributions of the anti-inflammatory activity of these fragments were compute by using the model 2. These quantitative contributions are given in Table 5.

If we analyze the series of fragments from $\mathbf{F_{18}}$ to $\mathbf{F_{20}}$ we can find an interesting behaviour. The three fragments increase the anti-inflammatory property, although at decreases the electronegativity of the substituent in *para* position decreases their contribution. Hashimoto et al.[55] show this behaviour in a family of 4-aryl-5-phenyloxazole compounds. In this series with the introduction of a fluorine atom on the 4-position of the phenyl group was showed high anti-inflammatory potency. Furthermore, this author demonstrated that an increase of the

Table 5. Contribution of some selected fragments to the anti-inflammatory activity

Fragment	Contribution	Fragment	Contribution	Fragment	Contribution
$\mathbf{F_1}$	0.173	$\mathbf{F_{14}}$	0.039	$\mathbf{F_{27}}$	0.095
$\mathbf{F_2}$	0.077	$\mathbf{F_{15}}$	−0.133	$\mathbf{F_{28}}$	0.118
$\mathbf{F_3}$	0.141	$\mathbf{F_{16}}$	0.022	$\mathbf{F_{29}}$	0.328
$\mathbf{F_4}$	0.066	$\mathbf{F_{17}}$	0.441	$\mathbf{F_{30}}$	0.321
$\mathbf{F_5}$	0.192	$\mathbf{F_{18}}$	0.104	$\mathbf{F_{31}}$	0.311
$\mathbf{F_6}$	−0.791	$\mathbf{F_{19}}$	0.232	$\mathbf{F_{32}}$	0.155
$\mathbf{F_7}$	0.124	$\mathbf{F_{20}}$	0.352	$\mathbf{F_{33}}$	0.249
$\mathbf{F_8}$	−0.130	$\mathbf{F_{21}}$	0.261	$\mathbf{F_{34}}$	0.304
$\mathbf{F_9}$	0.286	$\mathbf{F_{22}}$	0.218	$\mathbf{F_{35}}$	0.187
$\mathbf{F_{10}}$	−0.350	$\mathbf{F_{23}}$	0.217	$\mathbf{F_{36}}$	0.054
$\mathbf{F_{11}}$	0.144	$\mathbf{F_{24}}$	0.007	$\mathbf{F_{37}}$	−0.029
$\mathbf{F_{12}}$	−0.352	$\mathbf{F_{25}}$	0.029	$\mathbf{F_{38}}$	0.073
$\mathbf{F_{13}}$	0.189	$\mathbf{F_{26}}$	0.117	$\mathbf{F_{39}}$	0.391

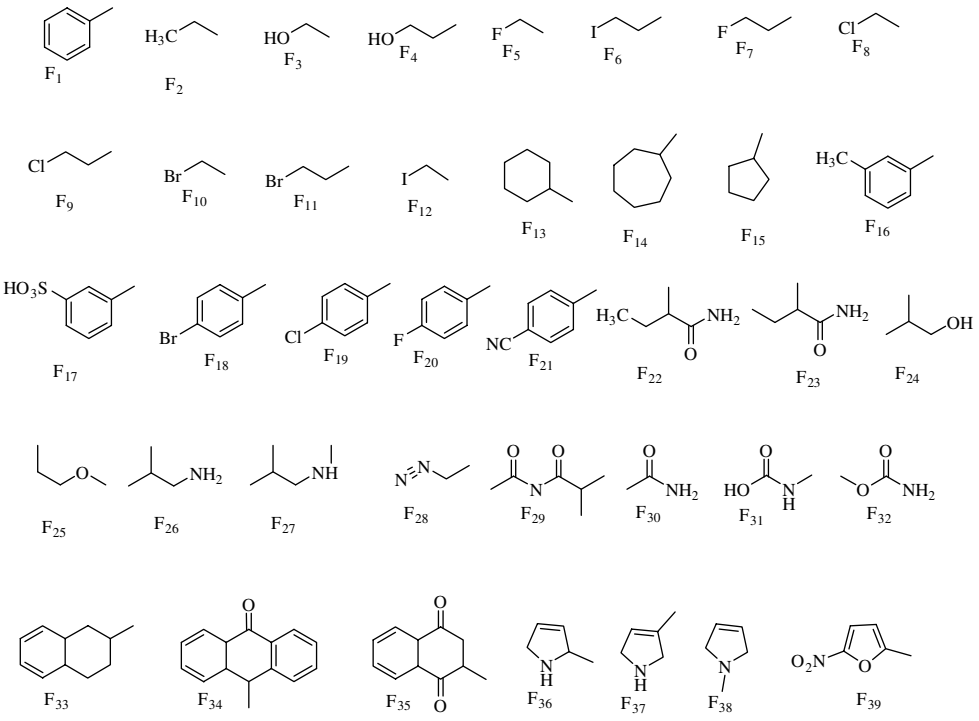

Figure 2. Structures of selected fragments for which their contributions to the anti-inflammatory activity are evaluated.

