QSPR modelling with the topological substructural molecular design approach: β-cyclodextrin complexation

ARTICLE *in* JOURNAL OF PHARMACEUTICAL SCIENCES · DECEMBER 2009

Impact Factor: 2.59 · DOI: 10.1002/jps.21747 · Source: PubMed

4 AUTHORS, INCLUDING:

Alfonso Pérez-Garrido
Universidad Católica San Antonio de Murcia
16 PUBLICATIONS **128** CITATIONS
SEE PROFILE

Aliuska Morales Helguera
50 PUBLICATIONS **1,080** CITATIONS
SEE PROFILE

Natália D. S. Cordeiro
University of Porto
245 PUBLICATIONS **3,040** CITATIONS
SEE PROFILE

QSPR Modelling With the Topological Substructural Molecular Design Approach: β-Cyclodextrin Complexation

ALFONSO PÉREZ-GARRIDO,[1,2] ALIUSKA MORALES HELGUERA,[3,4,5] M. NATÁLIA D.S. CORDEIRO,[5] AMALIO GARRIDO ESCUDERO[1]

[1]Environmental Engineering and Toxicology Department, Catholic University of San Antonio, Guadalupe, Murcia, C.P. 30107, Spain

[2]Department of Food and Nutrition Technology, Catholic University of San Antonio, Guadalupe, Murcia, C.P. 30107, Spain

[3]Faculty of Chemistry and Pharmacy, Department of Chemistry, Central University of Las Villas, Santa Clara, 54830 Villa Clara, Cuba

[4]Molecular Simulation and Drug Design Group, Chemical Bioactive Center, Central University of Las Villas, Santa Clara, 54830 Villa Clara, Cuba

[5]REQUIMTE, Faculty of Sciences, Chemistry Department, University of Porto, 4169-007 Porto, Portugal

Received 10 October 2008; revised 30 January 2009; accepted 11 February 2009

Published online 5 June 2009 in Wiley InterScience (www.interscience.wiley.com). DOI 10.1002/jps.21747

ABSTRACT: This study aims at developing a quantitative structure–property relationship (QSPR) model for predicting complexation with β-cyclodextrins (β-CD) based on a large variety of organic compounds. Molecular descriptors were computed following the *TOP*ological *S*ubstructural *MO*lecular *DE*sign (TOPS-MODE) approach and correlated with β-CD complex stability constants by linear multivariate data analysis. This strategy afforded a final QSPR model that was able to explain around 86% of the variance in the experimental activity, along with showing good internal cross-validation statistics, and also good predictivity on external data. Topological substructural information influencing the complexation with β-CD was extracted from the QSPR model. This revealed that the major driving forces for complexation are hydrophobicity and van der Waals interactions. Therefore, the presence of hydrophobic groups (hydrocarbon chains, aryl groups, etc.) and voluminous species (Cl, Br, I, etc.) in the molecules renders easy their complexity with β-CDs. To our knowledge, this is the first time a correlation between TOPS-MODE descriptors and complexing abilities of β-CDs has been reported.
 J Pharm Sci98:4557–4576, 2009
Keywords: QSPR; QSAR; drug design; cyclodextrins; complexation

INTRODUCTION

Cyclodextrins (CDs) are cyclic oligomers of β-D-glucose produced from starch by means of enzymatic conversion, and shaped like truncated cones with primary and secondary hydroxyl groups crowning the narrower rim and wider rim, respectively.[1] CDs have attracted much interest in many fields, because they are able to form host–guest complexes with hydrophobic molecules and greatly modify their physical and chemical properties, mostly in terms of water solubility. For instance, upon complexation with CDs, the drugs solubility strongly increases, making them available for a wide range of pharmaceutical applications. Different drugs are currently marketed as solid or solution-based CD complex

Additional Supporting Information may be found in the online version of this article.

Correspondence to: Alfonso Pérez-Garrido (Telephone: +34-968278755; Fax: +34-968278622;
E-mail: aperez@pdi.ucam.edu)

Journal of Pharmaceutical Sciences, Vol. 98, 4557–4576 (2009)

formulations.[2,3] In these pharmaceutical products, CDs are mainly used as complexing agents to increase the aqueous solubility of poorly water-soluble drugs, to increase their bioavailability and stability.[4–6] Poor solubility continues to impact the development of a large number of potential drug candidates.[7] These factors have had a significant impact on what is required from formulators given that the number of formulation options, and by extension excipients, has to be increased to address the larger number of challenges being presented.[8] CDs represent a true added value in this context.

In addition, CDs can also promote drug absorption across the dermal, nasal, or intestinal barrier by extracting cholesterol, phospholipids, or proteins from membranes,[9] reduce or prevent gastrointestinal and ocular irritation, reduce or eliminate unpleasant smells or tastes,[10,11] prevent drug–drug or drug–additive interactions, as well as to convert oils and liquid drugs into microcrystalline or amorphous powders.[12] Moreover, pharmacon–CD complexes often increase the bioavailability of the active substances and permit their controlled release.[13] An example of the latter is the CD encapsulation of *trans*-platinum complex where it has been found that the cytotoxicity *in vitro* of the novel inclusion complex indicated a much higher activity.[14]

The experimental determination of CD complex binding constants is often difficult and time consuming because of the low solubility of the guest molecules in aqueous solution. Previous studies have suggested five major types of interactions: (i) hydrophobic interactions, (ii) van der Waals interactions, (iii) hydrogen-bonding between polar groups of the *guest* and the hydroxyl groups of the *host*, (iv) relaxation by release of high-energy water from the CD cavity upon substrate inclusion, and (v) relief of the conformational strain in a CD–water adduct.

In contrast, computational methods have only recently been used for predicting binding constants and to study the driving forces involved in the process. An exhaustive set of these computational applications has been excellently reviewed by Lipkowitz.[15]

Group-contribution models, quantitative structure–activity/property relationships (QSAR/QSPR) methods (2D-QSAR, 3D-QSAR, CoMFA), molecular modelling computations (using Quantum Mechanics, Monte Carlo/Molecular Dynamics Simulations, Molecular Mechanics, etc.), statistical analysis tools, and artificial neural networks have all been applied to elucidate the most important factors influencing the host–guest interactions and to predict the thermodynamic stability of CDs inclusion complexes.[16–25]

Nevertheless, it is clear that knowledge of the complexation abilities of guest molecules with CDs is deemed necessary to decide whether or not a host–guest complexation is useful in a particular application using the knowledge of what kind of bonds contribute positively to this phenomenon. In this sense, Katritzky et al.[25] presented a QSAR study predicting the free energies of inclusion complexation between diverse *guest* molecules and CDs using (i) CODESSA descriptors and (ii) counts of different molecular fragments. The fragmental descriptors are more easily interpretable than CODESSA descriptors. One can select the fragments whose contributions are considerable and give reasonable explanations based on physical phenomena involved in host–guest complexation. However, QSPR models based on fragments generally comprise much more variables than those using traditional descriptors, which still remain as an important problem.

The aim of the present study was to build a QSPR regression-based model, which could correlate and predict the complex stability constant between diverse guest molecules and β-CDs using the *TOP*ological *S*ubstructural *MO*lecular *DE*sign (TOPS-MODE) descriptors.[26–28] There is evidence that these descriptors performed well in similar QSAR/QSPR modelling studies on which they have been used because they are easy to calculate, and one can draw from the derived models useful information regarding the type of structures that contribute favourably or not to the activity or property.[29–42] This approach is able to transform simple molecular descriptors, such as log *P*, polar surface area, molar refraction, charges, etc., into series of descriptors that account for the distribution of these characteristics (hydrophobicity, polarity, steric effects, etc.) across the molecule. Thus, we can obtain this structural information at a local scale from the models developed using global molecular descriptors. It has been recognised that the TOPS-MODE approach "provides a mechanistic interpretation at a bond level and enables the generation of new hypotheses such as structural alerts."[43] Such valuable information can then be used for the design of new drugs with increased bioavailability and solubility due to their complexation with β-CDs.

EXPERIMENTAL

Data Set

The overall data set of 233 substances comprised a large number of classes of organic compounds: aromatic hydrocarbons, alcohols, phenols, ethers,aldehydes, ketones, acids, esters, nitriles, anilines, halogenated compounds, heterocycles, nitro, sulphur and steroids and barbitals compounds. This set of guest molecules was extracted from the work of Suzuki,[16] and the experimental endpoint to be predicted is the β-CD complex stability constants (K), which have been measured at $T = 298.15$ K using water as solvent, taken from references therein. Two of such guest molecules, chemicals 214 and 215, are stereoisomers, which could not be distinguished by the present 2D descriptors but had nevertheless different K values. Thus, one of the isomers was discarded (chemical 215), the other one (chemical 214) being only considered in our study with an averaged value of K. Moreover, all K values were log-transformed ($\log K$) for being of practical use in the following QSPR modelling. Table 1 displays a complete list of the chemicals along with the reported experimental data.

The TOPS-MODE Descriptors

The TOPS-MODE descriptors are based on the calculation of the spectral moments of the so-called bond matrix.[28] The theoretical foundations of the spectral moments have been reported previously,[26,27] nevertheless an overview of this descriptor family will be given here. The spectral moments are defined as the traces of the bond adjacency matrix. That is, the sum of the main diagonal elements of different powers of such matrix. The bond adjacency matrix is a squared symmetric matrix whose entries are ones or zeros if the corresponding bonds are adjacent or not. The order of this matrix (m) is the number of bonds in the molecular graph, two bonds being adjacent if they are incident to a common atom. Furthermore, weights are introduced in the diagonal entries of this matrix to mirror fundamental physicochemical properties that might relate to the target endpoint being modelled. Here, several bond weights were used for computing the spectral moments, namely the standard bond distance (Std), standard bond dipole moments (Dip, Dip2), hydrophobicity (H), polar surface area (Pols), polarisability (Pol), molar refractivity (Mol), van der Waals radii (Van), Gasteiger–Marsilli charges (Gas), atomic masses (Ato), solute excess molar refraction (Ab-R_2), solute dipolarity/polarisability (Ab-π_2^H), effective hydrogen-bond basicity (Ab-$\sum \beta_2^O$, Ab-$\sum \beta_2^H$) and solute gas-hexadecane partition coefficient (Ab-$\log L^{16}$) were used for computing the spectral moments of the bond matrix.

