A topological substructural molecular design approach for predicting mutagenesis end-points of α, β-unsaturated carbonyl compounds

ARTICLE *in* TOXICOLOGY · DECEMBER 2009

Impact Factor: 3.62 · DOI: 10.1016/j.tox.2009.11.023 · Source: PubMed

5 AUTHORS, INCLUDING:

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

Aliuska Morales Helguera
50 PUBLICATIONS **1,043** CITATIONS
SEE PROFILE

Gabriel Caravaca López
Universidad Católica San Antonio de Murcia
8 PUBLICATIONS **75** CITATIONS
SEE PROFILE

Natália D. S. Cordeiro
University of Porto
245 PUBLICATIONS **2,911** CITATIONS
SEE PROFILE

Toxicology 268 (2010) 64–77

Contents lists available at ScienceDirect

Toxicology

journal homepage: www.elsevier.com/locate/toxicol

A topological substructural molecular design approach for predicting mutagenesis end-points of α, β-unsaturated carbonyl compounds

Alfonso Pérez-Garrido[a,b,*], Aliuska Morales Helguera[c,d,e], Gabriel Caravaca López[b], M.Natália D.S. Cordeiro[e], Amalio Garrido Escudero[a]

[a] Enviromental Engineering and Toxicology Dpt., Catholic University of San Antonio, Guadalupe, Murcia, C.P. 30107, Spain
[b] Department of Food and Nutrition Technology, Catholic University of San Antonio, Guadalupe, Murcia, C.P. 30107, Spain
[c] Department of Chemistry, Faculty of Chemistry and Pharmacy, Central University of Las Villas, Santa Clara, 54830 Villa Clara, Cuba
[d] Molecular Simulation and Drug Design Group, Chemical Bioactive Center, Central University of Las Villas, Santa Clara, 54830 Villa Clara, Cuba
[e] REQUIMTE, Chemistry Department, Faculty of Sciences, University of Porto, 4169-007 Porto, Portugal

ARTICLE INFO

Article history:
Received 20 October 2009
Received in revised form
29 November 2009
Accepted 30 November 2009
Available online 11 December 2009

Keywords:
QSAR
Mutagenicity
TOPS-MODE
α, β-Unsaturated carbonyl compounds
REACH

ABSTRACT

Chemically reactive, α, β-unsaturated carbonyl compounds are common environmental pollutants able to produce a wide range of adverse effects, including, e.g. mutagenicity. This toxic property can often be related to chemical structure, in particular to specific molecular substructures or fragments (alerts), which can then be used in specialized software or expert systems for predictive purposes. In the past, there have been many attempts to predict the mutagenicity of α, β-unsaturated carbonyl compounds through quantitative structure activity relationships (QSAR) but considering only one exclusive endpoint: the Ames test. Besides, even though those studies give a comprehensive understanding of the phenomenon, they do not provide substructural information that could be useful forward improving expert systems based on structural alerts (SAs). This work reports an evaluation of classification models to probe the mutagenic activity of α, β-unsaturated carbonyl compounds over two endpoints – the Ames and mammalian cell gene mutation tests – based on linear discriminant analysis along with the topological Substructure molecular design (TOPS-MODE) approach. The obtained results showed the better ability of the TOPS-MODE approach in flagging structural alerts for the mutagenicity of these compounds compared to the expert system TOXTREE. Thus, the application of the present QSAR models can aid toxicologists in risk assessment and in prioritizing testing, as well as in the improvement of expert systems, such as the TOXTREE software, where SAs are implemented.

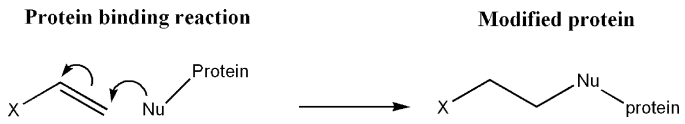
1. Introduction

α, β-Unsaturated carbonyl compounds are common environmental pollutants, often used in the synthesis of chemicals, solvents, food additives, disinfectants and dental restorative materials (Feron et al., 1991; Boelens and Gemert, 1987; van Noort et al., 1990). These compounds possess a strongly polarized carbon-oxygen double bond due to the presence of an additional double bond between carbons 2 and 3 (i.e. α and β), which makes them even more reactive than simple carbonyls. Because of their particular reactivity, they are able to interact with electron-rich biological macromolecules and a wide range of adverse effects has been reported, including for instance mutagenicity. In general, these mutagenic compounds act through a Michael type addition mechanism (see Fig. 1) but the type of substituents in the α or β-carbons of the unsaturated carbonyl moiety significantly affects the effectiveness of the reaction (Aptula and Roberts, 2006).

Regulatory bodies use various endpoints as standard tests for screening chemicals for potential mutagenic effects. The four common assay types employed are bacterial mutagenesis, mammalian mutagenesis, *in vitro* chromosome aberration and *in vivo* micronucleous. All have in common the ability to identify those substances that produce some sort of alteration on DNA. Owing to the cost in both resources and time required in such mutagenic assays, there has been a remarkable upsurge in interest in alternative non-animal approaches as tools for speeding up, at least, priority setting and risk assessment. One such set of tools, strongly encouraged under the framework of the European Union's REACH (Registration, Evaluation, Authorisation and restriction of CHemicals) legislation, comprises in silico prediction of mutagenicity, based on (Quantitative) Structure–Activity Relationships [(Q)SAR] modelling (OECD, 2007). QSAR modelling seeks to discover and use mathematical relationships between molecular properties of the compounds (descriptors) and the activity or property of interest.

