JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

Modeling Drug-Induced Anorexia by Molecular Topology

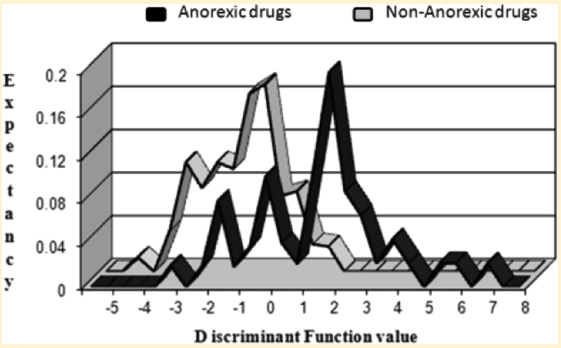
María Gálvez-Llompart,[†] Jorge Gálvez,[†] Ramón García-Domenech,[†,*] and Lemont B. Kier[‡]

[†]Molecular Connectivity and Drug Design Research Unit, Faculty of Pharmacy, Department of Physical Chemistry, University of Valencia Avd, V.A. Estellés, s/n 46100-Burjassot, Valencia, Spain

[‡]Center for the Study of Biological Complexity, Virginia Commonwealth University, P.O. Box 842030, Richmond, Virginia 23284, United States

Supporting Information

ABSTRACT: Molecular topology (MT) has demonstrated to be a very good technique for describing molecular structures and to predict physical, chemical, and biological properties of compounds. In this paper, a topological−mathematical model based on MT has been developed for identifying drug compounds showing anorexia as a side effect. An external validation (*test set*) has been carried out, yielding over an 80% correct classification in the active and inactive compounds. These results reinforce the role of MT as a potential useful tool for predicting drug side effects.



INTRODUCTION

A *side effect*, a term which is nowadays changing in favor of *adverse drug event* (ADE), could be define as "response to a drug which is noxious and unintended and which occurs at doses normally used in man for prophylaxis, diagnosis, or therapy of disease or for the modification of physiologic function".[1] ADE cause substantial morbidity and mortality, yet they remain underappreciated and misunderstood.[2]

Since 1969, 75 drugs and combination drugs products have been removed from the United States market for safety reasons. This represents less than 1% of all marketed drugs.[1] However, safety-related regulatory actions (e.g., labeling changes, such as the addition of precautions, contraindications, or black box warnings) are far more common, although less publicized. From 1969 to 2002, the Adverse Event Reporting System (AERS) of the Food and Drug Administration (FDA) received approximately 2.3 million reports of adverse events on more than 6000 drug products.[1]

Over 770,000 people are injured or die each year in hospitals from ADEs, which may cost up to $5.6 million each year per hospital depending on hospital size. An estimation of national hospital expenses to treat patients who suffer ADEs during hospitalization are between $1.56 and $5.6 billion annually.[3]

Anticipating the likely side effect profile of drugs and anticipating which drug will most likely suffer from a determinate side-effect is an aspect of key importance in current drug discovery, development, and marketing. This strategy will save millions dollars in health care.

One side effect having big repercussions in national health care is anorexia, defined as a loss of the desire to eat or a loss of appetite. Although it is often a symptom of an underlying disorder such as cancer, dementia, or depression, anorexia can be caused by exogenous sources, including medications as an undesirable effect.[4]

There are different types of drugs causing anorexia as a part of their side effects, such as anticonvulsivants (Topiramate, Zonisamide...), antiinfectives (Zidovudine, Ribavirin), psychotropics (Bupropion, Fluoxetine, Paroxetine...), stimulants (Methylphenidate, Phentermine...) and many other drugs (Etoposide, Lisinopril...).[4]

Several mechanisms of drug-induced anorexia have been determined in literature. Some of them include the increased release of hypothalamic neurotransmitters (norepinephrine and dopamine) inhibiting the reuptake of dopamine and serotonin, alteration in serum leptin concentrations leading to early satiety, and increased lipoprotein lipase activity in adipose tissue.[4]

Signs and symptoms typically associated with drug-induced anorexia are weight loss, vomiting, nutritional deficiencies, dental caries, nausea, loss of appetite, feeling of fullness with meals, difficulty swallowing, fatigue, loss of taste, and dry mouth.[4]

Anorexia is the most common symptom of cancer patients.[5,6] It is present in 15−25% of patients with cancer and is almost universal in patients with metastatic disease, occurring in more than 80% of cases.[7,8]

Unfortunately, weight loss has a marked significant impact on the quality of life of the cancer patients, including diminished response to chemotherapy,[9] radiotherapy, and surgery[10−12] and development of toxic effects of treatment[13,14] with morbidity

Received: December 19, 2011
Published: April 19, 2012

dx.doi.org/10.1021/ci2006092 | *J. Chem. Inf. Model.* 2012, 52, 1337−1344

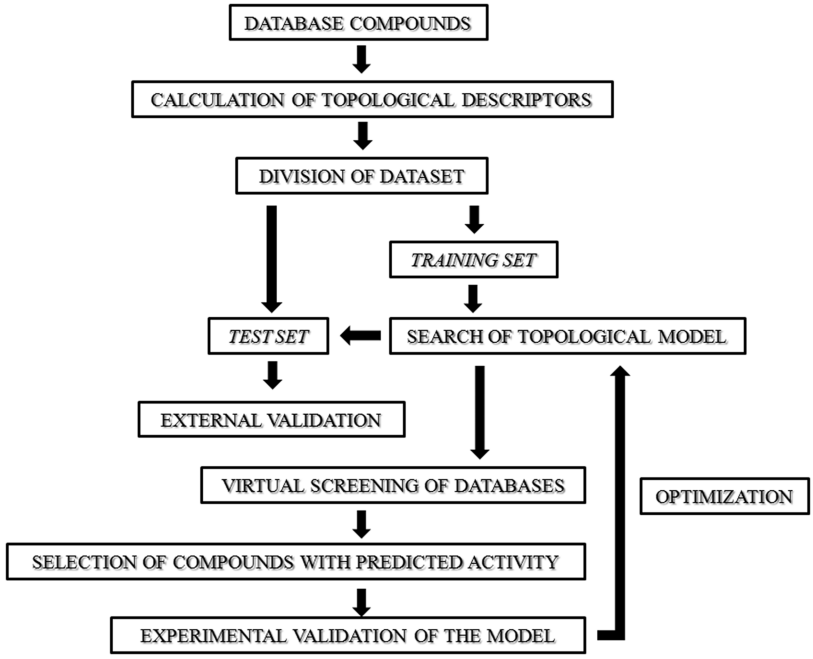


Figure 1. Scheme of drug research through molecular topology by virtual screening on databases.