Explicitly, we have calculated the first 15 spectral moments (μ_1–μ_{15}) for each bond weight and the number of bonds in the molecules (μ_0) without hydrogen. Also, we multiplied μ_0 and μ_1 for the first 15 spectral moments obtaining 30 new variables. Notice that in this way such variables might offset the linear approximation assumption of the model. As described previously,[44] the atomic contributions were then transformed into bond contributions as follows:

$$w_{(i,j)} = \frac{w_i}{\delta_i} + \frac{w_j}{\delta_j} \quad (1)$$

where w_i and δ_i are the atomic weight and vertex degree of the atom i. Calculation of the TOPS-MODE descriptors was carried out with the MODESLAB software (http://www.modeslab.-com)[45] from the SMILES (Simplified Molecular Input Line Entry System) notation available for each compound.[46] To develop the structure–property relationships, the following six-step path was adopted:

1. Select a small subset of the 233 chemicals to act as a test set. The remaining chemicals form the training set for QSPR modelling.
2. Draw the molecular graphs for each molecule included in the training set.
3. Compute the spectral moment's descriptors using an appropriate set of weights.
4. Find an adequate QSPR model from the training set by a regression-based approach. The task here is to obtain a mathematical function (see Eq. 2 below) that best describes the studied property P (in our case, the $\log K$ partitioning) as a linear combination of the X-predictor variables (the spectral moments μ_k), with the coefficients a_k. Such coefficients are to be optimised by means of multiple linear regression (MLR) analysis along with a variable subset selection procedure

$$P = a_0\mu_0 + a_1\mu_1 + a_2\mu_2 + \cdots + a_k\mu_k \quad (2)$$

Table 1. Names, CAS Number, Observed ($\log K_o$) and Predicted ($\log K_p$) Activity,* and Leverage (h) Values for the Compounds Used in this Study

No.	Name	CAS	$\log K_o$	$\log K_p$	Partition	h	Refs.
1	Carbon tetrachloride	56-23-5	2.20	2.29	Test	0.174[b]	20
2	Chloroform	67-66-3	1.43	0.66	Training	—	20
3	Methanol	67-56-1	−0.49	−0.35	Training	—	18
4	Acetonitrile	75-05-8	−0.27	−0.36	Training	—	20
5	Acetaldehyde	75-07-0	−0.64	−0.23	Training	—	20
6	Ethanol	64-17-5	−0.03	0.19	Test	0.055	18
7	1,2-Ethanediol	107-21-1	−0.19	−0.02	Training	—	18
8	Acetone	67-64-1	0.42	0.40	Training	—	20
9	1-Propanol	71-23-8	0.57	0.69	Test	0.039	18
10	2-Propanol	67-63-0	0.63	0.93	Training	—	20
11	1,3-Propanediol	504-63-2	0.67	0.46	Training	—	18
12	Tetrahydrofuran	109-99-9	1.47	1.10	Test	0.034	20
13	Cyclobutanol	2919-23-5	1.18	1.51	Training	—	18
14	1-Butanol	71-36-3	1.22	1.18	Training	—	79
15	2-Butanol	78-92-2	1.19	1.37	Training	—	18
16	2-Methyl-1-propanol	78-83-1	1.62	1.37	Training	—	18
17	2-Methyl-2-propanol	75-65-0	1.68	2.01	Training	—	18
18	1,4-Butanediol	110-63-4	0.64	0.93	Training	—	18
19	Diethylamine	109-89-7	1.36	1.22	Training	—	20
20	Cyclopentanol	96-41-3	2.08	1.70	Training	—	18
21	1-Pentanol	71-41-0	1.80	1.64	Training	—	79
22	2-Pentanol	6032-29-7	1.49	1.83	Training	—	18
23	3-Pentanol	584-02-1	1.35	1.78	Training	—	18
24	2-Methyl-1-butanol	1565-80-6	2.08	1.80	Training	—	18
25	2-Methyl-2-butanol	75-85-4	1.91	2.29	Training	—	18
26	3-Methyl-1-butanol	123-51-3	2.25	1.85	Test	0.023	18
27	3-Methyl-2-butanol	1517-66-4	1.92	1.92	Training	—	18
28	2,2-Dimethyl-1-propanol	75-84-3	2.71	2.39	Test	0.027	79
29	1,5-Pentanediol	111-29-5	1.22	1.38	Test	0.026	18
30	1,4-Dibromobenzene	106-37-6	2.97	2.78	Training	—	22
31	1,4-Diiodobenzene	624-38-4	3.17	3.63	Training	—	22
32	3,5-Dibromophenol	626-41-5	2.56	2.79	Training	—	18
33	3,5-Dichlorophenol	591-35-5	2.07	2.53	Test	0.045	18
34	1-Chloro-4-nitrobenzene	100-00-5	2.15	2.52	Training	—	22
35	Fluorobenzene	462-06-6	1.96	2.02	Test	0.026	22
36	Bromobenzene	108-86-1	2.50	2.42	Training	—	22
37	Iodobenzene	591-50-4	2.93	2.87	Training	—	22
38	3-Fluorophenol	372-20-3	1.70	2.03	Training	—	18
39	4-Fluorophenol	371-41-5	1.73	2.02	Training	—	18
40	3-Chlorophenol	108-43-0	2.28	2.28	Training	—	18
41	4-Chlorophenol	106-48-9	2.61	2.27	Training	—	22
42	3-Bromophenol	591-20-8	2.51	2.42	Test	0.031	18
43	4-Bromophenol	106-41-2	2.65	2.41	Test	0.031	22
44	3-Iodophenol	626-02-8	2.93	2.85	Training	—	18
45	4-Iodophenol	540-38-5	2.98	2.84	Training	—	22
46	Nitrobenzene	98-95-3	2.04	2.32	Training	—	20
47	4-Nitrophenol	100-02-7	2.39	2.29	Training	—	22
48	Benzene	71-43-2	2.23	2.05	Test	0.033	79
49	Phenol	108-95-2	1.98	2.05	Training	—	22
50	Hydroquinone	123-31-9	2.05	2.04	Test	0.024	22
51	4-Nitroaniline	100-01-6	2.48	2.35	Test	0.082	22

Table 1. (*Continued*)

No.	Name	CAS	$\log K_o$	$\log K_p$	Partition	h	Refs.
52	Aniline	62-53-3	1.60	2.12	Training	—	20
53	Sulphaanilamide	63-74-1	2.76	2.16	Training	—	21
54	Cyclohexanol	108-93-0	2.67	2.24	Training	—	79
55	1-Hexanol	111-27-3	2.33	2.09	Training	—	79
56	2-Hexanol	626-93-7	1.98	2.27	Training	—	18
57	2-Methyl-2-pentanol	590-36-3	1.99	2.73	Training	—	18
58	3-Methyl-3-pentanol	77-74-7	2.15	2.56	Training	—	18
59	4-Methyl-2-pentanol	108-11-2	2.04	2.48	Training	—	18
60	3,3-Dimethyl-2-butanol	464-07-3	2.75	2.94	Training	—	18
61	1,6-Hexanediol	629-11-8	1.69	1.80	Training	—	18
62	Benzonitrile	100-47-0	2.23	1.81	Training	—	22
63	Benzothiazole	95-16-9	2.38	1.92	Training	—	80
64	4-Nitrobenzoic acid	62-23-7	2.34	2.23	Training	—	22
65	Benzaldehyde	100-52-7	1.78	1.79	Training	—	20
66	Benzoic acid	65-85-0	2.12	2.03	Training	—	79
67	4-Hydroxybenzaldehyde	123-08-0	1.75	1.78	Training	—	18
68	4-Hydroxybenzoic acid	99-96-7	2.20	2.02	Training	—	22
69	Benzyl chloride	100-44-7	2.45	2.36	Training	—	22
70	Toluene	108-88-3	2.09	2.50	Training	—	20
71	Benzyl alcohol	100-51-6	1.71	2.05	Training	—	20
72	Anisole	100-66-3	2.32	2.12	Training	—	22
73	*m*-Cresol	108-39-4	1.98	2.49	Training	—	18
74	*p*-Cresol	106-44-5	2.40	2.48	Training	—	22
75	4-Methoxyphenol	150-76-5	2.21	2.10	Training	—	22
76	3-Methoxyphenol	150-19-6	2.11	2.11	Training	—	18
77	4-Hydroxybenzyl alcohol	623-05-2	2.16	2.01	Training	—	22
78	Hydrochlorothiazide	58-93-5	1.76	1.74	Training	—	21
79	*N*-methylaniline	100-61-8	2.12	2.14	Training	—	22
80	1-Butylimidazole	4316-42-1	2.19	2.27	Training	—	81
81	1-Heptanol	111-70-6	2.85	2.51	Test	0.026	18
82	Phenylacetylene	536-74-3	2.36	2.62	Training	—	22
83	Thianaphthene	95-15-8	3.23	2.49	Training	—	80
84	4-Fluorophenyl acetate	405-51-6	2.11	2.22	Test	0.027	18
85	3-Fluorophenyl acetate	701-83-7	1.91	2.23	Training	—	18
86	4-Chlorophenyl acetate	876-27-7	2.50	2.46	Training	—	18
87	3-Chlorophenyl acetate	13031-39-5	2.44	2.47	Training	—	18
88	4-Bromophenyl acetate	1927-95-3	2.68	2.59	Training	—	18
89	3-Bromophenyl acetate	35065-86-2	2.67	2.60	Test	0.032	18
90	4-Iodophenyl acetate	33527-94-5	3.00	2.93	Training	—	18
91	3-Iodophenyl acetate	61-71-2	3.07	2.94	Training	—	18
92	4-Nitrophenyl acetate	830-03-5	2.13	2.39	Training	—	18
93	Acetophenone	98-86-2	2.27	2.20	Training	—	22
94	Phenyl acetate	122-79-2	2.10	2.22	Training	—	18
95	Methyl benzoate	93-58-3	2.50	2.12	Training	—	22
96	3-Hydroxyacetophenone	121-71-1	2.06	2.19	Training	—	18
97	4-Hydroxyacetophenone	99-93-4	2.18	2.18	Training	—	22
98	Acetoanilide	103-84-4	2.20	1.92	Test	0.018	22
99	*p*-Xylene	106-42-3	2.38	2.92	Training	—	22
100	Ethylbenzene	100-41-4	2.59	2.80	Training	—	22
101	Phenetole	103-73-1	2.49	2.67	Training	—	22
102	2-Phenylethanol	60-12-8	2.15	2.48	Training	—	18
103	3-Ethylphenol	620-17-7	2.60	2.76	Training	—	18

(*Continued*)

Table 1. (*Continued*)