* Corresponding author at: Enviromental Engineering and Toxicology Dpt., Catholic University of San Antonio, Guadalupe, Murcia, C.P. 30107, Spain.
Tel.: +34 968 278 755.
E-mail address: Aperez@pdi.ucam.edu (A. Pérez-Garrido).

0300-483X/$ – see front matter © 2009 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.tox.2009.11.023



Double or triple bond with electron-withdrawing substituent X, such as –CHO, -COR, -CO_2R, -CN, -SO_2R, -NO_2, etc.Includes ortho- or para-quinones, often formed by oxidation of ortho- or para-dihydroxy aromatics acting as pro-Michael acceptors. X can also be a heterocyclic group such as ortho- or para-pyridino.

Fig. 1. Michael type addition reaction and corresponding applicability domain.

In the past, there have been several QSAR studies aimed at modelling the mutagenicity of α, β-unsaturated carbonyl compounds. These QSARs can be subdivided into two types of models, i.e. for (1) the gradation of potency of active chemicals (Yourtee et al., 2001; Holder and Ye, 2009; Helguera et al., 2004; González et al., 2005b and for (2) the discrimination between active (positive) and inactive (negative) chemicals (Benigni et al., 2005; Koleva et al., 2008). One should notice here that, previous work has shown that the structural effects that modulate the mutagenic potency are normally different from those that distinguish positive/negative active chemicals, and so the first type of models are not useful to set up mutagenic activity the first issue in risk assessment (Benigni et al., 1998).

The only attempts towards modelling the mutagenic activity of this family of substances have been based on data collected from the Salmonella typhimurium Ames test. Benigni et al. (2005) developed a QSAR prediction model for 25 α, β-unsaturated aldehydes by stepwise linear discriminant analysis based on data from TA100 strain assays. The results of this study indicated a dependency between mutagenicity, hydrophobicity and molecular volume. More recently, a study has appeared to model and predict the Ames TA100-derived mutagenicity for 45 α, β-unsaturated carbonyl compounds (Koleva et al., 2008). In this study, the set of compounds was divided into several sub-groups, namely: (1) halogenated derivatives, (2) nitro derivatives of cinnamaldehyde and (3) related acroleins, and a system of rules was developed for the classification based on their reactivity and mechanisms of action.

All the above studies only considered as endpoint the Ames test for estimating the mutagenic activity. Even though that is understandable for two of the other endpoints – *in vitro* chromosome aberration and *in vivo* micronucleus – since they do not provide enough data (only for 6 and 7 compounds, respectively) to set up valid QSAR models, that is not so for the MGGM endpoint (data for 45 compounds). What's more, information across multiple endpoints seems to be needed to reach more realistic predictions about mutagenicity (Yang et al., 2008). In addition, those studies resorted to global molecular descriptors (i.e.: molecular refractivity, the logarithm of the octanol/water partition coefficient log *P*, etc.) and many of them were unable to perceive how each fragment or functional group influence a particular molecular structure of interest. One way to overcome the latter problem is to use Structural Alerts-based SAR studies such as the recently implemented expert system: TOXTREE (Benigni and Bossa, 2008; Benigni et al., 2008). Structural Alerts (SA) are molecular substructures or functional groups that are related to the toxic properties of the chemicals, that is to say, a sort of "codes" embodying long series of studies aimed at highlighting their mechanisms of action. This SA-based expert system has already pulled together very good results in predictions of mutagenic/carcinogenic chemicals (Benigni and Bossa, 2008). Even so, it has been recognised also that further work is still required to improve the knowledge about modulating factors (Benigni and Bossa, 2008; Benigni et al., 2008), that is to say, about the chemical functionalities that may annihilate the toxic effects of the SAs when they are present simultaneously in the same molecule.

This work aims at discriminating the mutagenic activity of α, β-unsaturated carbonyl compounds, based on two different endpoints together with a substructural approach such as the TOPological Substructural MOlecular Design (TOPS MODE) (Estrada, 1996, 1997, 1995) descriptors. These descriptors can and have proved to be very useful in QSAR modelling of a broad range of toxicities (Estrada et al., 2001, 2003a,b, 2004; Estrada and Uriarte, 2001; González et al., 2005a, 2006; Helguera et al., 2005, 2006, 2007, 2008a,b; Estrada and Molina, 2006; García-Lorenzo et al., 2008; Sosted et al., 2004), including mutagenicity (González et al., 2004a,b, 2005b; Helguera et al., 2004; Pérez-Garrido et al., 2008). Furthermore, the TOPS MODE approach is able to transform simple molecular descriptors, such as Log *P*, polar surface area, charges, etc., into series of descriptors that account for the distribution of the related characteristics (hydrophobicity, polarity, electronic effects, etc.) across the molecule. In fact, such approach has been recognised recently by the Organisation for Economic Co-operation and Development (OECD) as providing "*a mechanistic interpretation at a bond level*" and enabling "*the generation of new hypotheses such as structural alerts*" (OECD, 2007). Thus, by gathering structural information at a local level from the models developed, we will be able to identify SAs as well as to quantify their accompanying modulating factors. The results of this work can then improve the expert systems where these SAs are implemented.