and mortality in more than 20% of cancer patients, along with being the major cause of death.[15]

Molecular topology (MT) has proved to be a very useful technique to encode molecular structure. Its scope is the topological characterization of molecules by means of numerical invariants, called topological indices (TIs), which are the descriptive variables of the molecular topological models.

These descriptors are able to characterize the most important features of molecular structure: molecular size, binding, cycles, and branching. The computation of topological indices is very swift, and they have also the advantage of being true structural invariants. This means that topological indices are independent of the spatial position of the atoms in the molecule. However, extensions of the TIs that give account of the three-dimensional structure have been also devised.[16,17]

Topological indices have been used for the prediction of physicochemical and biological properties for groups of compounds showing considerable structural diversity.[18−22] Several pharmacological activities can be pointed out among the properties modeled, such as anticonvulsants,[23] antimalarials,[24,25] antimicrobials,[26,27] antifungals,[28] antineoplastics,[29] antihistaminics,[30] bronchodilator agents,[31,32] cytostatics,[33] and also anti-inflammatory compounds.[34,35]

Whatever the field, molecular topology's main advantage is clear: Predicting with confidence specific activities or properties of molecules and thereby selecting or designing new ones, saving time and money.

The present work deals with the modeling of drugs suffering anorexia as a side effect. This is the first time that MT is applied in the field of adverse drug events (ADE).

MATERIAL AND METHODS

The application of the MT model involves the following steps (Figure 1):

- Selection of data set from the literature. These data comprise all compounds (drug-induced anorexia and non) used for building the model.
- Calculation of topological descriptors using Dragon software.[36]
- Splitting of the data set into two groups: *training set* and *test set*. The criteria applied for such a division was based on the structural heterogeneity of compounds.
- Application of linear discriminant analysis (LDA) to the *training set*.
- Validation of the LDA through an external test (*test set*).
- Application of the topological model to the identification of compounds without anorexia as a side effect.

Selection of Data Set and Indices Calculation. A total of 177 compounds showing anorexia as a side effect and not showing anorexia as a side effect were selected to build the data set. Compounds used here were from the side effect database named SIDER[37] (available at http://sideeffects.embl.de/) and from the literature. The active group consisted of compounds showing anorexia as a side effect with at least 8% incidents of anorexia after assumption and anorexic compounds. The inactive group consisted of compounds showing anorexia as a side effect with an incident rate less than 8% and drugs not causing anorexia.

Then, the data set was split into two, namely, *training* and *test sets*. The former was of 45 compounds causing anorexia as ADE (active set) and 77 compounds not showing such a side effect (inactive set). The latter was made of 20 compounds causing anorexia as ADE (active set) and 35 compounds not showing such a side effect (inactive set).

The chemical structure of each drug was depicted by using ChemBioDraw Ultra version 12.0 (CambridgeSoft Corporation, Cambridge, MA).

Each compound was characterized by a set of 438 topological indices obtained with the Dragon software, version 5.4.

Among the TIs and topological descriptors, walk and path counts indexes, edge adjacency indices, spectral moment

Table 1. Results of Prediction of Compounds Showing Anorexia as a Side Effect Obtained by Linear Discriminant Analysis for the *Training Set*[a]