No.	Name	CAS	$\log K_o$	$\log K_p$	Partition	h	Refs.
104	4-Ethylphenol	123-07-9	2.69	2.75	Training	—	22
105	4-Ethoxyphenol	622-62-8	2.33	2.62	Test	0.008	18
106	3-Ethoxyphenol	621-34-1	2.35	2.63	Test	0.008	18
107	3,5-Dimethoxyphenol	500-99-2	2.34	2.13	Training	—	18
108	*N*-ethylaniline	103-69-5	2.34	2.67	Test	0.013	22
109	*N*,*N*-dimethylaniline	121-69-7	2.36	2.49	Training	—	22
110	Barbital	57-44-3	1.78	2.05	Training	—	21
111	Cyclooctanol	696-71-9	3.30	3.00	Training	—	18
112	1-Octanol	111-87-5	3.17	2.92	Test	0.033	18
113	2-Octanol	123-96-6	3.13	3.09	Training	—	18
114	Quinoline	91-22-5	2.12	2.37	Training	—	80
115	3-Cyanophenyl acetate	55682-11-6	1.49	2.14	Training	—	18
116	4-Hydroxycinnamic acid	7400-08-0	2.83	2.51	Training	—	21
117	Ethyl benzoate	93-89-0	2.73	2.63	Training	—	22
118	4′-Hydroxypropiophenone	70-70-2	2.63	2.43	Training	—	18
119	3′-Hydroxypropiophenone	13103-80-5	2.61	2.44	Training	—	18
120	*p*-Tolyl acetate	140-39-6	2.49	2.60	Training	—	18
121	3-Methylphenyl acetate	122-46-3	2.21	2.61	Training	—	18
122	4-Methoxyphenyl acetate	1200-06-2	2.45	2.24	Training	—	18
123	4-Propylphenol	645-56-7	3.55	3.11	Training	—	18
124	3-Propylphenol	621-27-2	3.28	3.12	Training	—	18
125	4-Isopropylphenol	99-89-8	3.58	3.15	Training	—	18
126	3-Isopropylphenol	618-45-1	3.44	3.16	Training	—	18
127	4-Isopropoxyphenol	7495-77-4	2.86	3.11	Training	—	18
128	2-Norbornaneacetate	—	3.59	3.11	Test	0.052	79
129	1-Benzylimidazole	4238-71-5	2.61	2.75	Training	—	81
130	*m*-Methylcinnamic acid	3029-79-6	2.93	2.92	Training	—	21
131	4-Ethylphenyl acetate	3245-23-6	2.83	2.82	Test	0.017	18
132	3-Ethylphenyl acetate	3056-60-8	2.68	2.83	Training	—	18
133	4-Ethoxyphenyl acetate	69788-77-8	2.54	2.72	Training	—	18
134	3-Ethoxyphenyl acetate	151360-54-2	2.49	2.73	Test	0.023	18
135	Allobarbital	52-43-7	1.98	2.15	Training	—	21
136	4-*n*-butylphenol	1638-22-8	3.97	3.46	Test	0.023	18
137	3-*n*-butylphenol	4074-43-5	3.76	3.46	Training	—	18
138	3-Isobutylphenol	30749-25-8	4.21	3.65	Training	—	18
139	4-*sec*-butylphenol	99-71-8	4.18	3.46	Training	—	18
140	3-*sec*-butylphenol	3522-86-9	4.06	3.47	Training	—	18
141	4-*tert*-butylphenol	98-54-4	4.56	3.84	Test	0.045	18
142	3-*tert*-butylphenol	585-34-2	4.41	3.85	Training	—	18
143	Menadion	58-27-5	2.27	2.30	Test	0.027	21
144	Sulphapyridine	144-83-2	2.70	2.57	Training	—	21
145	Sulphamonomethoxine	1220-83-3	2.48	2.20	Training	—	21
146	Sulfisoxazole	127-69-5	2.32	2.69	Training	—	21
147	4-*n*-propylphenyl acetate	61824-46-2	3.15	3.13	Training	—	18
148	3-*n*-propylphenyl acetate	—	3.28	3.14	Training	—	18
149	4-Isopropylphenyl acetate	2664-32-6	2.88	3.16	Training	—	18
150	3-Isopropylphenyl acetate	36438-57-0	3.36	3.17	Training	—	18
151	4-*n*-amylphenol	14938-35-3	4.19	3.78	Test	0.031	18
152	4-*tert*-amylphenol	80-46-6	4.70	4.02	Training	—	18
153	Carbutamide	339-43-5	2.29	2.37	Training	—	21
154	Pentobarbital	76-74-4	3.01	3.16	Test	0.069	21
155	Amobarbital	57-43-2	3.07	3.07	Training	—	79

Table 1. (*Continued*)

No.	Name	CAS	$\log K_o$	$\log K_p$	Partition	h	Refs.
156	Thiopental	76-75-5	3.28	3.12	Training	—	21
157	Dibenzofuran	132-64-9	2.97	2.60	Training	—	80
158	Dibenzothiophene	132-65-0	3.48	2.86	Training	—	80
159	Phenazine	92-82-0	2.41	2.03	Training	—	80
160	Thianthrene	92-85-3	3.57	3.48	Test	0.091	80
161	Carbazole	86-74-8	2.44	3.01	Training	—	80
162	Phenoxazine	135-67-1	2.69	2.75	Test	0.060	80
163	Phenothiazine	92-84-2	2.73	3.08	Training	—	80
164	Furosemide	200-203-6	1.78	3.02	Test	0.071	21
165	Phenobarbital	50-06-6	3.22	2.70	Test	0.062	79
166	Sulfisomidine	515-64-0	2.10	2.66	Test	0.061	21
167	Sulphamethomidine	3772-76-7	2.33	2.46	Test	0.038	21
168	Sulphadimethoxine	122-11-2	2.26	2.20	Training	—	21
169	4-*n*-butylphenyl acetate	55168-27-9	3.62	3.42	Training	—	18
170	3-*n*-butylphenyl acetate	—	3.66	3.43	Training	—	18
171	3-Isobutylphenyl acetate	916728-77-3	3.83	3.62	Training	—	18
172	4-*tert*-butylphenyl acetate	3056-64-2	3.85	3.83	Training	—	18
173	Cyclobarbital	52-31-3	2.71	2.55	Training	—	21
174	Hexobarbital	56-29-1	3.08	2.86	Training	—	21
175	1-Adamantaneacetate	875907-32-7	4.32	4.56	Training	—	79
176	Acridine	260-94-6	2.33	2.70	Training	—	80
177	Phenanthridine	229-87-8	2.57	2.61	Training	—	80
178	Xanthene	92-83-1	2.71	3.32	Training	—	80
179	*N*-phenylanthranilic acid	91-40-7	2.89	3.06	Training	—	79
180	Mephobarbital	115-38-8	3.16	3.03	Training	—	21
181	4-*n*-amylphenyl acetate	202831-79-6	3.80	3.69	Training	—	18
182	Flufenamic acid	530-78-9	3.10	3.08	Training	—	79
183	Meclofenamic acid	644-62-2	2.67	3.15	Training	—	79
184	Nitrazepam	146-22-5	1.97	1.70	Training	—	21
185	Flurbiprofen	5104-49-4	3.69	3.48	Training	—	21
186	Sulphaphenazole	526-08-9	2.35	1.97	Training	—	21
187	Bendroflumethiazide	200-800-1	1.90	2.02	Training	—	21
188	Mefenamic acid	61-68-7	2.49	3.26	Training	—	21
189	Acetohexamide	968-81-0	2.94	2.62	Test	0.047	21
190	Fludiazepam	3900-31-0	2.33	1.76	Training	—	21
191	Nimetazepam	2011-67-8	1.73	1.63	Training	—	21
192	Fenbufen	252-979-0	2.63	3.26	Training	—	21
193	Ketoprofen	22071-15-4	2.85	3.26	Training	—	21
194	Medazepam	2898-12-6	2.40	2.98	Training	—	21
195	Progabide	62666-20-0	2.53	2.80	Test	0.084	21
196	Griseofulvin	126-07-8	1.47	1.91	Training	—	21
197	Tolnaftate	2398-96-1	3.83	3.68	Training	—	21
198	Prostacyclin	35121-78-9	2.94	3.26	Training	—	21
199	Triamcinolone	124-94-7	3.37	3.92	Test	0.095	82
200	Cortisone	53-06-5	3.35	3.19	Training	—	21
201	Prednisolone	50-24-8	3.56	3.36	Test	0.105	82
202	Hydrocortisone	50-23-7	3.60	3.42	Training	—	21
203	Corticosterone	50-22-6	3.85	3.17	Test	0.106	82
204	Dexamethasone	50-02-2	3.65	4.48	Training	—	21
205	Betamethasone	378-44-9	3.73	4.03	Training	—	82
206	Paramethasone	53-33-8	3.40	3.32	Training	—	82
207	Cortisone-21-acetate	50-04-4	3.62	3.26	Training	—	82

(*Continued*)

Table 1. (*Continued*)

No.	Name	CAS	$\log K_o$	$\log K_p$	Partition	h	Refs.
208	Prednisolone-21-acetate	52-21-1	3.76	3.21	Training	—	82
209	Hydrocortisone-21-acetate	50-03-3	3.51	3.35	Training	—	82
210	Fluocinolone acetonide	67-73-2	3.48	3.60	Training	—	82
211	Triamcinolone acetonide	76-25-5	3.51	3.95	Training	—	82
212	Spironolactone	52-01-7	4.44	4.20	Training	—	82
213	Dehydrocholic acid	81-23-2	3.38	3.45	Training	—	82
214	Chenodeoxycholic acid	474-25-9	4.36[a]	4.45	Training	—	82
215	Ursodeoxycholic acid	128-13-2	4.51[a]	—	Training	—	82
216	Cholic acid	81-25-4	3.50	3.90	Test	0.184[b]	82
217	Hydrocortisone-17-butyrate	237-093-4	3.23	3.04	Training	—	82
218	Cinnarizine	298-57-7	3.64	3.24	Training	—	21
219	Cycloheptanol	502-41-0	3.23	2.63	Training	—	18
220	2-Methoxyethanol	109-86-4	0.22	0.29	Test	0.051	18
221	3-Hydroxycinnamic acid	588-30-7	2.56	2.52	Training	—	21
222	Ethyl 4-hydroxybenzoate	120-47-8	3.01	2.59	Training	—	21
223	Ethyl 4-aminobenzoate	94-09-7	2.69	2.64	Test	0.036	21
224	4-Methylcinnamic acid	1866-39-3	2.65	2.91	Training	—	21
225	Sulphadiazine	68-35-9	2.52	2.01	Test	0.057	21
226	L-α-*O*-benzylglycerol	213458-77-6	2.11	2.55	Test	0.051	81
227	Sulphamerazine	127-79-7	1.97	2.30	Training	—	21
228	Butyl 4-hydroxybenzoate	94-26-8	3.39	3.20	Training	—	21
229	Butyl 4-aminobenzoate	94-25-7	3.19	3.25	Training	—	21
230	Benzidine	92-87-5	3.35	3.64	Test	0.111	21
231	Triflumizole	68694-11-1	2.66	2.28	Training	—	21
232	Diazepam	439-14-5	2.33	1.97	Training	—	21
233	Prostaglandin E2	363-24-6	3.09	3.18	Training	—	21

*β-CD complex stability constant (K_o, observed, K_p, predicted), then log-transformed ($\log K$).
[a]Chemicals 214 and 215 were replaced by only one compound (chemical 214) with an averaged $\log K$ value (=4.44).
[b]Chemicals 1 and 216 have leverage values above the threshold (0.129) and, for that reason, its predictions were not taken into account when calculating Q^2_{EXT}.