2. Materials and methods

2.1. Mutagenicity data sets

The two sets of data include various substances (220 for the Ames test mutagenicity -AMES-, and 48 for the mammalian cell gene mutation test -MCGM-) with α, β-unsaturated carbonyl moiety (Table 1). AMES data was derived from the Ames test classification made by Kazius et al. (2005) for mutagenicity, being their analysis restricted to the *Salmonella typhimurium* strains TA98, TA100, TA1535 and either TA1537 or TA97, performed with the standard plate or preincubation method either with or without a metabolic activation mixture. In that classification, *a compound was categorized as a mutagen if at least one Ames test result was positive and non-mutagen if exclusively negative Ames test results -one or more- were reported* (Kazius et al., 2005).

On the other hand, MCGM data was collected from compounds with published results from mammalian cells mutagenesis in L5178Y mouse lymphoma cells, CHO, AS52 and V79 lines of Chinese hamster cells extracted from the Chemical Carcinogenesis Research Information System (available at http://toxnet.nlm.nih.gov/cgi-bin/sis/htmlgen?CCRIS). The classification was performed in the same manner as the Ames test classification made by Kazius et al. (2005).

2.2. The TOPS-MODE approach

The TOPS-MODE approach is based on computing the spectral moments of the topological bond matrix (Estrada, 1995). The mathematical details of this approach have been documented in detail previously (Estrada, 1996, 1997), but an overview highlighting only the most important aspects will be given here.

Firstly, the molecular structure of each compound is represented by its molecular graph and then, the bond adjacency matrix (B) is derived. B 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, being two bonds 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

Table 1
CAS number, observed and predicted classification, and leverage values for the compounds used for obtaining the final QSAR models for the two endpoints (AMES: Eq. (7) and MCGM: Eq. (8)).

Compound no.	CAS	Observed		TOPS-MODE		TOXTREE	Observed		Predicted	
		AMES Class.	Partition	AMES Class.	Leverage	AMES Class.	MCGM Class.	Partition	MCGM Class.	Leverage
1	87406-72-2	1	Training	1	–	1	–	–	–	–
2	23282-20-4	−1	Training	−1	–	1	–	–	–	–
3	34807-41-5	−1	Training	−1	–	1	–	–	–	–
4	6379-69-7	−1	Training	−1	–	1	–	–	–	–
5	63166-73-4	−1	Training	−1	–	1	–	–	–	–
6	23246-96-0	1	Training	−1	–	−1	–	–	–	–
7	130-01-8	−1	Training	−1	–	−1	–	–	–	–
8	21794-01-4	−1	Training	−1	–	−1	–	–	–	–
9	61203-01-8	1	Training	1	–	1	–	–	–	–
10	2849-98-1	−1	Training	−1	–	−1	–	–	–	–
11	97-90-5	−1	Training	−1	–	−1	1	Training	1	–
12	4513-36-4	−1	Training	−1	–	−1	–	–	–	–
13	13675-34-8	−1	Training	−1	–	−1	–	–	–	–
14	868-77-9	−1	Training	−1	–	−1	–	–	–	–
15	5466-77-3	−1	Training	−1	–	−1	–	–	–	–
16	123-73-9	1	Training	1	–	1	–	–	–	–
17	96910-73-5	1	Training	1	–	1	–	–	–	–
18	14925-39-4	1	Training	1	–	1	–	–	–	–
19	2397-76-4	−1	Training	−1	–	−1	–	–	–	–
20	68162-37-8	1	Training	1	–	1	–	–	–	–
21	836-37-3	1	Training	1	–	1	–	–	–	–
22	29590-42-9	−1	Training	−1	–	−1	−1	Training	1	–
23	555-68-0	1	Training	1	–	1	–	–	–	–
24	58-54-8	−1	Training	1	–	1	–	–	–	–
25	3688-53-7	1	Training	1	–	1	1	Training	1	–
26	18829-55-5	1	Training	−1	–	1	–	–	–	–
27	1013-96-3	−1	Training	1	–	1	–	–	–	–
28	102059-18-7	1	Training	−1	–	1	–	–	–	–
29	7364-09-2	1	Training	1	–	1	–	–	–	–
30	499-12-7	−1	Training	−1	–	−1	–	–	–	–
31	10443-65-9	−1	Training	−1	–	−1	–	–	–	–
32	6281-23-8	1	Training	1	–	1	–	–	–	–
33	109460-96-0	−1	Training	−1	–	−1	–	–	–	–
34	5443-49-2	1	Training	1	–	1	–	–	–	–
35	645-62-5	−1	Training	−1	–	1	–	–	–	–
36	97-86-9	−1	Training	−1	–	−1	–	–	–	–
37	137-05-3	1	Training	1	–	−1	–	–	–	–
38	20426-12-4	−1	Training	−1	–	−1	–	–	–	–
39	104-55-2	−1	Training	−1	–	−1	–	–	–	–
40	488-11-9	1	Training	1	–	1	–	–	–	–
41	109-16-0	−1	Training	−1	–	−1	1	Training	1	–
42	434-07-1	−1	Training	−1	–	1	−1	Training	−1	–
43	126572-80-3	1	Training	1	–	1	–	–	–	–
44	13171-21-6	1	Training	1	–	1	–	–	–	–
45	141-32-2	−1	Training	−1	–	−1	1	Training	1	–
46	94-62-2	−1	Training	1	–	1	–	–	–	–
47	399-10-0	−1	Training	−1	–	−1	–	–	–	–
48	2998-23-4	−1	Training	−1	–	−1	–	–	–	–
49	2403-27-2	−1	Training	−1	–	−1	–	–	–	–
50	107-02-8	1	Training	1	–	1	1	Training	1	–
51	6606-59-3	−1	Training	−1	–	−1	–	–	–	–
52	97-63-2	−1	Training	−1	–	−1	–	–	–	–
53	1985-51-9	−1	Training	−1	–	−1	–	–	–	–
54	97055-37-3	1	Training	1	–	1	–	–	–	–
55	89811-25-6	1	Training	1	–	1	–	–	–	–
56	623-15-4	−1	Training	−1	–	−1	–	–	–	–
57	117823-31-1	1	Training	1	–	1	–	–	–	–
58	1774-66-9	−1	Training	−1	–	−1	–	–	–	–
59	999-55-3	−1	Training	−1	–	−1	–	–	–	–
60	2657-25-2	−1	Training	−1	–	−1	–	–	–	–
61	1107-26-2	−1	Training	−1	–	−1	–	–	–	–
62	125974-06-3	1	Training	1	–	1	–	–	–	–
63	6197-30-4	−1	Training	−1	–	−1	−1	Training	−1	–
64	2358-84-1	−1	Training	−1	–	−1	–	–	–	–
65	2403-28-3	−1	Training	−1	–	−1	–	–	–	–
66	683-51-2	1	Training	1	–	1	–	–	–	–
67	90147-21-0	1	Training	1	–	1	–	–	–	–
68	1466-88-2	1	Training	1	–	1	–	–	–	–
69	505-70-4	−1	Training	−1	–	−1	–	–	–	–
70	5234-68-4	−1	Training	1	–	1	–	–	–	–
71	97055-38-4	1	Training	1	–	1	–	–	–	–
72	2873-97-4	−1	Training	−1	–	1	–	–	–	–