Compound	Prob (*A*)	DF	Class	Compound	Prob (*A*)	DF	Class
			Active group				
Acecainide[47]	0.451	−0.198	I	Irinotecan	0.887	2.063	A
Atovaquone	0.250	−1.097	I	Lamivudine	0.823	1.534	A
Bortezomib	0.600	0.403	A	Lenalidomide[53]	0.606	0.430	A
Busulfan	0.854	1.762	A	Lisinopril	0.857	1.790	A
Carmustine[48]	0.922	2.468	A	Methylphenidate	0.511	0.043	A
Cerulenin[49]	0.635	0.553	A	Nabilone	0.983	4.055	A
Cholecystokinin[50]	0.989	3.481	A	Nelarabine	0.881	2.005	A
Cidofovir	0.923	2.488	A	Orlistat[54]	0.998	5.995	A
Cysteamine	0.992	4.877	A	Pamidronate	0.989	4.472	A
Cytarabine	0.965	3.308	A	Pemetrexed[55]	0.252	−1.090	I
Daunorubicin	0.885	2.034	A	Pentostatin	0.908	2.292	A
Desvenlafaxine[51]	0.931	2.606	A	Phentermine[56]	0.959	3.141	A
Donepezil	0.493	−0.028	I	Ribavirin	0.976	3.704	A
Doxorubicin	0.893	2.109	A	Risperidone	0.900	2.194	A
Emtricitabine	0.773	1.224	A	Seliciclib[57]	0.206	−1.347	I
EGCG[49]	0.231	−1.209	I	Sibutramine[54]	0.875	1.949	A
Etoposide	0.912	2.328	A	Sorafenib[58]	0.272	−0.985	I
Felbamate	0.174	−1.561	I	Sunitinib[59]	0.635	0.555	A
Fentanyl[52]	0.511	0.045	A	Thioacetazone	0.054	−2.867	I
Flolan	0.944	2.825	A	Tobramycin	0.999	7.217	A
Fluoxetine	0.525	0.099	A	Topiramate[60]	0.998	6.386	A
Fluvoxamine	0.934	2.654	A	Vinorelbine	0.945	2.846	A
Galantamine	0.874	1.935	A				
			Inactive group				
Alimta	0.252	−1.090	I	Midodrine	0.260	−1.047	I
Alprazolam	0.073	−2.542	I	Mirtazapine	0.439	−0.245	I
Altretamine	0.425	−0.301	I	Mitoxantrone	0.661	0.664	A
Amifostine	0.849	1.730	A	Mycophenolic acid	0.133	−1.877	I
Amiodarone	0.549	0.198	A	Nabumetone	0.047	−3.003	I
Amprenavir	0.230	−1.209	I	Naproxen	0.167	−1.605	I
Anagrelide	0.272	−0.986	I	Naratriptan	0.840	1.659	A
Argatroban	0.601	0.408	A	Nevirapine	0.270	−0.994	I
Asa	0.014	−4.222	I	Nicardipine	0.195	−1.422	I
Atenolol	0.391	−0.443	I	Nicotine	0.409	−0.367	I
Baclofen	0.202	−1.373	I	Nitricoxide	0.466	−0.134	I
Bexarotene	0.410	−0.363	I	Nitroglicerin	0.804	1.410	A
Celecoxib	0.190	−1.447	I	Norfloxacin	0.686	0.780	A
Ciprofloxacin	0.722	0.954	A	Olanzapine	0.370	−0.534	I
Clomipramine	0.087	−2.356	I	Omeprazole	0.405	−0.385	I
Clonazepam	0.051	−2.921	I	Ondansetron	0.215	−1.297	I
Cyclosporine	0.488	−0.062	I	Oxcarbazepine	0.089	−2.329	I
Desloratadine	0.136	−1.846	I	Oxybutynin	0.706	0.874	A
Dofetilide	0.290	−0.896	I	Pindolol	0.066	−2.644	I
Doxazosin	0.392	−0.442	I	Posaconazole	0.293	−0.883	I
Entacapone	0.382	−0.480	I	Propafenone	0.365	−0.555	I
Estazolam	0.046	−3.032	I	Propofol	0.314	−0.781	I
Fenofibrate	0.533	0.132	A	Ritonavir	0.097	−2.231	I
Flurbiprofen	0.107	−2.120	I	Selegiline	0.363	−0.562	I
Gadoversetamide	0.626	0.512	A	Sertralinehcl	0.533	0.133	A
Hycamtin	0.767	1.190	A	Sotalol	0.385	−0.468	I
Indapamide	0.173	−1.562	I	Sulfasalazine	0.075	−2.520	I
Invirase	0.355	−0.598	I	Tacrine Hcl	0.084	−2.391	I
Iopromide	0.586	0.342	A	Terbinafine	0.432	−0.274	I
Isradipine	0.415	−0.343	I	Theophylline	0.307	−0.814	I
Itraconazole	0.282	−0.936	I	Tizanidine	0.073	−2.540	I
Lamotrigine	0.054	−2.855	I	Tracleer	0.124	−1.954	I
Lansoprazole	0.440	−0.240	I	Valdecoxib	0.126	−1.934	I
Leflunomide	0.713	0.911	A	Valsartan	0.504	0.016	A
Loratadine	0.295	−0.872	I	Verteporfin	0.147	−1.761	I

dx.doi.org/10.1021/ci2006092 | J. Chem. Inf. Model. 2012, 52, 1337−1344

Table 1. continued

Compound	Prob (A)	DF	Class	Compound	Prob (A)	DF	Class
Lorazepam	0.071	−2.567	I	Zafirlukast	0.241	−1.152	I
Meloxicam	0.038	−3.238	I	Zaleplon	0.168	−1.601	I
Mesalamine	0.117	−2.022	I	Zolpidem	0.053	−2.886	I
Metaproterenol	0.306	−0.818	I				

[a]Prob (A) = probability of activity. DF = discriminant function. Class = classification as active or inactive.

indices, and other graph-theoretical descriptors are not outlined here, as they were not selected for the final model.

Linear Discriminant Analysis. Linear discriminant analysis,[38,39] is a statistical tool providing a classification model based on the combination of variables that best predicts the category or group to which a given compound belongs. In our case, the *training set* compounds were allocated to active or inactive groups according to their capability of producing anorexia as a side effect, respectively. Hence, the discriminatory property was the capability of producing anorexia as a side effect and the independent variables were the TIs.

In order to compensate for the difference in the number of molecules conforming the *training set*, more weight was assigned to the group of active (0.6) and less (0.4) to the inactive group.

The discriminatory capability was assessed as the percentage of correct classifications in each set of compounds. The classification criterion was the Mahalanobis minimal distance (distance of each case to the mean of all the cases in a category).[40] The quality of the discriminant function was evaluated by using the Wilks parameter,[41] λ, which was obtained by multivariate analysis of variance, that tests what the equality of the group means for the variable in the model.

The method used to select the descriptors was based on the Fisher−Snedecor parameter (F),[42,43] which determines the relative importance of candidate variables. The variables used to compute the linear classification function are chosen in a stepwise manner: at each step, the variable that makes the largest contribution to the separation of the groups is entered into the discriminant equation (or the variable that makes the smallest contribution is removed).

Another important parameter that usually provides a balanced evaluation of the models prediction is the Matthews correlation coefficient (MCC).[44] This coefficient is based on the fact that in any prediction process there can be four different possibilities to account for:

TP: True positive, a drug-induced anorexia correctly classified or predicted.

FP: False positive, a drug not inducing anorexia predicted as induced or when there was none to predict.

TN: True negative, a drug not inducing anorexia correctly classified.

FN: False negative, a drug-induced anorexia predicted as not inducing or when there was none to predict.