5. Subject the derived QSPR model to rigorous internal and external validation, thereby assessing the performance of the model in what concerns its applicability and predictive power.
6. Compute the contribution of the different substructures to determine their quantitative contribution to the complexation of the studied molecules.

Variable Selection

Nowadays, there is a vast amount and wide range of molecular descriptors with which one can model the activity of interest. This makes the search for gathering the most suitable subset quite complicated and time consuming because of the many possible combinations, especially if one tries to define an accurate, robust, and (above all) interpretable model. For this reason, we applied the Genetic Algorithm (GA) procedure[47] for selecting the variables, as implemented in the Mobydigs software (v1.0).[48] The particular GA simulation applied here resorted to the generation of 100 regression models, ordered according to their increased internal predictive performance (verified by leave one out cross-validation). First of all, models with one to two variables were developed by the variable subset selection procedure in order to explore all low combinations. The number of descriptors was subsequently increased one by one, and new models formed. The GA was stopped when further increments in the size of the model did not increase internal predictivity in any significant degree. Furthermore, the following GA simulation conditions were used: the maximum number of variables in a model was 10, the number of best retained models for each size was 5, the trade-off between crossovers and mutation parameter (T) was from 0.3 to 0.7, and selection bias ($B\%$) was from 30 to 90.

Model Validation

Two kinds of diagnostic statistical tools were used for evaluating the performance of our regression model: the so-called goodness of fit and goodness of the prediction. In the first case, attention is given to the fitting properties of the model, whereas in the second case attention is paid to the predictive power of the model (i.e., the model adequacy for describing new compounds). In this work, *k*-Means Cluster Analysis (*k*-MCA) was used to split the original data set of chemicals into training and test sets. On doing so, 186 of the 233 compounds were selected as the training set and the remaining 47 taken as the external test set. Full details of this partition can be found in our previous work.[49]

Goodness of fit of the models was assessed by examining the determination coefficient (R^2), the standard deviation (s), the Fisher's ratio (F), and the ratio between the number of cases and the number of adjustable parameters in the model (known as the ρ statistics; notice that ρ should be ± 4).[50] Other important statistics, namely the Kubinyi function (FIT)[51,52] and Akaike's information criteria (AIC)[53,54] were taken into account, as they give enough criteria for comparing models with different parameters, numbers of variables, and numbers of chemicals.

As to the robustness and predictivity of the models, these were evaluated by means of cross-validation, basically leave-one-out (CV-LOO) and bootstrapping testing techniques, by looking to the outcome statistics of both techniques (i.e., Q^2_{LOO} and Q^2_{boot}) as well as to the Q^2_{EXT} values obtained with the test set substances that fall within the applicability domain of the model. Bootstrapping simulates what would happen if the data set were to be randomly resampled several times (here 5000 times), then deriving the all squared difference between the true and predicted responses by using predictive residual sum of squares (PRESS). The average predictive power is expressed as Q^2_{boot}.[55] Further, the stability under heavy perturbations in the training set was checked by examining the outcome statistics of a response randomisation procedure (Y scrambling) for the training and test sets ($\alpha(r^2)$ and $\alpha(Q^2)$ values). The randomisation procedure was repeated 300 times. All these calculations were carried out with software Mobydigs (v1.0).[48]

To sum up, good quality of the models is indicated by high F, FIT, and ρ values, by low s and AIC values, as well as by values closed to one for R^2, Q^2_{LOO}, Q^2_{boot}, and Q^2_{EXT} (save for $\alpha(r^2)$ and $\alpha(Q^2)$ values, which check random correlations).

The spectral moments are inherently collinear. From the point of view of QSPR modelling, the main drawback of collinearity is that it increases the standard errors associated with the individual regression coefficients, thereby decreasing their value for purposes of interpretability. To overcome this problem, we have employed here the Randić's method of orthogonalisation.[56–60] Firstly, one has to select the appropriate order of orthogonalisation, which, in this case, is the order of significance of the variables in the model. The first variable (v_1) is taken as the first orthogonal descriptor ($\Omega^1 v_1$). The second one (v_2) is orthogonalised with respect to it by taking the residual of its correlation with $\Omega^1 v_1$. The process is repeated until all variables are completely orthogonalised, after which they are further standardised. Orthogonal standardised variables are then used to obtain a new model. For extracting of the information contained in the orthogonalised descriptors, we followed the procedure reported by Estrada and Molina.[29]

Structural Alerts Identification

The identification of structural alerts (fragment contribution) to the β-CD complexation is based on bond contributions. This procedure, implemented in MODESLAB software, consists in transforming a QSAR/QSPR model into a bond additive scheme. Then, by summing up bonds contributions, one can detect the fragments on a given molecule that contribute positively or negatively to the underlying property and forward an interpretation of their effects in terms of physicochemical properties. Bond contributions are derived from the local spectral moments. They are defined as the diagonal entries of the different powers of the weighted bond matrix (B):

$$\mu_k^T(i) = b_{ii}^k(T) \tag{3}$$

where $\mu_k^T(i)$ is the kth local spectral moment of the bond i, $b_{ii}^k(T)$ are the diagonal entries of the weighted B matrix, and T is the type of bond weight. For a given molecule, we can substitute the values of the local spectral moments computed by Eq. (3) into Eq. (4) and thus gather the total contribution to the complexation of its different bonds

$$P = b_0 + \sum_k a_k \mu_k^T \tag{4}$$

Since the activity modelled is expressed as $\log K$, positive bond contributions increase

the K value and increase the complexation and vice versa. The structural information highlighted by the bond contributions may allow, along with other theoretical and experimental data, for a better understanding of the mechanisms of complexation of the involved chemicals.

Applicability Domain of the Models

Given that the real utility of a QSAR/QSPR model relies on its ability to accurately predict the modelled activity/property for new chemicals, careful assessment of the model's true predictive power is a must. This includes the model validation but also the definition of the applicability domain of the model in the space of molecular descriptors used for deriving the model. There are several methods for assessing the applicability domain of QSAR/QSPR models,[61,62] but the most common one encompasses determining the leverage values for each compound.[63] A Williams plot, that is, the plot of standardised residuals versus leverage values (h), can then be used for an immediate and simple graphical detection of both the response outliers and structurally influential chemicals in the model. In this plot, the applicability domain is established inside a squared area within $\pm x$ standard deviations and a leverage threshold h^* (h^* is generally fixed at $3\kappa/n$, where n is the number of training compounds and κ the number of model parameters, whereas $x = 2$ or 3), lying outside this are (vertical lines) the outliers and (horizontal lines) the influential chemicals. For future predictions, only predicted complex stability constant data for chemicals belonging to the chemical domain of the training set should be proposed and used.[64] So, calculations of Q^2_{EXT} were performed only for those substances that had a leverage value below the threshold h^*.

RESULTS AND DISCUSSION

QSPR Model

According to the strategy outlined before, we began by seeking the best linear model relating the complex stability with the TOPS-MODE descriptors for the training set. The resulting best-fit model (a 11-variable equation) is given below along with the MLR statistics. As seen, this model is good both statistical significance and goodness of fit.

$$\begin{aligned}\log K = &-1.44\times10^{-3}(\pm7.34\times10^{-5})\mu_1\mu_2^{\mathrm{Std}} + 3.95\times10^{-7}(\pm2.14\times10^{-8})\mu_{10}^{\mathrm{Std}}\\ &-1.50\times10^{-2}(\pm1.18\times10^{-3})\mu_5^{\mathrm{Ab-R_2}} + 0.42(\pm3.57\times10^{-2})\mu_1^{\mathrm{Hyd}}\\ &-0.25(\pm2.49\times10^{-2})\mu_1^{\mathrm{Dip2}} + 1.10\times10^{-2}(\pm1.52\times10^{-3})\mu_3^{\mathrm{Van}}\\ &+2.42\times10^{-4}(\pm3.90\times10^{-5})\mu_1\mu_4^{\mathrm{Dip2}} + 9.33\times10^{-3}(\pm1.68\times10^{-3})\mu_4^{\mathrm{Ab-log}L^{16}}\\ &+1.50\times10^{-2}(\pm3.21\times10^{-3})\mu_4^{\mathrm{Ab-}\sum\beta_2^{\mathrm{O}}} + 5.03\times10^{-7}(\pm1.62\times10^{-7})\mu_4^{\mathrm{Pols}}\\ &-0.55(\pm0.12)\end{aligned} \tag{5}$$

$$N = 185,\quad R^2 = 0.870,\quad Q^2_{\mathrm{LOO}} = 0.849,\quad s = 0.329,\quad F = 116.76,\quad \mathrm{AIC} = 0.122,\quad \mathrm{FIT} = 4.106,$$
$$Q^2_{\mathrm{boot}} = 0.825,\quad a(r^2) = 0.021,\quad a(Q^2) = -0.114,\quad Q^2_{\mathrm{EXT}} = 0.827$$

Another aspect deserving special attention is the degree of collinearity of the variables of the model, which can readily be diagnosed by analysing the cross-correlation matrix (Tab. 2). Rather than deleting any of these descriptors, it is of interest to examine the performance of orthogonal complements.