Table 1 (*Continued*)

Compound no.	CAS	Observed		TOPS-MODE		TOXTREE	Observed		Predicted	
		AMES Class.	Partition	AMES Class.	Leverage	AMES Class.	MCGM Class.	Partition	MCGM Class.	Leverage
73	68053-32-7	1	Training	1	–	1	−1	Training	−1	–
74	1070-13-9	1	Training	1	–	1	–	–	–	–
75	19660-16-3	1	Training	1	–	1	–	–	–	–
76	104-28-9	−1	Training	−1	–	−1	–	–	–	–
77	3524-68-3	−1	Training	−1	–	−1	1	Training	1	–
78	142-09-6	−1	Training	−1	–	−1	–	–	–	–
79	14129-84-1	1	Training	1	–	1	–	–	–	–
80	122-57-6	1	Training	−1	–	−1	–	–	–	–
81	24140-30-5	1	Training	−1	–	−1	–	–	–	–
82	125974-08-5	1	Training	1	–	1	1	Training	1	–
83	15625-89-5	1	Training	−1	–	−1	1	Test	1	0.139
84	126572-78-9	1	Training	1	–	1	–	–	–	–
85	2223-82-7	−1	Training	−1	–	−1	–	–	–	–
86	710-25-8	1	Training	1	–	1	–	–	–	–
87	2213-00-5	1	Training	1	–	1	–	–	–	–
88	614-47-1	−1	Training	−1	–	−1	–	–	–	–
89	13088-34-1	1	Training	1	–	−1	–	–	–	–
90	4823-47-6	1	Training	1	–	1	–	–	–	–
91	17831-71-9	−1	Training	−1	–	−1	1	Training	1	–
92	1629-58-9	1	Training	−1	–	1	–	–	–	–
93	2206-89-5	1	Training	1	–	1	–	–	–	–
94	127072-60-0	1	Training	−1	–	−1	–	–	–	–
95	90147-18-5	1	Training	1	–	1	–	–	–	–
96	110-17-8	−1	Training	−1	–	−1	1	Training	1	–
97	6728-26-3	1	Training	1	–	1	–	–	–	–
98	62674-12-8	1	Training	1	–	−1	–	–	–	–
99	1576-87-0	1	Training	1	–	1	–	–	–	–
100	924-42-5	−1	Training	−1	–	1	–	–	–	–
101	5392-40-5	−1	Training	−1	–	−1	–	–	–	–
102	3695-86-1	−1	Training	−1	–	−1	–	–	–	–
103	619-89-6	1	Training	1	–	1	–	–	–	–
104	79-10-7	−1	Training	−1	–	−1	1	Training	1	–
105	2274-11-5	−1	Training	−1	–	−1	1	Training	1	–
106	78-85-3	1	Training	1	–	1	1	Training	1	–
107	959-23-9	−1	Training	−1	–	−1	–	–	–	–
108	91134-58-6	1	Training	1	–	1	–	–	–	–
109	122-40-7	−1	Training	−1	–	−1	–	–	–	–
110	97461-40-0	1	Training	1	–	1	–	–	–	–
111	90147-19-6	1	Training	1	–	1	–	–	–	–
112	96-33-3	−1	Training	1	–	−1	1	Training	1	–
113	79-41-4	−1	Training	−1	–	−1	–	–	–	–
114	112309-61-2	1	Training	1	–	−1	–	–	–	–
115	2393-18-2	−1	Training	−1	–	1	–	–	–	–
116	1152-48-3	1	Training	1	–	1	–	–	–	–
117	97461-41-1	1	Training	1	–	1	–	–	–	–
118	1222-98-6	1	Training	1	–	1	−1	Training	−1	–
119	7085-85-0	−1	Training	−1	–	−1	–	–	–	–
120	2082-81-7	−1	Training	−1	–	−1	–	–	–	–
121	2499-95-8	−1	Training	−1	–	−1	–	–	–	–
122	104-98-3	−1	Training	−1	–	−1	–	–	–	–
123	6755-13-1	1	Training	1	–	1	–	–	–	–
124	2157-01-9	−1	Training	−1	–	−1	–	–	–	–
125	25870-67-1	1	Training	1	–	1	–	–	–	–
126	557-48-2	−1	Training	−1	–	−1	–	–	–	–
127	3160-37-0	−1	Training	1	–	−1	–	–	–	–
128	97461-42-2	1	Training	1	–	1	–	–	–	–
129	140-10-3	−1	Training	−1	–	−1	1	Training	1	–
130	91642-47-6	1	Training	1	–	1	–	–	–	–
131	1565-94-2	−1	Training	−1	–	−1	–	–	–	–
132	6923-22-4	1	Training	1	–	1	–	–	–	–
133	147151-67-5	1	Training	1	–	1	–	–	–	–
134	28564-83-2	1	Training	1	–	1	–	–	–	–
135	142438-64-0	1	Training	1	–	−1	–	–	–	–
136	63-75-2	1	Training	1	–	−1	–	–	–	–
137	55557-02-3	1	Training	1	–	1	–	–	–	–
138	2210-28-8	−1	Training	−1	–	−1	–	–	–	–
139	97461-43-3	1	Training	1	–	1	–	–	–	–
140	125973-99-1	1	Training	1	–	1	–	–	–	–
141	3179-47-3	−1	Training	−1	–	−1	–	–	–	–
142	78-94-4	1	Training	1	–	1	–	–	–	–
143	101-39-3	−1	Training	−1	–	−1	–	–	–	–
144	90-65-3	−1	Training	−1	–	1	–	–	–	–
145	13048-33-4	−1	Training	−1	–	−1	–	–	–	–
146	36840-85-4	−1	Training	−1	–	−1	–	–	–	–