It is clear, therefore, that any single number that represents the predictive power of the method must account for all of the possibilities listed above. One such factor is the Matthews correlation coefficient, given by

$$\mathrm{MCC} = \frac{(\mathrm{TP} \times \mathrm{TN}) - (\mathrm{FP} \times \mathrm{FN})}{\sqrt{(\mathrm{TN} + \mathrm{FN}) \times (\mathrm{TN} + \mathrm{FP}) \times (\mathrm{TP} + \mathrm{FP}) \times (\mathrm{TP} + \mathrm{FN})}} \quad (1)$$

The Matthews correlation coefficient ranges from $-1 \leq \mathrm{MCC} \leq 1$. A value of MCC = 1 indicates the best possible prediction, where every drug-induced anorexia was correctly predicted, and only true drug-induced anorexia was predicted. A MCC = −1 indicates the worst possible prediction (or anticorrelation), where no drug-induced anorexia was correctly predicted, and the largest number of incorrect drugs inducing anorexia was predicted. Finally, a Matthews correlation coefficient of MCC = 0 is what would be expected for a random prediction scheme.

The predictive power of the model was also assessed by using internal and external validation *(test set)*. A cross-validation, as an internal validation test, was carried out by taking randomly 15% of active compounds and 20% of inactive compounds from the *training set* and *test set* and changing their roles to see if the model continued to have good prediction of the remaining compounds in the *training* and *test sets*.

The equation obtained for each of the *training sets*, including the same descriptors, was used to predict the corresponding values of the *test set*.

The software employed to perform the LDA was the StatSoft version 9.0.[45]

Arrangement of the Pharmacological Distribution Diagram. The corresponding distribution diagram of compounds causing anorexia as a sideeffect, PDD,[46] was drawn from the outcome of the discriminant function. This diagram is arranged to establish the intervals of the discriminant function in which the expectancy (*E*) of finding compounds causing anorexia as a side effect is maximum. PDDs are histogram-like plots of connectivity functions in which expectancies appear on the ordinate axis. For an arbitrary interval of values of a given function, we define the expectancy of activity as $Ea = a/(i + 1)$; where a is the number of active compounds in the interval divided by the total number of active compounds, and i is the number of inactive compounds in the interval divided by the total number of inactive compounds. The expectancy of inactivity is defined in a similar way as $Ei = i/(a + 1)$. This plot provides a good visualization of minimum overlap regions and allows the selection of frames in which the probability of finding active compounds is at a maximum.

RESULTS AND DISCUSSION

In order to obtain the discriminant function (DF), LDA was applied to a *training set* comprised of 122 compounds distributed in two subsets: active (45 compounds causing anorexia as a side effect) and inactive (77 compounds not exhibiting this side effect, or at least not described as such in the literature) (Table 1).

The discriminant function selected was

$$\begin{aligned} \mathrm{DF} = {} & -0.134 + 0.0003 \times T(\mathrm{O..O}) - 24.228 \times \mathrm{piPC02} \\ & + 7.066 \times \mathrm{EEig02x} + 1.794 \times \mathrm{EEig12x} \\ & + 15.272 \times \mathrm{ESpm02u} \end{aligned}$$

$$N = 122 \qquad \lambda = 0.604 \qquad F = 15.2 \qquad p < 0.00001 \quad (2)$$

Where DF = discriminant function; T(O..O) = sum of topological distances between pairs of oxygen atoms in the molecule; piPC02 = molecular multiple path count of order 2;EEig02x, EEig12x = eigenvalues 02 and 12 from edge adjacency matrix weighted by edge degrees; ESpm02u = spectral moment 02 from edge adjacency matrix; N = number of data compounds; λ = Wilks' lambda; F = Fisher−Snedecor parameter; and p = statistic significance.

From this equation, a given compound will be selected as *active* or as a potential producer of anorexia if DF > 0, otherwise it is classified as *inactive*, i.e., not producing anorexia. The classification matrix for DF is very significant for the *training set*: 80% of correct prediction for the active group in DF, i.e., 36 out of 45 compounds, were correctly classified in the active set, whereas 61 out of 77 were correct (79%) in the inactive group (Table 1).

As an additional way of measuring the model's prediction, the Matthews'coefficient, which returns a value between −1 and +1, has been calculated. The higher its value the more reliable is the model. However, we calculated the Matthews' coefficient in a slightly different way, i.e., we just added +1 to each threshold value, so that the outcome is expressed as % accuracy. In other words, 0 would mean no correlation at all, 1 represents 50%, and 2 accounts for 100% correlation. By doing so, the output resulted to be 80% of overall accuracy (modified Matthews' coefficient = 1.6), which fits well with the results from the discriminant analysis (Table 2).

Table 2. Classification Matrix Obtained with the Selected Discriminant Function (DF) and Internal Cross-Validation Analysis

		Training Set		*Test Set*		Total
DF	λ (Wilks)	Active (%)	Inactive (%)	Active (%)	Inactive (%)	%
Model selected	**0.604**	**80.0**	**79.0**	**80.0**	**83.0**	**80.5**
CV1	0.598	80.0	76.6	80.0	83.3	80.0
CV2	0.581	80.0	81.8	80.0	80.0	80.5
CV3	0.603	82.0	81.8	75.0	77.1	79.0
CV4	0.569	82.2	83.1	60.0	82.9	77.1
CV5	0.596	82.2	79.2	75.0	74.3	77.7
CV average	**0.589**	**81.3**	**80.5**	**74.0**	**79.5**	**78.8**

An internal cross-validation analysis was also carried out to check the quality of the selected discriminant function. Table 2 shows the values of λ (Wilks' lambda) and the classification matrix for compounds in the *training* and *test sets*. The variability of λ is little for each subset, and the average, after five trials following the procedure described in Materials and Methods (λ = 0.589), is similar to that obtained with the model selected (λ = 0.604). Moreover, the average percentage of correct classification is similar to that obtained with the model selected (78.8 and 80.5%, *training* and *test set*, respectively).