Following Randić's technique, we determined orthogonal complements for all variables of the above nonorthogonalised model. On doing so, variables $\Omega^2\mu_{10}^{\mathrm{Std}}$ and $\Omega^3\mu_5^{\mathrm{Ab-R_2}}$ were found to be not statistically significant ($p = 0.189$ and 0.496; Tab. 3), most likely because the information contained in these variables is common to the information contained in other molecular descriptors. In addition, the significance of adding these two variables to the model remains unclear as seen from the modest improvement in R^2 on going from step 8 to step 9 and to step 10 (see in Tab. 3, ΔR^2 for those steps). So after eliminating these uninformative variables, further proceed to refitting and

Table 2. Correlation matrix for Intercorrelations among the 10 Variables of the Initial Model (Eq. 5)

	μ_{10}^{Std}	μ_{1}^{Dip2}	μ_{1}^{Hyd}	μ_{4}^{Pols}	μ_{3}^{Van}	$\mu_{5}^{Ab-R_2}$	$\mu_{4}^{Ab-\sum\beta_2^O}$	$\mu_{4}^{Ab-\log L^{16}}$	$\mu_1\mu_2^{Std}$	$\mu_1\mu_4^{Dip2}$
μ_{10}^{Std}	1.000	—	—	—	—	—	—	—	—	—
μ_{1}^{Dip2}	**0.737**	1.000	—	—	—	—	—	—	—	—
μ_{1}^{Hyd}	−0.205	−0.062	1.000	—	—	—	—	—	—	—
μ_{4}^{Pols}	0.405	0.352	−0.333	1.000	—	—	—	—	—	—
μ_{3}^{Van}	**0.885**	**0.801**	0.059	0.516	1.000	—	—	—	—	—
$\mu_{5}^{Ab-R_2}$	**0.974**	**0.760**	−0.142	0.509	**0.945**	1.000	—	—	—	—
$\mu_{4}^{Ab-\sum\beta_2^O}$	**0.941**	**0.782**	−0.062	0.532	**0.980**	**0.989**	1.000	—	—	—
$\mu_{4}^{Ab-\log L^{16}}$	**0.929**	**0.761**	−0.033	0.540	**0.980**	**0.984**	**0.997**	1.000	—	—
$\mu_1\mu_2^{Std}$	**0.935**	**0.770**	0.003	0.442	**0.965**	**0.952**	**0.966**	**0.962**	1.000	—
$\mu_1\mu_4^{Dip2}$	**0.921**	**0.894**	−0.152	0.376	**0.878**	**0.916**	**0.904**	**0.885**	**0.907**	1.000

Significant correlations are marked in bold.

orthogonalisation, the following QSPR model was obtained:

$$\begin{aligned}\log K = {} & 0.379(\pm0.024)\Omega^1\mu_1\mu_2^{Std} + 0.509(\pm0.024)\Omega^4\mu_1^{Hyd} - 0.063(\pm0.025)\Omega^5\mu_1^{Dip2} \\ & + 0.475\times10^{-2}(\pm0.024)\Omega^6\mu_3^{Van} + 0.080(\pm0.025)\Omega^7\mu_1\mu_4^{Dip2} + 0.177(\pm0.025)\Omega^8\mu_4^{Ab-\log L^{16}} \\ & + 0.105(\pm0.024)\Omega^9\mu_4^{Ab-\sum\beta_2^O} + 0.078(\pm0.025)\Omega^{10}\mu_4^{Pols} + 2.537(\pm0.024)\end{aligned} \quad (6)$$

$$N = 185, \quad R^2 = 0.868, \quad Q^2_{LOO} = 0.851, \quad s = 0.329, \quad F = 145.61, \quad \text{AIC} = 0.12, \quad \text{FIT} = 4.656,$$
$$Q^2_{boot} = 0.845, \quad a(r^2) = 0.007, \quad a(Q^2) = -0.1, \quad Q^2_{EXT} = 0.8341$$

where the symbol ${}^i\Omega X$ means the orthogonal complement of variable X, the superscript referring to followed order in the orthogonalisation process.

Table 3. Step-by-Step Analysis of the Forward Stepwise Process

Step	Variable Included	R^2	ΔR^2	p-Level
1	${}^4\Omega\mu_1^{Hyd}$	0.323	0.323	3.25×10^{-17}
2	${}^6\Omega\mu_3^{Van}$	0.608	0.285	$2.56\times10^{-}$
3	$\Omega\mu_1\mu_2^{Std}$	0.798	0.191	6.26×10^{-28}
4	${}^8\Omega\mu_4^{Ab-\log L^{16}}$	0.833	0.035	5.21×10^{-9}
5	${}^9\Omega\mu_4^{Ab-\sum\beta_2^O}$	0.848	0.015	5.59×10^{-5}
6	${}^{10}\Omega\mu_4^{Pols}$	0.855	0.008	2.12×10^{-3}
7	${}^7\Omega\mu_1\mu_4^{Dip2}$	0.863	0.008	1.65×10^{-3}
8	${}^5\Omega\mu_1^{Dip2}$	0.869	0.005	8.35×10^{-3}
9	${}^2\Omega\mu_{10}^{Std}$	0.870	1.3×10^{-3}	**0.189**
10	${}^3\Omega\mu_5^{Ab-R_2}$	0.870	3.5×10^{-4}	**0.496**

Significant correlations are marked in bold.

As can be seen in Table 3, removal of $\Omega^2\mu_{10}^{Std}$ and $\Omega^3\mu_5^{Ab-R_2}$ had little effect on the overall fitness of the model as the statistics are as robust as before, and further, by comparing Eq. (5) with Eq. (6), one can see that there are no changes in either the sign of the regression coefficients. Nevertheless, the relative contributions of the variables in the orthogonal model are quite different from those related to the nonorthogonal model.

Their direct interpretation of these complex topological indices is rather difficult, considering that they essentially condense a large amount of topological and atomic property information into a single number. However, some indirect links between those descriptors and the physical phenomena involved in host–guest complexation might be suggested.

The variables weighted with hydrophobicity and van der Waals radii explained, respectively, 32.3% and 28.5% of the variance for this specific training set of chemicals (Tab. 3). Thus, hydrophobicity and van der Waals seem to be the main driving forces of the complexation of β-CDs for the molecules under study.

The variables weighted with standard distance, solute gas-hexadecane partition coefficient, effective hydrogen-bond basicity, polar surface, and dipole moment accounted for 19.1%, 3.5%, 1.5%, 0.8%, and 1.3% of the variance, respectively, therefore, although to a lesser extent, interactions due to the polarity (hydrogen bonding) also appear to influence complexation.

TOPS-MODE Structural Interpretation

Recently Katritzky et al.[25] presented a QSAR study predicting the free energies of inclusion complexation between diverse *guest* molecules and CDs using (i) CODESSA descriptors and (ii) counts of different molecular fragments. The first of them (the Hansch-type approach[65]) uses as descriptors certain physicochemical parameters calculated either by quantum mechanical methods or by some empirical techniques. The second (the Free-Wilson-type approach[66]) uses counts of different molecular fragments as variables in a multiple regression analysis. Both techniques have their advantages and disadvantages.[25] Generally, fragmental descriptors (Free-Wilson-type method) are more interpretable than CODESSA descriptors (Hansch-type method). However, the main disadvantage of QSPR methods based on counts of different molecular fragments is related to the fact that they generally use more variables than CODESSA descriptors, thus leading to smaller values of Fisher criterion (less robust models). Another problem of the fragment-based approach is related to molecules containing fragments of "rare" occurrence (i.e., found in a single molecule), which should be excluded from the training or test sets, thus reducing the number of treated compounds.[25] The last problem arises when we attempt to study heterogeneous data sets of organic molecules. In this case there is not necessarily an atomic/bond pattern, which is repeated in all the molecules under study. As a consequence is most adequate to use molecular descriptors like the electronic chemical potential, the molecular electronegativity, the chemical hardness, or other global molecular indices.

This question immediately poses another: can we obtain structural information at a local scale from the models developed using global molecular descriptors? The only information that we need to transform the global model into the atomic/bond contributions is the mathematical relationship between the global molecular descriptor and the local contributions.[67]

In this article, the TOPS-MODE approach has been used to account for the contributions of molecular parts to the global molecular properties. The main advances of using the TOPS-MODE approach to study complex stability constant between diverse guest molecules and β-CDs as compared with other approaches, such as CODESSA,[25] is twofold. On one hand, TOPS-MODE permits the development of robustness and predictive QSPR models in a similar way to those approaches using molecular descriptors, such as CODESSA. On the other hand, it permits the interpretation of the results in terms of fragment contribution identifying those groups, fragments, or molecular regions that can be responsible for the studied property in a similar way as fragmental descriptors does. To do this, fragmental descriptors needs to collect a significant amount of data for each kind of compounds while TOPS-MODE is able to recognise this structural pattern from only one compound present in the data set.[67] This is possible due to the nature of these descriptors. They describe the molecular structure as a whole in terms of hydrophobic, steric, and electronic characteristics of the molecules that can be transformed into local contributions. In addition to that, new hypothesis can be obtained with TOPS-MODE approach, which can form the basis for new structural interpretation after experimental confirmation.

Thus, TOPS-MODE approach let us to detect fragments that contribute positively or negatively to a particular target endpoint and their effects been interpreted in terms of physicochemical properties.[68] Specifically in our case, the contributions to the β-CD complex stability constant for each of the selected fragments (see Fig. 1) were extracted from the final orthogonal-descriptor model; these are shown in Table 4. A careful look at these values might allow us to find functional groups, fragments, or molecular regions that either hamper the inclusion phenomenon or enhance it. Further, it might lead us to design molecular structures that have a better profile for the phenomenon or to a rapid selection of the most favourable substance among a long list of substances.

The importance of hydrophobicity in predicting β-CD complexation is also demonstrated here if one looks at the contributions of fragments from F_9 to F_{15} and F_{19} to F_{21}, which have a large hydrophobic character. Clearly, their presence in

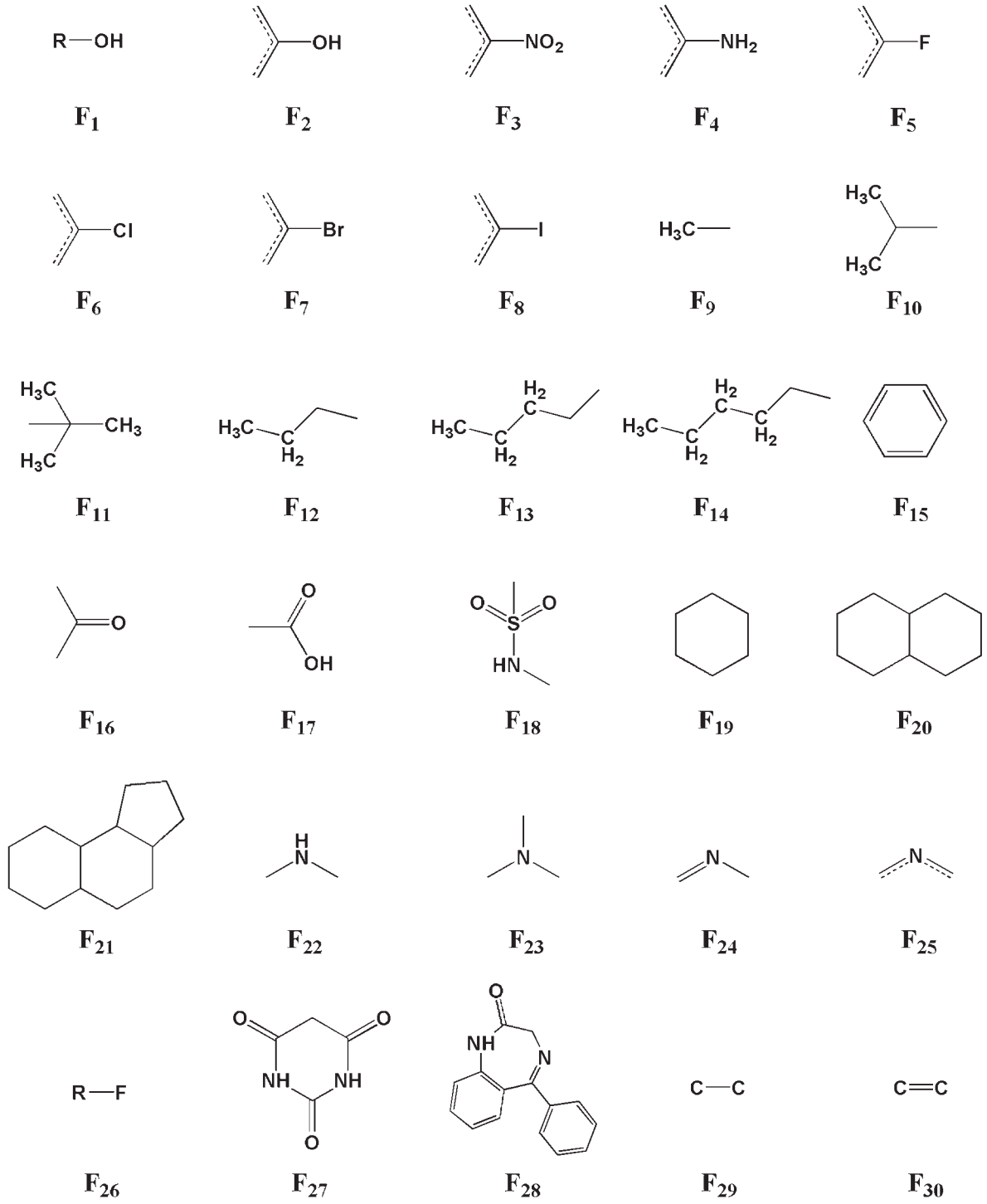

Figure 1. Selected molecular fragments (substructures) for which their contributions to the complexation with β-CD were calculated.