Table 1 (*Continued*)

Compound no.	CAS	Observed		TOPS-MODE		TOXTREE	Observed		Predicted	
		AMES Class.	Partition	AMES Class.	Leverage	AMES Class.	MCGM Class.	Partition	MCGM Class.	Leverage
147	110-26-9	1	Training	−1	–	1	–	–	–	–
148	14308-65-7	1	Training	1	–	1	–	–	–	–
149	14901-07-6	−1	Training	−1	–	1	–	–	–	–
150	97461-38-6	1	Training	1	–	1	–	–	–	–
151	53175-28-3	1	Training	1	–	1	–	–	–	–
152	3787-28-8	1	Training	1	–	1	–	–	–	–
153	103-11-7	−1	Training	−1	–	−1	1	Training	1	–
154	18031-40-8	−1	Training	−1	–	1	–	–	–	–
155	78-59-1	−1	Training	−1	–	1	−1	Test	−1	0.046
156	959-33-1	−1	Training	−1	–	−1	–	–	–	–
157	1193-54-0	1	Training	1	–	1	–	–	–	–
158	458-37-7	−1	Training	−1	–	−1	–	–	–	–
159	1615-02-7	−1	Training	−1	–	−1	–	–	–	–
160	614-48-2	−1	Training	1	–	1	–	–	–	–
161	555-66-8	1	Training	−1	–	1	–	–	–	–
162	6755-16-4	1	Training	1	–	1	–	–	–	–
163	122-69-0	1	Training	Outlier	–	−1	–	–	–	–
164	1663-39-4	−1	Training	−1	–	−1	–	–	–	–
165	15743-13-2	−1	Training	−1	–	1	–	–	–	–
166	97461-39-7	1	Training	1	–	1	–	–	–	–
167	39965-42-9	1	Training	1	–	1	–	–	–	–
168	31876-38-7	−1	Training	−1	–	−1	–	–	–	–
169	19337-19-0	−1	Training	−1	–	1	–	–	–	–
170	122551-89-7	1	Training	1	–	1	–	–	–	–
171	541-59-3	1	Training	−1	–	1	1	Test	1	0.142
172	2177-18-6	−1	Training	−1	–	1	–	–	–	–
173	137-66-6	−1	Training	−1	–	−1	–	–	–	–
174	89-65-6	−1	Training	−1	–	−1	–	–	–	–
175	80-71-7	−1	Training	1	–	1	–	–	–	–
176	97-88-1	−1	Training	−1	–	−1	–	–	–	–
177	23255-69-8	−1	Test	−1	0.183 [a]	1	–	–	–	–
178	480-81-9	1	Test	−1	0.068	−1	–	–	–	–
179	2439-35-2	−1	Test	−1	0.012	−1	–	–	–	–
180	17341-40-1	−1	Test	1	0.050	1	–	–	–	–
181	1070-70-8	−1	Test	−1	0.018	−1	–	–	–	–
182	37962-27-9	1	Test	1	0.103	1	–	–	–	–
183	2154-67-8	−1	Test	−1	0.065	−1	–	–	–	–
184	514-78-3	−1	Test	−1	0.127 [a]	−1	–	–	–	–
185	584-79-2	1	Test	−1	0.048	1	–	–	–	–
186	1135-24-6	−1	Test	−1	0.018	−1	–	–	–	–
187	1608-51-1	−1	Test	−1	0.016	−1	–	–	–	–
188	497-23-4	−1	Test	1	0.016	−1	–	–	–	–
189	7473-93-0	−1	Test	1	0.027	1	–	–	–	–
190	1609-93-4	1	Test	1	0.015	−1	–	–	–	–
191	87-56-9	1	Test	1	0.032	1	1	Training	1	–
192	1734-79-8	1	Test	1	0.025	1	1	Training	−1	–
193	331-39-5	−1	Test	−1	0.021	−1	1	Training	1	–
194	6203-18-5	−1	Test	−1	0.025	1	–	–	–	–
195	129401-88-3	−1	Test	−1	0.013	1	–	–	–	–
196	3066-70-4	1	Test	1	0.021	1	–	–	–	–
197	585-07-9	−1	Test	−1	0.044	−1	–	–	–	–
198	1874-12-0	1	Test	1	0.023	1	–	–	–	–
199	1030-27-9	−1	Test	−1	0.023	1	–	–	–	–
200	818-61-1	−1	Test	−1	0.014	−1	1	Training	1	–
201	766-40-5	1	Test	1	0.020	−1	–	–	–	–
202	91182-09-1	1	Test	1	0.017	1	–	–	–	–
203	105-76-0	−1	Test	−1	0.023	−1	–	–	–	–
204	623-30-3	−1	Test	−1	0.017	−1	–	–	–	–