Moreover, an external validation of the model *(test set)* is necessary to check the performance of the equation. To achieve it, a set of 55 compounds outside the *training set*, 20 compounds known to produce anorexia, and 35 compounds not producing it or with a minor incidence than 8%, were selected to carry out the *test set*. The DF equation was used to classify every compound in the test. As shown in Table 2, the percentage of correct classification was 80% and 83% within the active and inactive sets, respectively. Table 3 outlines the results of the prediction obtained for every compound of the external test.

In the DF function, it appears several topological descriptors, one of them, namely, T(O..O), evaluates the position of the oxygen atoms in the molecule. Specifically, it is defined as the sum of topological distances between pairs of oxygen in the graph. The rest of the indices are related to pure structural features; for instance, piPC02 is the molecular multiple path count of order 2 and is related to molecular branching and size or in general to the graph's molecular complexity. When the molecule is bigger and its elemental composition is more complex, this descriptor decreases.

EEig02x and EEig12x represent the eigenvalue 02 and 12 from edge adjacency matrix weighted by edge degrees. In general, both indices contribute in a positive way to the equation and they are sensing the presence of condensed alicyclic moieties in the molecules.

ESpm02u (the spectral moment 02 from edge adj. matrix) takes into account the presence of molecular fragments of topological length 2. In general, the spectral moments are linear combinations of a series of structural fragments in the molecular graph, but not all of these fragments have a direct influence on the physical property. However, the spectral moments of edge adjacency matrix are very easy to calculate, and they can be combined in appropriate ways in order to produce very good correlations without loss of structural information. The decomposition of spectral moments of the E matrix into substructural fragments permits the structural interpretation of the correlation found with them to describe the property under analysis. In short, spectral moments are related with the number of different fragments in the graph and link molecular mass, branching, and steric factors to properties.[62] As in the case of the eigenvalues mentioned above, the ESpm02u index seems to be sensible to the presence of condensed rings.

In summary, as was to be expected, anorexia as a side effect is a complex property, and hence, it is beyond the aim of this paper to discuss in detail the events in terms of structure−activity. However, some issues seem to be significant, for instance, the presence and location of the oxygen atoms in the molecule, as well as pure structural features such as the multiple and simple path counts, branching degree, cyclation, etc. Regarding the presence of oxygen atoms, it is a noteworthy fact that there are a significant number of compounds showing anorexia and containing sugar moieties, for instance, cytarabine, daunorubicin, doxorubicin, emtricitabine, and etoposide, among others. Apart from this, on average, both the active and inactive compounds show a similar elemental composition (similar number of C, O, S, etc.) and similar size, that reinforces the idea that pure structural or topological features are playing a key role in this property.

Once the function's robustness is confirmed, it can be applied to the search of new anorexia-free drugs. A good way to proceed is to use the pharmacological distribution diagrams or PDD. Figure 2 shows the PDD obtained for our data set. As shown in the diagram, those compounds with DF values between 1 and 8 are clearly anorexic inducers, those between −1.5 and −5 do not produce anorexia, and finally, those compounds located between 1 and −1.5 are uncertain (not classified). The model was thought to be used so that no one possible anorexia-inducer compound would be left out, i.e., in this case we chose sensitivity rather than specificity to prevent the risk of anorexia.

dx.doi.org/10.1021/ci2006092 | J. Chem. Inf. Model. 2012, 52, 1337−1344

Table 3. Results of Prediction of Anorexia through the Discriminant Function Obtained Using the Molecular Topology for the *Test Set*[a]

Compound	Prob (A)	Class	DF	Compound	Prob (A)	Class	DF
Active group							
Amphotericin	0.743	A	1.030	Penicillamine[61]	0.999	A	7.581
Aripiprazole	0.829	A	1.582	Pergolide	0.740	A	1.047
Capecitabine	0.960	A	3.177	Pramlintide[61]	0.998	A	5.203
Clofarabine	0.875	A	1.942	Riluzole	0.736	A	1.027
Dexrazoxane	0.967	A	3.378	Thalidomide	0.325	I	−0.730
Fenfluramine	0.889	A	2.081	Tocainide[61]	0.253	I	−1.085
Imatinib	0.119	I	−1.999	Valproic Acid[61]	0.971	A	3.527
Levetiracetam	0.967	A	3.366	Venlafaxine	0.960	A	3.171
Linezolid	0.882	A	2.015	Zidovudine	0.812	A	1.462
Paliperidone	0.901	A	2.203	Zoledronic Acid	0.876	A	1.950
Inactive group							
Acitretin	0.018	I	−4.011	Indomethacin	0.072	I	−2.553
Allopurinol	0.031	I	−3.438	Iodixanol	0.893	A	2.093
Ambrisentan	0.236	I	−1.178	Isocarboxacid	0.173	I	−1.562
Amlodipine	0.334	I	−0.693	Ketoprofen	0.101	I	−2.184
Balsalazide	0.092	I	−2.294	Latanoprost	0.929	A	2.572
Buspirone	0.973	A	3.567	Porfimer	0.251	I	−1.096
Carbamazepine	0.055	I	−2.843	Rofecoxib	0.169	I	−1.594
Chantix	0.492	I	−0.031	Sprycel	0.321	I	−0.749
Clomipramine	0.419	I	−0.327	Temazepam	0.056	I	−2.818
Clozapine	0.373	I	−0.518	Temozolomide	0.280	I	−0.943
Deferasirox	0.118	I	−2.011	Tretinoin	0.201	I	−1.382
Diclofenac	0.148	I	−1.747	Verapamil	0.781	A	1.273
Didanosine	0.481	I	−0.078	Voriconazole	0.228	I	−1.220
Femara	0.069	I	−2.600	Ziagen	0.480	I	−0.081
Flecainide	0.958	A	3.129	Zolmitriptan	0.305	I	−0.824
Goserelin	0.025	I	−3.674	Zonisamide	0.238	I	−1.161
Guanfacine	0.062	I	−2.712	Zopiclone	0.418	I	−0.332
Imiquimod	0.038	I	−3.224				

[a]Prob (*A*) = probability of activity. DF = discriminant function. Class = classification as active or inactive.