the molecule produces a significant improvement in the stability of the complex (Fig. 2). Indeed, an increase in the length of the hydrocarbon chain raises the stability of the complex because of the greater hydrophobic character of the molecule (fragments F_9, F_{12}, F_{13}, and F_{14}). Additionally, the amount of branching can affect the complexation (fragments F_9–F_{11}). A certain degree of branching

Table 4. The Contributions of Different Structural Fragments to the Complex Stability Constant

Fragment	Contribution
F_1	−0.361
F_2	−0.081
F_3	−0.062
F_4	0.048
F_5	−0.208
F_6	0.066
F_7	0.126
F_8	0.627
F_9	0.276
F_{10}	0.611
F_{11}	0.912
F_{12}	0.363
F_{13}	0.521
F_{14}	0.685
F_{15}	0.464
F_{16}	−0.410
F_{17}	−0.421
F_{18}	−1.078
F_{19}	0.598
F_{20}	0.911
F_{21}	1.454
F_{22}	−0.156
F_{23}	−0.116
F_{24}	−0.587
F_{25}	−0.178
F_{26}	0.042
F_{27}	−1.162
F_{28}	−0.558
F_{29}	0.152
F_{30}	0.064

may be necessary to achieve optimal van der Waals contacts with the β-CD interior. However, an excess of branching could lead to steric clashes between the compound and the β-CD interior. Furthermore, for fragments F_{19}, F_{20}, and F_{21}, one can see that an increase in the hydrophobic cyclic framework for steroids makes possible the complexation, thus facilitating oral, bucal, or transdermal administration for these highly insoluble molecules.[69–75]

On the other hand, the increasing flexibility or degrees of freedom in a guest molecule leads to a more favourable complexation entropy, since more of the possible "conformers" can fit properly into the cavity so the presence of unsaturated bonds reduces this flexibility and their chance of inclusion (fragments F_{29} and F_{30}).[76]

The negative contribution for oxygen and nitrogen containing groups (except aromatic amine) in the β-CD system can be assigned to the possibility of the competitive interactions with the solvent as discussed by Park and Nah.[20] One can also state, by taking into account the negative sign of contributions from fragments F_1 and F_2, that the presence of hydroxyl groups hinders inclusion in the β-CD. As we have previously deduced the phenomenon of inclusion in the β-CD for this set of molecules is dominated mainly by hydrophobic interactions, so the presence of hydrophilic groups diminishes the ability of the molecules to go into the hydrophobic cavity of β-CD (Fig. 2). Notice also the differences between alcoholic and phenolic groups. Maybe these are due to the fact that, even though aliphatic hydroxyl groups can form hydrogen bonds to the peripheral hydroxyls of CD, these interactions are not as strong as those formed by phenolic hydroxyl groups.[76]

For other hydrophilic groups such as amines (F_4, F_{22}, F_{23}, F_{24}, and F_{25}) something similar could happen. The aromatic amines form a hydrogen bond to the peripheral hydroxyls of the CD stronger than aliphatic amines (Fig. 3). It is worth to note the higher values of the amine groups regarding the hydroxyl groups. Possibly, this suggests better hydrogen bonding formed between guest's amine groups and the host's hydroxyl groups than that formed between guest's and host's hydroxyl groups or guest's amine groups may come to form hydrogen bonds with various host's hydroxyl groups.

On another fragments like halogenated derivatives of benzilic group (F_5–F_8), complexity is enhanced by increased volumes of substituent, confirming what has been observed by Liu and Guo[77] where the increased volume and polarisability of the guest substituent can increase the stability of the complex due to the stronger van der Waals interactions. It is important to note that the F_5 fragment (Ar-F) makes a negative contribution to complexation with the β-CD while the rest of halogenated aromatic fragments make a positive one. In the cases of the Ar-F fragment should be considered another additional intermolecular force of attraction: hydrogen bonding. The hydrogen bonding formed between Ar-F and water (solvent) could be more powerful than that formed between Ar-F and β-CD and then it could justify the negative contribution of F_5. Therefore, the presence of Ar-F fragment in a molecule decreases the stability of the complex.

By analysing whole of these contributions, one might explore other situations in which drugs with low activity can be enhanced if they have a

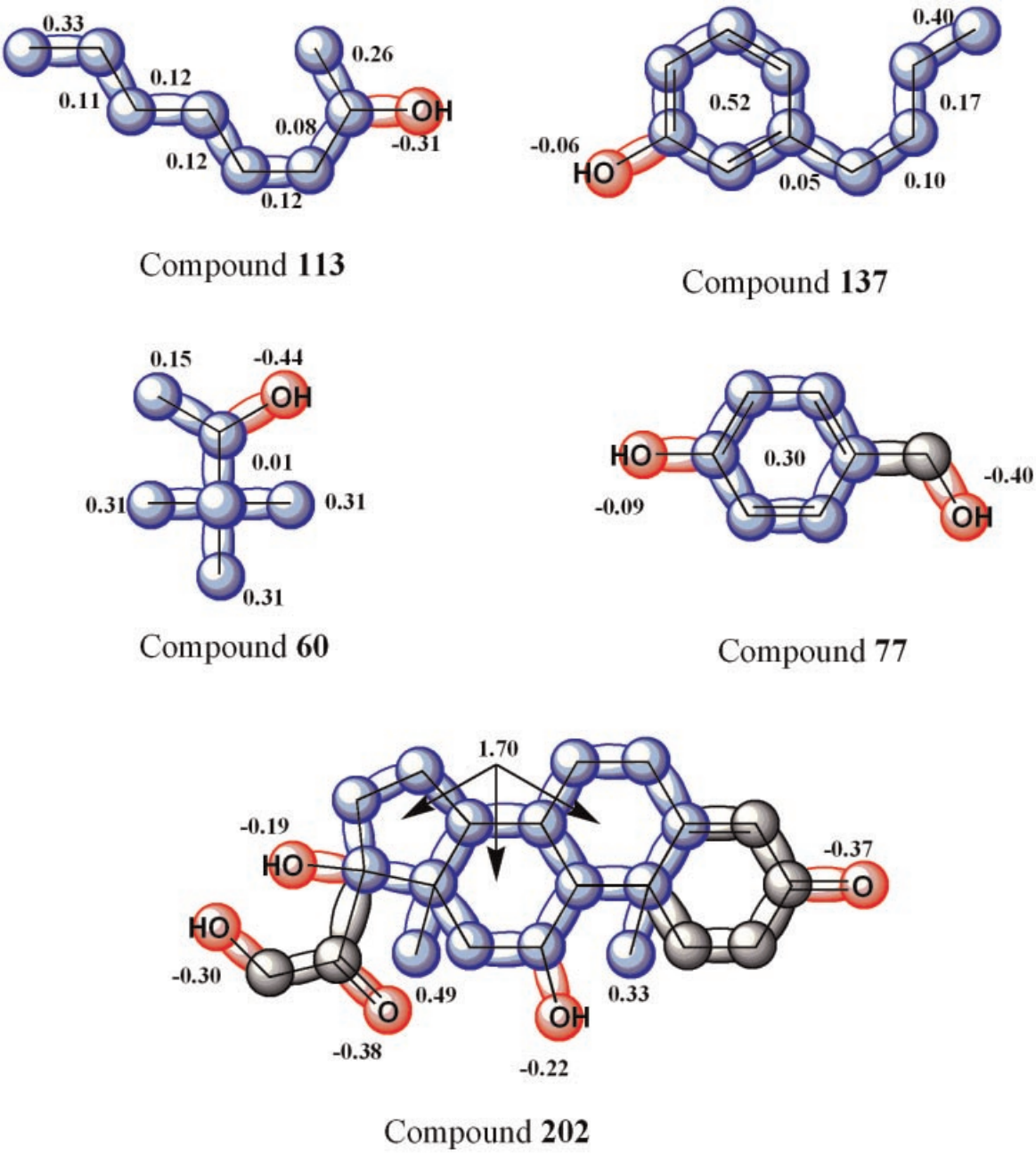

Figure 2. Contributions of the hydrophobic and hydroxyl groups. The red-coloured spheres represent the negative contributions to the complex stability constants whereas the blue spheres are those with positive contributions.

good complexation with the β-CD or facilitate their administration (as we have seen with steroids) or its bioavailability. For example, looking at the values of IC_{50} taste reported for benzodiazepines on the work of Sutherland et al.,[78] for which in compounds **191** and **232** (Fig. 4) the replacement of a nitro group by a chlorine entails a reduction of its power, but an increase in their complexation with β-CD (fragments F_3 and F_6).