205	142-83-6	1	Test	1	0.021	1	1	Training	−1	–
206	106-63-8	−1	Test	−1	0.011	−1	–	–	–	–
207	90147-31-2	1	Test	1	0.018	1	–	–	–	–
208	1951-56-0	1	Test	1	0.020	1	1	Test	1	0.177
209	2648-51-3	1	Test	1	0.035	1	–	–	–	–
210	3290-92-4	1	Test	−1	0.068	−1	1	Training	1	–
211	1874-22-2	1	Test	1	0.031	1	–	–	–	–
212	108-31-6	−1	Test	−1	0.032	−1	–	–	–	–
213	77439-76-0	1	Test	1	0.029	1	1	Training	1	–
214	4655-34-9	−1	Test	−1	0.011	−1	–	–	–	–
215	4074-88-8	−1	Test	−1	0.025	−1	–	–	–	–
216	1874-24-4	1	Test	1	0.022	1	–	–	–	–
217	96910-71-3	1	Test	1	0.181 [a]	1	−1	Training	−1	–
218	125974-01-8	1	Test	1	0.025	1	–	–	–	–
219	922-63-4	1	Test	1	0.019	1	–	–	–	–
220	327-97-9	−1	Test	−1	0.092	−1	−1	Training	1	–

Table 1 (*Continued*)

Compound no.	CAS	Observed		TOPS-MODE		TOXTREE	Observed		Predicted	
		AMES Class.	Partition	AMES Class.	Leverage	AMES Class.	MCGM Class.	Partition	MCGM Class.	Leverage
221	4170-30-3	–	–	–	–	–	−1	Training	1	–
222	303-34-4	–	–	–	–	–	1	Training	1	–
223	14371-10-9	–	–	–	–	–	1	Training	1	–
224	110-44-1	–	–	–	–	–	−1	Training	−1	–
225	108893-54-5	–	–	–	–	–	−1	Training	−1	–
226	18409-46-6	–	–	–	–	–	1	Training	1	–
227	33118-34-2	–	–	–	–	–	−1	Training	−1	–
228	3588-17-8	–	–	–	–	–	−1	Training	1	–
229	5956-39-8	–	–	–	–	–	−1	Training	−1	–
230	62966-21-6	–	–	–	–	–	1	Training	1	–
231	120-57-0	–	–	–	–	–	1	Test	1	0.178
231	140-88-5	–	–	–	–	–	1	Test	1	0.055
233	79-06-1	–	–	–	–	–	1	Test	1	0.079
234	37841-91-1	–	–	–	–	–	1	Test	−1	0.142
235	50656-61-6	–	–	–	–	–	−1	Test	−1	0.089

[a] Chemicals with leverage values above the threshold (0.120 for AMES mutagenicity and 0.307 for MCGM) and, for that reason, its predictions were not taken into account.

being modelled. In this work, the weights included the standard bond distance (Std), standard bond dipole moments (Dip, Dip2), as well as contributions from the following atomic properties: hydrophobicity (Hyd), polar surface area (Pols), polarizability (Pol), molar refractivity (Mol), van der Waals radii (vdW), Gasteiger–Marsilli charges (Gas), atomic masses (Ato), solute excess molar refraction (Ab-R_2), solute dipolarity/polarizability (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}). As described previously (Estrada et al., 2003b), the atomic contributions are transformed into bond weight contributions – $w(i,j)$ – as follows:

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

where w_i and δ_i stand for the atomic weight and vertex degree of the atoms i and j, respectively. Finally, the spectral moments are defined as the traces (i.e., the sum of the main diagonal elements) of the different powers of the weighted B matrix.