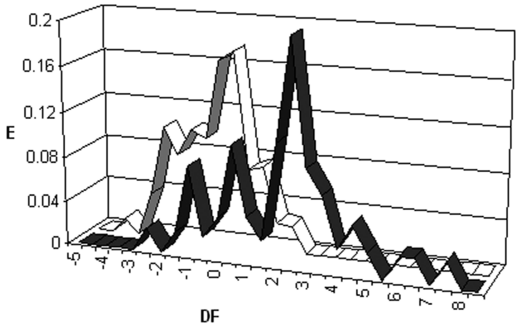


Figure 2. Pharmacological distribution diagram for compounds causing anorexia as a side effect by plotting expectancy (*E*) versus the DF function (black and white bars represent the active and inactive compounds, respectively).

The applicability domain of the model will be reduced to compounds showing a DF value between 1 and 8, the range in which we obtained the greatest expectancy in finding the anorexia effect, and between −1.5 and −5, which showed the greatest probability of finding nonanorexic compounds. Therefore, a virtual screening of anorexic compounds or nonanorexic compounds could be carried out applying this model, depending on the purpose of the study, as this model could be applied either to search for new anorexic compounds or for identifying compounds suffering for anorexia as a side effect.

CONCLUSIONS

The results outlined in this work clearly point toward the efficacy of molecular topology in the prediction of a very important drug side effect, the induction of anorexia. In fact, the mathematical model described herein is capable of distinguishing satisfactorily between drugs showing and not showing anorexia. This predictive capability performed well in both the *training set* and the *test set* of drug compounds. As far as we know, this is the first time that a side effect has been so accurately predicted using topological indices and discriminant analysis as main tools, and it could hopefully be a first step to proceed with profiling other side effects of drugs, which can save a lot of time and money in the process of drug design and discovery.

ASSOCIATED CONTENT

Supporting Information

Value of the descriptors used in the drug-induced anorexia model for all the molecules conforming the *training* and *test set*. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*Tel: 9635 44291. Fax: 9635 44892. E-mail: ramon.garcia@uv.es.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank the Ministerio de Economia y competitividad, Spain (Project SAF2009-13059-C03-02) for support of this work. M. Galvez Llompart acknowledges the "V Segles" Fellowship provided by the University of Valencia to carry out this study.

■ REFERENCES

(1) Wysowski, D. K.; Swartz, L. Adverse drug event surveillance and drug withdrawals in the United States, 1969−2002: The importance of reporting suspected reactions. *Arch. Intern. Med.* **2005**, *12*, 1363−1369.

(2) Nebeker, J. R.; Barach, P.; Samore, M. H. Clarifying adverse drug events: A clinician's guide to terminology, documentation, and reporting. *Ann. Intern. Med.* **2004**, *10*, 795−801.

(3) *Reducing and Preventing Adverse Drug Events to Decrease Hospital Costs*; Research in Action, Issue 1; AHRQ Publication Number 01-0020; Agency for Healthcare Research and Quality: Rockville, MD, March 2001. http://www.ahrq.gov/qual/aderia/aderia.htm (accessed March 2, 2012).

(4) Tisdale, J. E., Miller, D. A., Eds.; *Drug-Induced Diseases: Prevention, Detection and Management*, 2th ed.; American Society of Health-System Pharmacists: Bethesda, MD, 2010.

(5) Inui, A. Recent development in research and management of cancer anorexia-cachexia syndrome. *Gan to Kagaku Ryoho* **2005**, *6*, 743−749.

(6) Neary, N. M.; Small, C. J.; Wren, A. M.; Lee, J. L.; Druce, M. R.; Palmieri, C.; Frost, G. S.; Ghatei, M. A.; Coombes, R. C.; Bloom, S. R. Ghrelin increases energy intake in cancer patients with impaired appetite: Acute, randomized, placebo-controlled trial. *J. Clin. Endocrinol. Met.* **2004**, *6*, 2832−2836.

(7) Von Roenn, J. H.; Knopf, K. Anorexia/cachexia in patients with HIV: Lessons for the oncologist. *Oncology* **1996**, *7*, 1049−1056.

(8) Bruera, E. ABC of palliative care: Anorexia, cachexia, and nutrition. *Br. Med. J.* **1997**, *7117*, 1219−1222.

(9) Ströhle, A.; Zänker, K.; Hahn, A. Nutrition in oncology: The case of micronutrients (Review). *Oncol. Rep.* **2010**, *4*, 815−828.

(10) Blackburn, G. L.; Bistrian, B. R.; Maini, B. S.; Schlamm, H. T.; Smith, M. F. Nutritional and metabolic assessment of the hospitalized patient. *J. Parenter. Enteral Nutr.* **1977**, *1*, 11−22.

(11) Glare, P.; Sinclair, C.; Downing, M.; Stone, P.; Maltoni, M.; Vigano, A. Predicting survival in patients with advanced disease. *Eur. J. Cancer* **2008**, *8*, 1146−1156.

(12) Isabel, T. The impact of malnutrition on morbidity, mortality, length of hospital stay and costs evaluated through a multivariate model analysis. *Clin. Nutr.* **2003**, *3*, 235−239.

(13) Arrieta, O.; Ortega, R. M.; Villanueva-Rodríguez, G.; Serna-Thomé, M.; Flores-Estrada, D.; Diaz-Romero, C.; Rodríguez, C.; Martínez, L.; Sánchez-Lara, K. Association of nutritional status and serum albumin levels with development of toxicity in patients with advanced non-small cell lung cancer treated with paclitaxel−cisplatin chemotherapy: A prospective study. *BMC Cancer* **2010**, *1*, 50.

(14) Slaviero, K. A.; Read, J. A.; Clarke, S. J.; Rivory, L. P. Baseline nutritional assessment in advanced cancer patients receiving palliative chemotherapy. *Nutr. Cancer* **2003**, *2*, 148−157.