Suzuki[16] found similar results when they used a group contribution model (GCM) for predicting free energies of complexation between guest molecules with β-CDs based on the same robust training set of 218 diverse ligands. In general, the presence of carbon, halogen, and sulphur results

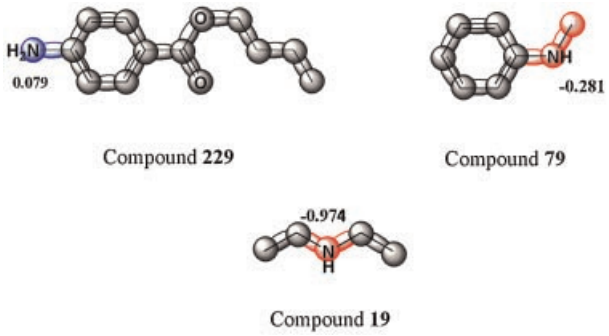

Figure 3. Contributions of the amine groups.

in an increase of the complex stability. In contrast, the presence of most oxygen and nitrogen containing groups (except –OH (phenol), –O–(ring), >C=O (ring), –COOH, $–NH_2$, and $–NO_2$) decreases this one. In this technique, a molecule is analysed for the presence of certain predefined fragments or functional groups. Each group has a specific contribution to the overall value of the binding free energy, which is obtained by summing those individual contributions. After that, specific group contributions are used as descriptors for generating the QSAR model. This technique limits the types of compounds that can be evaluated. A molecule that contains very little or none of the fragments in the model training set cannot be properly analysed. However, this concern is not an issue for our model. TOPS-MODE descriptors describe the molecular structure in a global way, permitting to find the contribution of any fragment in the molecular structure to the complexation with the β-CD. Thus, by using QSPR model and TOPS-MODE descriptors you easily obtain the contribution of any fragment in training/test set to the property, which is an advantage of this work. Actually, in Figure 1 and Table 4 we show some simple molecular fragments (F_1–F_9, F_{16}, F_{17}, F_{22}–F_{25}) similar to Suzuki et al., and their fragment contributions, respectively. But we can also obtain new and more complex hypothesis (e.g., F_{21}, F_{18}, F_{27}, F_{28}) with TOPS-MODE approach, which can form the basis for new structural interpretation after experimental confirmation.

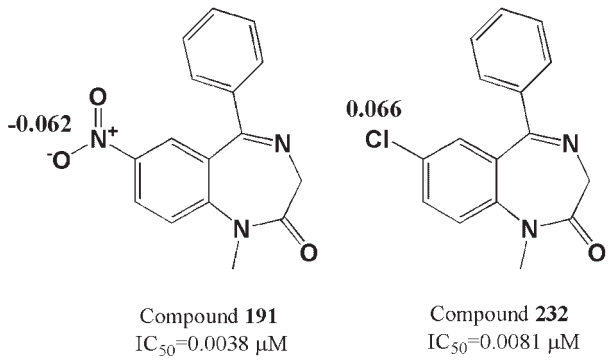

Figure 4. Fragment contributions of two benzodiazepines.

Applicability Domain

It would be very interesting to have a predictive model for the vast majority of chemicals, particularly for those which have not been tested and, therefore, with unknown $\log K$ values. Since this is usually not possible, one should define the applicability domain of the QSPR model, that is, the range within which the model bears a new compound.

For that purpose, we built a Williams plot using the leverage values calculated for each compound. As seen in Figure 5, most of the compounds of the test set are within the applicability domain covered by ± 3 times the standard residual (σ) and the leverage threshold h^* (=0.129), save for compounds 31, 53, 175, 197, 202, 204, 205, 207, 211, 214, 218, and 233. Even so, the latter should not be considered outliers but influential chemicals.[61]

Nevertheless, all evaluations pertaining to the external set were performed by taking into account the applicability domain of our QSPR model. So, if a chemical belonging to the test set

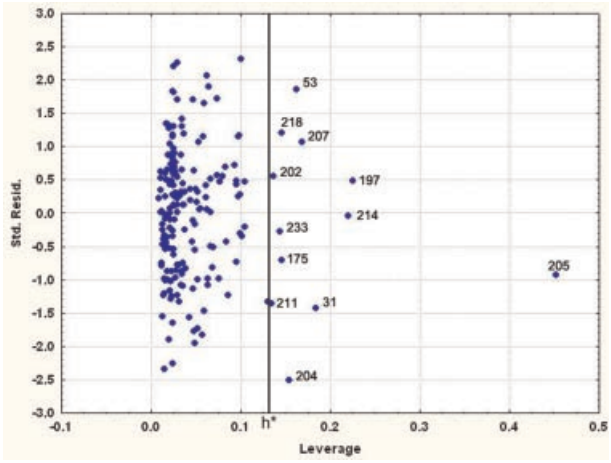

Figure 5. Williams plot based on Eq. (6), that is, plot of standardised residuals versus leverage values with a warning leverage of $h^* = 0.129$.

had a leverage value greater than h^*, we consider that this means that the prediction is the result of substantial extrapolation and therefore may not be reliable.[62]

CONCLUSION

Due to the beneficial effects arising from the complexation of drugs with β-CDs, we have applied here a QSPR regression-based approach to a diverse set of 233 organic compounds with known complex stability constant (K) values. By means of k-MCA, 80% of these compounds were selected as the training set and the remaining as the external evaluation set. With regard to the QSPR modelling, the combination of multivariate data analysis in conjunction with a TOPS-MODE representation and the genetic selection algorithm was found to produce a final regression model with good accuracy, internal cross-validation statistics, and predictivity on the external data.

The analysis of the most frequent descriptors implicated in the final QSPR model afforded model interpretation in terms of chemical features influencing complexation with β-CD. The major driving forces for complexation, extracted from the model, were hydrophobicity and van der Waals interactions, and thus the presence of hydrophobic groups (hydrocarbon chains, aryl groups, etc.) and voluminous species (Cl, Br, I, etc.) in the molecule facilitate their complexation by β-CD, while possibly increasing the beneficial effects (solubility and bioavailability) derived from this. The final QSPR model was further used to collect effective information about what kinds of groups favour such complexation.

In summary, the information gathered by these descriptors given in the form of bond contributions provide valuable information for future use in drug design and other applications related to complexation with β-CDs.

ACKNOWLEDGMENTS

The authors acknowledge to MODESLAB 1.0 software owners for delivering a free copy of such program and the anonymous reviewers for their-comments. A.M.H. acknowledges the *Portuguese Fundaão para a Ciência e a Tecnologia* (FCT, Lisboa) (SFRH/BD/22692/2005) for financial support.

REFERENCES

1. Saenger W, Jacob J, Gessler K, Steiner T, Daniel S, Sanbe H, Koizumi K, Smith SM, Tanaka T. 1998. Structures of the common cyclodextrins and their larger analogues—Beyond the doughnut. Chem Rev 98:1787–1802.
2. Loftsson T, Brewster M, Masson M. 2004. Role of cyclodextrins in improving oral drug delivery. Am J Drug Deliv 2:261–275.
3. Davis ME, Brewster M. 2004. Cyclodextrin-based pharmaceutics: Past, present and future. Nat Rev Drug Discov 3:1023–1035.
4. Avdeef A, Bendels S, Tsinman O, Tsinman K, Kansy M. 2007. Solubility excipient classification gradient maps. Pharm Res 24:530–545.
5. Kim C, Park J. 2004. Solubility enhancers for oral drug delivery. Am J Drug Deliv 2:113–130.
6. Loftsson T, Jarho P, Masson M, Jarvinen T. 2005. Cyclodextrins in drug delivery. Expert Opin Drug Deliv 2:335–351.
7. Kola I, Landis J. 2004. Can the pharmaceutical industry reduce attrition rates? Nat Rev Drug Discov 2:711–715.
8. Liu R. 2000. Water-insoluble drug formulation. Englewood: CO Interpharm Press.
9. Irie T, Uekama K. 1997. Pharmaceutical applications of cyclodextrins. iii. Toxicological issues and safety evaluation. J Pharm Sci 86:147–162.
10. Szejtli J. 1998. Introduction and general overview of cyclodextrin chemistry. Chem Rev 98:1743–1753.
11. Lantz A, Rodriguez M, Wetterer S, Armstrong D. 2006. Estimation of association constants between oral malodor components and various native and derivatized cyclodextrins. Anal Chim Acta 557:184–190.
12. Uekama K. 1999. Cyclodextrins in drug delivery. Adv Drug Deliv Rev 36:1–2.
13. Duchêne D, editor. 1987. Cyclodextrins and their industrial uses. Paris: Editions de Santé Paris.
14. Horvath G, Premkumar T, Boztas A, Lee E, Jon S, Geckeler KE. 2008. Supramolecular nanoencapsulation as a tool: Solubilization of the anticancer drug trans-dichloro(dipyridine)platinum(ii) by complexation with beta-cyclodextrin. Mol Pharm 5:358–363.
15. Lipkowitz KB. 1998. Applications of computational chemistry to the study of cyclodextrins. Chem Rev 98:1829–1873.
16. Suzuki T. 2001. A nonlinear group contribution method for predicting the free energies of inclusion complexation of organic molecules with α- and β-cyclodextrins. J Chem Inf Comput Sci 41:1266–1273.
17. Pérez F, Jaime C, Sánchez-Ruiz X. 1995. Mm2 calculations on cyclodextrins: Multimodel inclusion complexes. J Org Chem 60:3840–3845.