In this work, these graph-based descriptors were computed with the MODESLAB software (http://www.modeslab.com) (Gutierrez and Estrada, 2002), using the SMILES (Simplified Molecular Input Line Entry System) notation available for each compound (Weininger, 1988).

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), excluding the hydrogen atoms. We have also multiplied μ_0 and μ_1 for the first 15 spectral moments, obtaining therefore 30 new variables. Be aware that, these variables might offset the linear approximation assumption of the model. As to the modelling technique, we opted for building discriminant functions able of classifying the chemicals as actives or inactives This was attained by the Linear Discriminant Analysis (LDA) technique implemented in STATISTICA software 8.0 (Frank, 2002).

To summarize, the following three-stage procedure was adopted to develop the structure–activity relationships.

Stage 1: Model selection. This proceeds as follows.

1. Select a small subset from the total chemicals to act as "test" set (for AMES: 44 from the total 220 chemicals; for MCGM: 9 from the total 48 chemicals; see Pérez-Garrido et al., in press). The remaining chemicals form the "training" set for QSAR modelling.
2. Draw the molecular graphs for each molecule included in the training and test sets.
3. Compute the spectral moment's descriptors using an appropriate set of weights.
4. Find an adequate QSAR model from the training set by a discriminant-based approach. The task here is to obtain a mathematical function (see Eq. (2) below) that best describes the studied activity P (in our case, the mutagenicity) as a linear combination of the X-predictor variables (the k-spectral moments μ_k), with the coefficients a_k. Such coefficients are to be optimized by means of LDA.
5. The QSAR model is subjected 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 mutagenicity of the studied molecules.

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

P values of +1 and −1 were assigned to active and inactive compounds, respectively. Moreover, several models were first obtained by forward stepwise LDA and then, the best of them was improved by the replacement method (RM) (Duchowicz et al., 2006).

Stage 2: Model validation. Assess the performance of the derived QSAR model by rigorous internal and external validation, looking in particular to its applicability and predictive power.

Stage 3: SAs identification. Identify one or more critical SAs for mutagenesis. That is to say, compute the contribution of different selected substructures and determine their quantitative contribution to the mutagenicity of the studied chemicals.

2.3. Model validation

Two kinds of diagnostic statistical tools were used for evaluating the performance of our discriminant 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 dataset of chemicals into training and an external validation test set. Full details of this partition can be found in our previous work (Pérez-Garrido et al., in press).

Measures of goodness of fit have been estimated by computing standard statistics such as the Mahalanobis distance (D^2), the Wilks' lambda (λ), the Fisher's statistic (F), and the corresponding p-level (p), as well as the percentage of correct classifications (accuracy). One should mention in particular that, the Mahalanobis distance shows whether the model has an adequate discriminatory power for differentiating between the two respective groups – active and inactive chemicals – whereas Wilks' lambda takes values in the range of zero (perfect discrimination) to one (no discrimination at all). Also, it should be remarked here that we minimized precisely the statistic when using the RM, not the standard deviation as it is done in linear regression analysis. Furthermore, similarly to the FIT statistic used in regression analysis (Kubinyi, 1994a,b), which allows comparing models with different number of variables (p) and cases (n), we have employed a new statistical parameter, FIT(λ), defined by:

$$FIT(\lambda) = \frac{(1-\lambda)(n-p-1)}{(n+p^2)\lambda} \quad (3)$$

Goodness-of-prediction for both the training and test sets was assessed by the following statistical measures:

- Accuracy: the percentage of chemicals correctly classified.
- Sensitivity: the percentage of toxicologically active chemicals (positives) correctly predicted as positives (calculated out of the total number of positives).
- Specificity: the percentage of toxicologically inactive chemicals (negatives) correctly predicted as negatives (calculated out of the total number of negatives).
- Kappa (K) (Cohen, 1960): The kappa index excludes matching due solely to chance. The maximum possible agreement is $K = 1$. $K = 0$ is obtained when the agreement observed is that expected exclusively by chance. If the agreement is higher than

Table 2
Interpretation of kappa

Kappa	Agreement
< 0	Less than chance agreement
0.01–0.20	Slight agreement
0.21–0.40	Fair agreement
0.41–0.60	Moderate agreement
0.61–0.80	Substancial agreement
0.81–0.99	Almost perfect agreement

Table 3
Results of the classification (%) of compounds in the training and external test sets, according to the TOPS-MODE models obtained here.

	AMES test (Eq. (4))	AMES test Cross val.	MCGM (Eq. (5))	MCGM Cross val.	AMES test (Eq. (6))	AMES test Cross val.
Training						
Sensitivity	85.54	85.47	92.59	93.35	86.59	86.34
Specificity	90.32	90.57	75.00	75.17	91.40	90.95
Accuracy	88.07	88.16	87.18	87.67	89.14	88.79
Test						
Sensitivity	85.00	85.71	85.71	89.87	85.00	86.13
Specificity	80.95	88.00	100.00	69.67	85.71	88.46
Accuracy	82.93	86.92	88.89	84.10	85.37	87.36

expected simply because of chance, $K > 0$, while if it is less, $K < 0$. However, a commonly cited scale is represented in Table 2 (Landis and Koch, 1977).