substituent size should be to negatively influence in this property, reason why show a perfectly fit to the contribution that realize this fragments to the property under study. Additionally, Hashimoto et al. describe as the cyclohexyl imparts as much activity as phenyl ring, whereas cycloheptyl is less active and cyclopentyl is nonactive. At this conclusion also can arrived if inspect the fragments $\mathbf{F_1}$, $\mathbf{F_{15}}$, $\mathbf{F_{13}}$ and $\mathbf{F_{14}}$ where the phenyl (0.173) and the cyclohexyl ring (0.187) introduce positive and similar contributions to the power anti-inflammatory and the cyclopentyl ring (−0.132) decrease the activity of the compound where this fragment appear. Therefore, the series of contributions of this rings as well establish Hashimoto should be defined as phenyl ≈ cyclohexyl > cycloheptyl > cyclopentyl.

On the other hand, when increase the aliphatic chain (ethyl (0.077) → propyl (0.145)) increase the activity of the fragment. This phenomenon was reported by Kalgutkar et al.[56] in a family of amide derivates of Meclofenamic acid inhibitors of COX. Where, an extension of the alkyl chain length in the alkylamine series significantly improved COX inhibitory potency.

In recent papers were highlight the importance of the halogens in the anti-inflammatory activity.[55–57] Kalgutkar et al.[56] establish that an incorporation of terminal halogens in the alkyl substituents improved the COX activity. Here, we compared the pairs of fragments $\mathbf{F_7}$, $\mathbf{F_2}$; $\mathbf{F_9}$, $\mathbf{F_2}$; $\mathbf{F_{11}}$, $\mathbf{F_2}$; and $\mathbf{F_6}$, $\mathbf{F_2}$ found that all of them in this position increase the activity, less the iodine atom. This phenomenon should be to the big size of this atom, the lead to decreases of the activity.[55] However, an opposite effect should found in the fragments $\mathbf{F_2}$ y $\mathbf{F_4}$, where the terminal hydroxyl group spread to decrease the contribution to the anti-inflammatory property of this fragment, behaviour that coincide with the Kalgulkar et al.[56]

Nevertheless, when we compare the fragments $\mathbf{F_3}$ and $\mathbf{F_4}$ found that the last one present an additional methyl group and their contribution and the activity is lower than $\mathbf{F_3}$. This behaviour is not logic if we take into consideration the previous conclusion that when the alkyl chain increase (ethyl → propyl) increase the fragment contribution. Apparently, a lengthening of the alkyl chain impede to make hydrogen bonding at hydroxyl group with potential residues of the enzyme active site such as His 90, Arg 120 and Tyr 355.[57] Another explanation should be that fragments of this type ($\mathbf{F_3}$) influence in the disruption of the hydrogen-bond between Arg 120, Arg 513 and Glu 524 contributes to the competitive binding inhibition.[58]

Finally, analyzing the contribution of fragments $\mathbf{F_{39}}$, it seems that the group nitro in position 5 of the furyl ring has a great positive effect on the anti-inflammatory activity. Therefore, a simple analysis of that contribution carry out to synthesized compounds with this type of sub-structure. However, the nitro group in this position is highly mutagenic. This type of sub-structure that is found in the 5-nitro-2-furylethylenes give positive to test de AMES and to test of induction of SOS,[59,60] therefore should be had special attention at the time of select the appropriate fragment for design the potentials candidates.

This last position recommends that computational models where combined are toxicological and pharmacology properties are necessary; with the objective of decreases the laboratory expense.

7. Concluding remarks

In spite of some criticism, there is an increase necessity of topological-indices-based QSAR models in order to rationalize the drug discovery process. In this sense, the TOPS-MODE approach has been extended not only to the discovery of novel leads but also to the study of the physicochemical and absorption properties of drugs.[49,50] On the other hand, QSAR make use of reduced or homologous series of compounds. Consequently, decays the model capability to predict the activity of different structural features. In the present paper the TOPS-MODE approach has been largely, probe to generate good predictive linear models in order to account for anti-inflammatory activity. Thence, we can assert that the TOPS-MODE approach may be used as an efficient alternative to massive screening of whether drugs.

Acknowledgements

The authors would like to express their gratitude to FAPESP (Brazil) and CNPq (Brazil) for financial support and kindness. Maykel Pérez thanks the owner of the software MODESLAB 1.0 for the donation of this valuable tool for the realization of this work.

Last but yet importantly, the authors would like to express they more sincerely gratitude to unknown referees and the editor Prof. Dr. Chi-Huey Wong for useful comments and kind attention.