18. Matsui Y, Nishioka T, Fujita T. 1985. Quantitative structure-reactivity analysis of the inclusion mechanism by cyclodextrins. Top Curr Chem 128: 61–89.
19. Davis DM, Savage JR. 1993. Correlation analysis of the host-guest interaction of α-cyclodextrin and substituted benzenes. J Chem Res-S 94–95.
20. Park JH, Nah TH. 1994. Binding forces contributing to the complexation of organic molecules with β-cyclodextrin in aqueous solution. J Chem Soc [Perkin Trans] 2:1359–1362.
21. Klein CT, Polheim D, Viernstein H, Wolschann P. 2000. A method for predicting the free energies of complexation between β-cyclodextrin and guest molecules. J Inclusion Phenom Macrocyclic Chem 36:409–423.
22. Liu L, Guo QX. 1999. Wavelet neural network and its application to the inclusion of β-cyclodextrin with benzene derivatives. J Chem Inf Comput Sci 39:133–138.
23. Suzuki T, Ishida M, Fabian WMF. 2000. Classical QSAR and comparative molecular field analyses of the host-guest interaction of organic molecules with cyclodextrins. J Comput Aided Mol Des 14:669–678.
24. Cramer IRD, Patterson DE, Bunce JD. 1988. Comparative molecular field analysis (COMFA). 1. Effect of shape on binding of steroids to carrier proteins. J Am Chem Soc 110:5959–5967.
25. Katritzky AR, Fara DC, Yang HF, Karelson M, Suzuki T, Solov'ev VP, Varnek A. 2004. Quantitative structure-property relationship modelling of β-cyclodextrin complexation free energies. J Chem Inf Comput Sci 44:529–541.
26. Estrada E. 1996. Spectral moments of the edge adjacency matrix in molecular graphs. 1. Definition and applications to the prediction of physical properties of alkanes. J Chem Inf Comput Sci 36:844–849.
27. Estrada E. 1997. Spectral moments of the edge-adjacency matrix of molecular graphs. 2. Molecules containing heteroatoms and QSAR applications. J Chem Inf Comput Sci 37:320–328.
28. Estrada E. 1995. Edge adjacency relationships and a novel topological index related to molecular volume. J Chem Inf Comput Sci 35:31–33.
29. Estrada E, Molina E. 2006. Automatic extraction of structural alerts for predicting chromosome aberrations of organic compounds. J Mol Graph Model 25:275–288.
30. Estrada E, Patlewicz G, Gutierrez Y. 2004. From knowledge generation to knowledge archive. A general strategy using tops-mode with Derek to formulate new alerts for skin sensitization. J Chem Inf Comput Sci 44:688–698.
31. González MP, Dias L, Helguera AM. 2004. A topological sub-structural approach to the mutagenic activity in dental monomers. 2. Cycloaliphatic epoxides. Polymer 15:5353–5359.
32. González MP, Helguera AM, Molina R, Garca JR. 2004. A topological substructural approach of the mutagenic activity in dental monomers. 1. Aromatic epoxides. Polymer 45:2773–2779.
33. González MP, Helguera AM, Cabrera MA. 2005. Quantitative structureactivity relationship to predict toxicological properties of benzene derivative compounds. Bioorg Med Chem 13:1775–1781.
34. Helguera AM, Pérez MAC, Combes RD, González MP. 2006. Quantitative structure activity relationship for the computational prediction of nitrocompounds carcinogenicity. Toxicology 220:51–62.
35. Helguera AM, González MP, Cordeiro MNDS, Cabrera MA. 2007. Quantitative structure carcinogenicity relationship for detecting structural alerts in nitrosocompounds. Toxicol Appl Pharmacol 221: 189–202.
36. Helguera AM, González MP, Cordeiro MNDS, Cabrera MA. 2008. Quantitative structure-carcinogenicity relationship for detecting structural alerts in nitroso compounds: Species, rat; sex, female; route of administration, gavage. Chem Res Toxicol 21: 633–642.
37. González MP, Terán C, Teijeira M. 2006. A topological function based on spectral moments for predicting affinity toward a3 adenosine receptors. Bioorg Med Chem Lett 16:1291–1296.
38. González MP, Helguera AM, Collado IG. 2006. A topological substructural molecular design to predict soil sorption coefficients for pesticides. Mol Divers 10:109–118.
39. González MP, Diáz HG, Ruiz RM, Cabrera MA, Ramos de Armas R. 2003. Tops-mode based QSARs derived from heterogeneous series of compounds. Applications to the design of new herbicides. J Chem Inf Comput Sci 43:1192–1199.
40. Pérez-Garrido A, González MP, Escudero AG. 2008. Halogenated derivatives QSAR model using spectral moments to predict haloacetic acids (haa) mutagenicity. Bioorg Med Chem 16:5720–5732.
41. Helguera AM, Cordeiro MNDS, Cabrera MA, Combes RD, González MP. 2008. Quantitative structure carcinogenicity relationship for detecting structural alerts in nitroso-compounds species: Rat; sex: male; route of administration: water. Toxicol Appl Pharmacol 231:197–207.
42. Helguera AM, Pérez MAC, González MP, Ruiz RM, Díaz HG. 2005. A topological substructural approach applied to the computational prediction of rodent carcinogenicity. Bioorg Med Chem 13: 2477–2488.
43. Environment Directorate OECD. 2007. Guidance Document of the Validation of (Quantitative) Structure-Activity Relationships (Q)SAR Models. Enviromental Health and Safety Publications, Series on Testing and Assessment No. 69.

44. Estrada E, Uriarte E, Gutierrez Y, González H. 2003. Quantitative structure-toxicity relationships using tops-mode. 3. Structural factors influencing the permeability of commercial solvents through living human skin. SAR QSAR Environ Res 14: 145–163.
45. Gutierrez Y, Estrada E. 2002. Modes Lab, Version 1.0.
46. Weininger D. 1988. Smiles, a chemical language and information system. 1. Introduction to methodology and encoding rules. J Chem Inf Comput Sci 28:31–36.
47. Goldberg D. 1989. Genetic algorithms in search, optimization, and machine learning. USA: Addison-Wesley.
48. Todeschini R, Ballabio D, Consonni V, Mauri A, Pavan M. 2004. Mobydigs Computer Software. Milano: TALLETE SRL.
49. Pérez-Garrido A, Helguera AM, Cordeiro MNDS, Abellán A, Escudero AG. 2008. Convenient QSAR model for predicting the complexation of structurally diverse compounds with β-cyclodextrins. Bioorg Med Chem 17:896–904.
50. Garcia-Domenech R, Julian-Ortiz JV. 1998. Antimicrobial activity characterization in a heterogeneous group of compounds. J Chem Inf Comput Sci 38:445–449.
51. Kubinyi H. 1994. Variable selection in QSAR studies. 1. An evolutionary algorithm. Quant Struct Act Relat 13:285–294.
52. Kubinyi H. 1994. Variable selection in QSAR studies. 2. A highly efficient combination of systematic search and evolution. Quant Struct Act Relat 13: 393–401.
53. Akaike H. 1973. Information theory and an extension of the maximum likelihood principle. In Proceedings of the Second International Symposium on Information Theory. Budapest: Akademiai Kiado, pp 267–281.
54. Akaike H. 1974. New look at statistical-model identification. IEEE Trans Automat Control AC-19: 716–723.
55. Lučić B, Nikolić S, Trinajstić N, Jurić D. 1995. The structure-property models can be improved using the orthogonalized descriptors. J Chem Inf Comput Sci 35:532–538.
56. Todeschini R, Consonni V. 2000. Handbook of molecular descriptors. Mannheim: Wiley-VCH, 667p.
57. Klein D, Randić M, Babić D, Lučić B, Nikolić S, Trinajstić N. 1997. Hierarchical orthogonalization of descriptors. Int J Quantum Chem 63:215–222.
58. Randić M. 1991. Orthogonal molecular descriptors. N J Chem 15:517–525.
59. Randić M. 1991. Resolution of ambiguities in structure-property studies by use of orthogonal descriptors. J Chem Inf Comput Sci 31:311–320.
60. Randić M. 1991. Correlation of enthalpy of octanes with orthogonal connectivity indices. J Mol Struct (Theochem) 233:45–59.
61. Eriksson L, Jaworska J, Worth AP, Cronin MTD, McDowell RM, Gramatica P. 2003. Methods for reliability and uncertainty assessment and for applicability evaluations of classification- and regression-based QSARs. Environ Health Perspect 111:1361–1375.
62. Netzeva TI, Worth AP, Aldenberg T, Benigni R, Cronin MTD, Gramatica P, Jaworska JS, Kahn S, Klopman P, Marchant CA, Myatt G, Nikolova-Jeliazkova N, Patlewicz GY, Perkins R, Roberts DW, Schultz TW, Stanton DT, van de Sandt JJH, Tong W, Veith G, Yang C. 2005. Current status of methods for defining the applicability domain of (quantitative) structure activity relationships. ATLA 33:155–173.
63. Gramatica P. 2007. Principles of QSAR models validation: Internal and external. QSAR Comb Sci 26:1–9.
64. Vighi M, Gramatica P, Consolaro F, Todeschini R. 2001. QSAR and chemometrics approaches for setting water quality objectives for dangerous chemicals. Ecotoxicol Environ Saf 49:206–220.
65. Hansch C, Leo A, Hoekman DH. 1995. Exploring QSAR fundamentals and applications in chemistry and biology. In ACS professional reference book. Washington, DC: American Chemical Society, p 580.
66. Free SM, Wilson JW. 1964. A mathematical contribution to structure-activity studies. J Med Chem 7:395–399.
67. Estrada E. 2008. How the parts organize in the whole? A top-down view of molecular descriptors and properties for QSAR and drug design. Mini Rev Med Chem 8:213–221.
68. Benigni R, Giuliani A. 2003. Putting the predictive toxicology challenge into perspective: Reflections on the results. Bioinformatics 19:1194–1200.
69. Seo H, Tsuruoka M, Hashimoto T, Fujinaga T, Otagiri M, Uekama K. 1983. Enhancement of oral bioavailability of spironolactone by betacyclodextrin and gamma-cyclodextrin complexations. Chem Pharm Bull 31:286–291.
70. Pitha J, Harman SM, Michel ME. 1986. Hydrophilic cyclodextrin derivatives enable effective oral-administration of steroidal hormones. J Pharm Sci 75:165–167.
71. Uekama K, Fujinaga T, Otagiri M, Seo H, Tsuruoka M. 1981. Enhanced bioavailability of digoxin by gamma-cyclodextrin complexation. J Pharmacobiodyn 4:735–737.
72. Uekama K, Fujinaga T, Hirayama F, Otagiri M, Yamasaki M, Seo H, Hashimoto T, Tsuruoka M. 1983. Improvement of the oral bioavailability of digitalis glycosides by cyclodextrin complexation. J Pharm Sci 72:1338–1341.

73. Taylor GT, Weiss J, Pitha J. 1989. Testosterone in a cyclodextrin containing formulation—Behavioral and physiological-effects of episode like pulses in rats. Pharm Res 6:641–646.
74. Loftsson T, Olafsdottir BJ, Bodor N. 1991. The effects of cyclodextrins on transdermal delivery of drugs. Eur J Pharm Biopharm 37:30–33.
75. Uekama K, Arimori K, Sakai A, Masaki K, Irie T, Otagiri M. 1987. Improvement in percutaneous-absorption of prednisolone by betacyclodextrin and gamma-cyclodextrin complexations. Chem Pharm Bull 35:2910–2913.
76. Rekharsky MV, Inoue Y. 1998. Complexation thermodynamics of cyclodextrins. Chem Rev 98:1875–1917.
77. Liu L, Guo QX. 2002. The driving forces in the inclusion complexation of cyclodextrins. J Inclusion Phenom Macrocyclic Chem 42:1–14.
78. Sutherland JJ, O'Brien LA, Boztas A, Weaver DF. 2003. Spline-fitting with a genetic algorithm: A method for developing classification structure-activity relationships. J Chem Inf Comput Sci 43: 1906–1915.
79. Inoue Y, Hakushi T, Liu Y, Tong LH, Shen BJ, Jin DS. 1993. Thermodynamics of molecular recognition by cyclodextrins. 1. Calorimetric titration of inclusion complexation of naphthalenesulfonates with α-, β-, and γ-cyclodextrins: Enthalpy-entropy compensation. J Am Chem Soc 115:475–481.
80. Carpignano R, Marzona M, Cattaneo E, Quaranta S. 1997. QSAR study of inclusion complexes of heterocyclic compounds with β-cyclodextrin. Anal Chim Acta 348:489–493.
81. Rekharsky MV, Goldberg RN, Schwarz FP, Tewari YB, Ross PD, Yamashoji Y, Inoue Y. 1995. Thermodynamic and nuclear magnetic resonance study of the interactions of α- and β-cyclodextrin with model substances: Phenethylamine, ephedrines, and related substances. J Am Chem Soc 117:8830–8840.
82. Wallimann P, Marti T, Fürer A, Diederich F. 1997. Steroids in molecular recognition. Chem Rev 97: 1567–1608.