(15) Sánchez-Lara, K.; Sosa-Sánchez, R.; Green-Renner, D.; Méndez-Sánchez, N. Observaciones sobre la patogénesis de la anorexia asociada a cáncer y su regulación por el sistema nervioso central. *Nutr. Hosp.* **2011**, *4*, 677−684.

(16) de Julian-Ortiz, J. V.; de Gregorio Alapont, C.; Ríos-Santamarina, I.; García-Doménech, R.; Gálvez, J. Prediction of properties of chiral compounds by molecular topology. *J. Mol. Graphics Modell.* **1998**, *1*, 14−18.

(17) Golbraikh, A.; Bonchev, D.; Tropsha, A. Novel ZE-isomerism descriptors derived from molecular topology and their application to QSAR analysis. *J. Chem. Inf. Comput. Sci.* **2002**, *4*, 769−787.

(18) Kier, L. B.; Hall, L. H. In *Molecular Connectivity in Chemistry and Drug Research*; Academic Press: New York; 1976; Vol. *2*.

(19) Kier, L. B.; Hall, L. H. In *Molecular Connectivity in Structure−Activity analysis*; Research Studies Press John Wiley and Sons: Chichester, U.K.; 1986; Vol. *9*.

(20) Kier, L. B.; Hall, L. H.; Hall, K. In *Molecular Structure Description: The Electrotopological State*; Academic Press: San Diego; 1999.

(21) Galvez, J. Prediction of molecular volume and surface of alkanes by molecular topology. *J. Chem. Inf. Comput. Sci.* **2003**, *4*, 1231−1239.

(22) Pla-Franco, J.; Gálvez-Llompart, M.; Gálvez, J.; García-Domenech, R. Application of molecular topology for the prediction of reaction yields and anti-inflammatory activity of heterocyclic amidine derivatives. *Int. J. Mol. Sci.* **2011**, *2*, 1281−1292.

(23) Bruno-Blanch, L.; Galvez, J.; Garcia-Domenech, R. Topological virtual screening: A way to find new anticonvulsant drugs from chemical diversity. *Bioorg. Med. Chem. Lett.* **2003**, *16*, 2749−2754.

(24) García-Domenech, R.; Gálvez, J.; de Julián-Ortiz, J. V.; Pogliani, L. Some new trends in chemical graph theory. *Chem. Rev.* **2008**, *3*, 1127−1169.

(25) Mahmoudi, N.; de Julián-Ortiz, J. V.; Ciceron, L.; Gálvez, J.; Mazier, D.; Danis, M.; Derouin, F.; García-Domenech, R. Identification of new antimalarial drugs by linear discriminant analysis and topological virtual screening. *J. Antimicrob. Chemother.* **2006**, *3*, 489−497.

(26) Mahmoudi, N.; Garcia-Domenech, R.; Galvez, J.; Farhati, K.; Franetich, J. F.; Sauerwein, R.; Hannoun, L.; Derouin, F.; Danis, M.; Mazier, D. New active drugs against liver stages of Plasmodium predicted by molecular topology. *Antimicrob. Agents Chemother.* **2008**, *4*, 1215−1220.

(27) de Gregorio Alapont, C.; Garcia-Domenech, R.; Gálvez, J.; Ros, M.; Wolski, S.; García, M. Molecular topology: a useful tool for the search of new antibacterials. *Bioorg. Med. Chem. Lett.* **2000**, *17*, 2033−2036.

(28) Mishra, R. K.; Garcia-Domenech, R.; Galvez, J. Getting discriminant functions of antibacterial activity from physicochemical and topological parameters. *J. Chem. Inf. Comput. Sci.* **2001**, *2*, 387−393.

(29) Pastor, L.; García-Domenech, R.; Galvez, J.; Wolski, S.; García, M. D. New antifungals selected by molecular topology. *Bioorg. Med. Chem. Lett.* **1998**, *18*, 2577−2582.

(30) Jasinski, P.; Welsh, B.; Galvez, J.; Land, D.; Zwolak, P.; Ghandi, L.; Terai, K.; Dudek, A. Z. A novel quinoline, MT477: Suppresses cell signaling through Ras molecular pathway, inhibits PKC activity, and demonstrates in vivo anti-tumor activity against human carcinoma cell lines. *Invest. New Drugs* **2008**, *3*, 223−232.

(31) Duart, M. J.; García-Domenech, R.; Gálvez, J.; Alemán, P. A.; Martín-Algarra, R. V.; Antón-Fos, G. M. Application of a mathematical topological pattern of antihistaminic activity for the selection of new drug candidates and pharmacology assays. *J. Med. Chem.* **2006**, *12*, 3667−3673.

(32) Ríos-Santamarina, I.; Garcia-Domenech, R.; Gálvez, J. New bronchodilators selected by molecular topology. *Bioorg. Med. Chem. Lett.* **1998**, *5*, 477−482.

(33) Rios-Santamarina, I.; Garcia-Domenech, R.; Galvez, J.; Morcillo, E.; Santamaria, P.; Cortijo, J. Getting new bronchodilator compounds from molecular topology. *Eur. J. Pharm. Sci.* **2004**, *4*, 271−277.

(34) Galvez, J.; Garcia-Domenech, R.; Gomez-Lechon, M.; Castell, J. Use of molecular topology in the selection of new cytostatic drugs. *J. Mol. Struct.: THEOCHEM* **2000**, *1−3*, 241−248.

(35) Galvez-Llompart, M.; Giner, M.; Recio, C.; Candeletti, S.; Garcia-Domenech, R. Application of molecular topology to the search of novel NSAIDs: Experimental validation of activity. *Lett. Drug Des. Discovery* **2010**, *6*, 438−445.

(36) *Dragon*, version 5.4; Talete srl; Milano, Italy, 2006. http://www.talete.mi.it/ (accessed March 5, 2012).