References and notes

1. Allison, M. C.; Howatson, A. G.; Torrance, C. J.; Lee, F. D.; Russel, R. I. N. *Engl. J. Med.* **1992**, *327*, 749–754.
2. Smith, W. L.; Gravito, R. M.; Dewitt, D. L. *J. Biol. Chem.* **1996**, *271*, 33157–33160.
3. Ferreira, S.; Greene, L.; Alabaster, V.; Bakhle, Y.; Vane, J. *Nature* **1970**, *225*, 379–380.
4. Murray, M. D.; Brater, D. C. *Ann. Rev. Pharmacol. Toxicol.* **1993**, *32*, 435–465.
5. Vane, J. R.; Botting, R. M. *Scand. J. Rheumatol.* **1996**, *25*, 9–21.
6. Tannenbaum, H.; Davis, P.; Russell, A. S.; Atkinson, M. H.; Maksymowych, W.; Huang, S. H.; Bell, M.; Hawker, G. A.; Juby, A.; Vanner, S.; Sibley, J. *Can. Med. Assoc. J.* **1996**, *155*, 77–88.
7. Taha, A. S. *New Engl. J. Med.* **1996**, *334*, 1435–1439.
8. Cuzzolin, L.; Conforti, A.; Adami, A.; Lussignoli, S.; Menestrina, F.; Del Soldato, P.; Benoni, G. *Pharmacol. Res.* **1995**, *31*, 61–65.

9. DeMeo, M. *Nutr. Rev.* **1995**, *53*, 13–16.
10. Delmas, P. D. *Br. J. Rheumatol.* **1995**, *34*, 25–28.
11. Harris, R. C.; McKanna, J. A.; Akai, Y.; Jacobson, H. R.; Dubois, R. N.; Breyer, M. D. *J. Clin. Invest.* **1994**, *94*, 2504–2510.
12. Kaufmann, W. E.; Worley, P. F.; Pegg, J.; Bremmer, P.; Isakson, P. C. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 2317–2321.
13. Beiche, F.; Scheuerer, S.; Brune, K.; Geisslinger, G.; Goppeltstruebe, M. *FEBS Lett.* **1996**, *390*, 165–169.
14. Zimmermann, K. C.; Sarbia, M.; Schror, K.; Weber, A. A. *Mol. Pharmacol.* **1998**, *54*, 536–540.
15. Papp, M.; Kase, H.; Gigler, G. *Acta Physiol. Hung.* **1992**, *80*, 423–425.
16. Schrier, D. J.; Lesch, M. E.; Wright, C. D.; Gilbertsen, R. B. *J. Immunol.* **1990**, *14*, 1874–1875.
17. Vigorita, M. G.; Grasso, S.; Zappala, M.; Ottana, R.; Monforte, M. T.; Barbera, R.; Trovato, A. *Farmaco* **1994**, *49*, 271–276.
18. Noguchi, T.; Shimazawa, R.; Nagasawa, K.; Hashimoto, Y. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 1043–1046.
19. Kumar, P.; Knaus, E. E. *Drug Des. Discov.* **1994**, *11*, 15–22.
20. Wilkerson, W. W.; Galbraith, W.; Gans-Brangs, K.; Grubb, M.; Hewes, W. E.; Jaffee, B.; Kenney, J. P.; Kerr, J.; Wong, N. *J. Med. Chem.* **1994**, *37*, 988–998.
21. Steven, S.; Wesolowski, L.; William, L. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 267–270.
22. Unangst, P. C.; Connor, D. T.; Cetenko, W. A.; Sorenson, R. J.; Kostlan, C. R.; Sircar, J. C.; Wright, C. D.; Schrier, D. J. *J. Med. Chem.* **1994**, *37*, 322–334.
23. Kubinyi, H. *QSAR: Hansch Analysis and Related Approaches*; VCH: New York, NY, 1993.
24. Estrada, E. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 844–849.
25. Estrada, E. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 320–328.
26. Estrada, E. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 23–27.
27. Estrada, E. *J. Chem. Soc., Faraday Trans.* **1998**, *94*, 1407–1411.
28. Estrada, E.; Peña, A.; García-Domenech, R. *J. Comput. Aided Mol. Des.* **1998**, *12*, 583–595.
29. Estrada, E.; Gutiérrez, Y. *J. Chromatogr. A* **1999**, *858*, 187–199.
30. Estrada, E.; Gutiérrez, Y.; González, H. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 1386–1399.
31. González, M. P.; González, H. D.; Molina, R. R.; Cabrera, M. A.; Ramos, R. A. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1192–1199.
32. Estrada, E.; Uriarte, E.; Montero, A.; Tejeira, M.; Santana, L.; De Clercq, E. *J. Med. Chem.* **2000**, *43*, 1975–1985.
33. Estrada, E.; Molina, E. QSPR/QSAR by Graph Theoretical Descriptors Beyond the Frontiers. In *QSAR/QSPR Studies by Molecular Descriptors*; Diudea, M., Ed.; Nova Science: New York, 2000; pp 83–107.
34. Estrada, E.; Uriarte, E. *SAR QSAR Environ. Res.* **2001**, *12*, 309–324.
35. Estrada, E. et al. *SAR QSAR Environ. Res* **2003**, *14*, 145–163.
36. Estrada, E.; Uriarte, E. *Curr. Med. Chem.* **2001**, *8*, 1573–1588.
37. Estrada, E. *Edge J. Chem. Inf. Comput. Sci.* **1995**, *35*, 31–33.
38. Gutierrez, Y.; Estrada, J. Modes Lab®, 2002, version 1.0 b.
39. Negwer, N. *Organic-Chemical Drugs and their Synonyms*; Akademie: Berlin, 1987.
40. Fu, J. Y.; Masferrer, J. L.; Seibert, K.; Raz, A.; Needleman, P. *J. Biol. Chem.* **1990**, *265*, 16737–16740.
41. Statsoft, Inc. STATISTICA 6.0, 2002, version 6.0.
42. Waterbeemd, H. Van. Ed.; Discriminant Analysis for Activity Prediction, In *Method and Principles in Medicinal Chemistry*; Manhnhold, R., Krogsgaard-Larsen, Timmerman, V., Eds.; VCH: Weinhiem, 1995; Vol. 2, pp 265–282.
43. Kowalski, R. B.; Wold, S. Pattern Recognition in Chemistry. In *Handbook of Statistics*; Krishnaiah, P. R., Kanal, L. N., Eds.; North Holland Publishing Company: Amsterdam, 1982; pp 673–697.
44. Mc Farland, J. W.; Gans, D. J. Cluster Significance Analysis. In *Method and Principles in Medicinal Chemistry*; Manhnhold, R., Krogsgaard-Larsen, Timmerman, V., Eds.; Van Waterbeemd, H., Eds. VCH: Weinhiem, 1995; Vol. 2, pp 295–307.
45. Johnson, R. A.; Wichern, D. W. *Applied Multivariate Statistical Analysis*; Prentice-Hall: NJ, 1988.
46. González, H.; Olazabal, E.; Castañedo, N.; Hernandez, I.; Morales, A.; Serrano, H. S.; González, J.; Ramos, R. *J. Mol. Mod.* **2002**, *1*, 83–95.
47. Franke, R. *Theoretical Drug Design Methods*; Elsevier: Amsterdam, 1984.
48. García-Domenech, R.; Gálvez, J.; Moliner, R.; García-March, F. *Drug Invest.* **1991**, *3*, 344–350.
49. González, M. P.; Morales, A. H.; González, D. H. *Polymer* **2004**, *45*, 2073–2079.
50. Cabrera, M. A.; Bermejo, M.; González, M. P.; Ramos, R. *J. Pharm. Sci.* **2004**, *93*, 1701–1717.
51. Kier, L. B.; Hall, L. H. *Eur. J. Med. Chem. Chim. Ther.* **1981**, *16*, 399–407.
52. Gálves, J.; Gracía-Domenech, R.; Julián-Ortiz, J. V.; Soler, R. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 272–284.
53. Randi, M. *J Comput. Chem.* **1993**, *4*, 363–370.
54. Estrada, E.; González, H. D. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 75–84.
55. Hashimoto, H.; Maeda, K.; Ozawa, K.; Haruta, J.; Wakitani, K. *Bioorg. Med. Chem.* **2002**, *12*, 65–68.
56. Kalgutkar, A. S.; Rowlinson, S. W.; Crews, B. C.; Marnett, L. *J. Bioorg. Med. Chem.* **2002**, *12*, 521–524.
57. Park, C. A.; Siomboing, X.; Yous, S.; Gressier, B.; Luyckx, M.; Chavatte, P. *Eur. J. Med. Chem.* **2002**, *37*, 461–468.
58. Pouplana, R.; Lozano, J. J.; Perez, C.; Ruiz, J. *J. Comput. Aided Mol. Des.* **2002**, *16*, 683–709.
59. Sturdik, E.; Rosenberg, M.; Stibranyi, L.; Balaz, S.; Chreño, O.; Ebringer, L.; Ilavsky, D.; Vegh, D. *Chem. Biol. Interact.* **1985**, *53*, 145–153.
60. Sturdik, E.; Benova, M.; Miertus, S.; Balaz, S.; Rosenberg, M.; Sturdikova, M.; Ebringer, L.; Stibranyi, L.; Ilavsky, D.; Vegh, D. *Chem. Biol. Interact.* **1986**, *54*, 69–78.