(37) Kuhn, M.; Campillos, M.; Letunic, I.; Jensen, L. J.; Bork, P. A side effect resource to capture phenotypic effects of drugs. *Mol. Syst. Biol.* **2010**, *6*, 343.

(38) Klecka, W. R. *Discriminant Analysis*; Sage Publications, Inc.: Los Angeles, 1980; Vol. *19*.
(39) Hastie, T.; Tibshirani, R. Discriminant analysis by Gaussian mixtures. *J. R. Stat. Soc., Ser. B* **1996**, *58*, 155−176.
(40) De Maesschalck, R.; Jouan-Rimbaud, D.; Massart, D. The Mahalanobis distance. *Chemom. Intell. Lab. Syst.* **2000**, *1*, 1−18.
(41) Everitt, B.; Dunn, G. *Applied Multivariate Data Analysis*; Arnold: London, 2001.
(42) Eadie, W. T.; Drijard, D.; James, F. E.; Roos, M.; Sadoulet, B. *Statistical Methods in Experimental Physics*, North-Holland: Amsterdam, 1971.
(43) Furnival, G. M. All possible regressions with less computation. *Technometrics* **1971**, *13*, 403−408.
(44) Matthews, B. W. Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochim. Biophys. Acta Protein Struct.* **1975**, *2*, 442−451.
(45) *Statistica (data analysis software system)*, version 9.0; Statsoft: Tulsa, OK, 2009.
(46) Gálvez, J.; García-Domenech, R.; de Gregorio Alapont, C.; de Julián-Ortiz, J.; Popa, L. Pharmacological distribution diagrams: A tool for de novo drug design. *J. Mol. Graphics* **1996**, *5*, 272−276.
(47) Aronson, J. *Meyler's Side Effect of Drugs: The International Encyclopedia of Adverse Drug Reactions and Interactions*, 15th ed.; Elsevier Science: New York, 2006.
(48) Schilsky, R. L.; Dolan, M. E.; Bertucci, D.; Ewesuedo, R. B.; Vogelzang, N. J.; Mani, S.; Wilson, L. R.; Ratain, M. J. Phase I clinical and pharmacological study of O6-benzylguanine followed by carmustine in patients with advanced cancer. *Clin. Cancer Res.* **2000**, *8*, 3025−3031.
(49) Oliveras, G.; Blancafort, A.; Urruticoechea, A.; Campuzano, O.; Gómez-Cabello, D.; Brugada, R.; López-Rodríguez, M.; Colomer, R.; Puig, T. Novel anti-fatty acid synthase compounds with anti-cancer activity in HER2 breast cancer. *Ann. N.Y. Acad. Sci.* **2010**, *1*, 86−92.
(50) Daun, J. M.; McCarthy, D. O. The role of cholecystokinin in interleukin-1-induced anorexia. *Physiol. Behav.* **1993**, *2*, 237−241.
(51) Liebowitz, M. R.; Yeung, P. P.; Entsuah, R. A randomized, double-blind, placebo-controlled trial of desvenlafaxine succinate in adult outpatients with major depressive disorder. *J. Clin. Psychiatry* **2007**, *11*, 1663−1672.
(52) Pekcan, Z.; Koc, B. The post-operative analgesic effects of epidurally administered morphine and transdermal fentanyl patch after ovariohysterectomy in dogs. *Vet. Anaesth. Analg.* **2010**, *6*, 557−565.
(53) Wryter, L. E.; Solimando, J., D.A.; Waddell, J. A. Cancer chemotherapy update − Deferasirox; lenalidomide. *Hosp. Pharm.* **2006**, *5*, 420−425.
(54) Leung, W.; Neil, T. G.; Chan, J. C. N.; Tomlinson, B. Weight management and current options in pharmacotherapy: Orlistat and sibutramine. *Clin. Ther.* **2003**, *1*, 58−80.
(55) Fossella, F. V.; Gatzemeier, U.; Fossella, F. V.; Gatzemeier, U. Phase I trials of pemetrexed. *Semin Oncol.* **2002**, *29*, 8−16.
(56) Ukraintsevac, V. N. A. S. V.; Anikina, I. V.; Zabezhinskia, I. G. P. M. A.; Bertseina, L. M.; Ingramd, A. V. A. D. K.; Laned, M. A.; Rothd, G. S. Effects of phentermine and phenformin on biomarkers of aging in rats. *Gerontology* **2005**, *51*, 19−28.
(57) Aldoss, I. T.; Tashi, T.; Ganti, A. K. Seliciclib in malignancies. *Expert Opin. Invest. Drugs* **2009**, *12*, 1957−1965.
(58) Pawlik, T. M.; Reyes, D. K.; Cosgrove, D.; Kamel, I. R.; Bhagat, N.; Geschwind, J. F. H. Phase II trial of sorafenib combined with concurrent transarterial chemoembolization with drug-eluting beads for hepatocellular carcinoma. *J. Clin. Oncol.* **2011**, *30*, 3960−3967.
(59) Suyama, T.; Ueda, T.; Fukasawa, S.; Komaru, A.; Kobayashi, M.; Naya, Y.; Nihei, N.; Ichikawa, T. Efficacy and safety of sunitinib in Japanese patients with metastatic renal cell carcinoma. *J. Cancer Ther.* **2011**, *2*, 335−341.
(60) Cho, Y. J.; Heo, K.; Kim, W. J.; Jang, S. H.; Jung, Y. H.; Ye, B. S.; Song, D. B.; Inlee, B. Long-term efficacy and tolerability of topiramate as add-on therapy in refractory partial epilepsy: An observational study. *Epilepsia* **2009**, *8*, 1910−1919.
(61) FDA Professional Drug Information. http://www.drugs.com/pro/ (accessed March 6, 2012).
(62) Estrada, E. Spectral moments of the edge adjacency matrix of molecular graphs. 1. Definition and applications to the prediction of physical properties of alkanes. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 844−849.