In addition, we carried out a cross-validation procedure on the training set. Specifically, the leave-group-out (LGO) procedure was applied, leaving out 20% of the training set by random extraction and then recalculating the model and the statistics with the remaining chemicals. This LGO procedure was repeated 300 times. The mean values of the accuracy, sensitivity, and specificity for both training and test sets, as well as the mean values of Wilk's λ (λ_{Cross}) and squared Mahalanobis distances (D^2_{Cross}), are reported.

In summary, good overall quality of the models is indicated by small values of λ, λ_{Cross} along with high values of FIT(λ), D^2, F and Kappa.

The spectral moments are inherently collinear. From the point of view of QSAR modelling, the main drawback of collinearity is that it increases the standard errors associated with the individual coefficients, thereby decreasing their value for purposes of interpretability. To overcome this problem, we have employed here the Randić's method of orthogonalisation (Lucic et al., 1995; Klein et al., 1997; Randić, 1991b,c,a). 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 Ωv_1 and the second one is orthogonalised with respect to it by taking the residual of its correlation with Ωv_1. The process is repeated until all variables are completely orthogonalised. For extracting of the information contained in the orthogonalised descriptors we followed the procedure reported by Estrada and Molina (2006).

2.4. Applicability domain of the models

The utility of a QSAR model is its ability to accurately predict activity for new substances, which requires a careful assessment of the true predictive ability of models. 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 (Eriksson et al., 2003; Netzeva et al., 2005) but the most common one encompasses determining the leverage values for each compound (Gramatica, 2007). A Williams plot, i.e. the plot of standardized 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 $3p/n$, where n is the number of training compounds and p the number of model parameters, whereas $x = 2$ or 3), lying outside this area (vertical lines) the outliers and (horizontal lines) influential chemicals. For future predictions, only predicted mutagenicity for chemicals belonging to the chemical domain of the training set should be proposed and used (Vighi et al., 2001). So, calculations of validation set clasifications were performed only for those substances that had a leverage value below the threshold h^*.

3. Results and discussion

3.1. QSAR models

Following the computational strategies outlined in the previous section, the best model obtained for each of the chosen endpoints we can see in the following equations and in Table 3. As seen, these model is good both statistical significance and goodness of fit and prediction.

$$\begin{aligned}\text{AMES} = {} & 1.758 + 1.691\mu_1^{\text{Dip2}} - 3.399\times10^{-3}\mu_6^{\text{Dip2}} \\ & - 1.564\times10^{-2}\mu_5^{\text{Hyd}} + 8.557\mu_1^{\text{Gas}} + 3.341\times10^{-2}\mu_5^{Ab-\pi_2^H} \\ & - 6.858\times10^{-12}\mu_0\mu_{15}^{\text{Pol}} + 1.338\times10^{-2}\mu_1\mu_3^{\text{Hyd}} \qquad (4)\end{aligned}$$

$N = 176$(83 positives, 93 negatives); $\lambda = 0.482$;

$D^2 = 4.259$; $\lambda_{Cross} = 0.478$; $D^2_{Cross} = 4.349$; $p < 10^{-5}$;

$F = 25.770$; FIT = 0.766; $K = 0.659$

$$\begin{aligned}\text{MCGM} = {} & 4.143 - 2.548\mu_2^{\text{Dip}} + 3.012\mu_2^{\text{Dip2}} \\ & - 1.54\times10^{-4}\mu_7^{\text{Pol}} + 5.271\mu_1^{\text{Gas}} \qquad (5)\end{aligned}$$

$N = 39$(26 positives, 13 negatives); $\lambda = 0.412$; $D^2 = 6.343$;

$\lambda_{Cross} = 0.399$; $D^2_{Cross} = 6.756$; $p < 10^{-5}$; $F = 12.107$;

$FIT = 0.881$; $K = 0.727$

Then, we search for the presence of outliers that might be distorting these models. The high value found for the standard residual (> 3) of chemical 163 (i.e. of cinnamyl cinnamate) suggests that it could be an outlier. In general the cinnamyl derivatives were not mutagenic in the Ames test and their metabolism *in vivo* is usually to hippuric acid (Belsito et al., 2007). Therefore the derived AMES model (Eq. (4)) is not able to predict the mutagenicity of this chemical since it does not follow the general pattern of cinnamyl derivatives, thus being an outlier. If we remove this chemical from the training set and further proceed to refitting, we obtained the AMES model shown below.

$$\begin{aligned}\text{AMES} = {} & 1.864 + 1.882\mu_1^{\text{Dip2}} - 3.757\times10^{-3}\mu_6^{\text{Dip2}} \\ & - 1.734\times10^{-2}\mu_5^{\text{Hyd}} + 9.489\mu_1^{\text{Gas}} + 3.705\times10^{-2}\mu_5^{Ab-\pi_2^H} \\ & - 7.534\times10^{-12}\mu_0\mu_{15}^{\text{Pol}} + +1.426\times10^{-2}\mu_1\mu_3^{\text{Hyd}} \qquad (6)\end{aligned}$$

$N = 175$ (82 positives, 93 negatives); $\lambda = 0.451$;

$D^2 = 4.818$; $\lambda_{Cross} = 0.448$; $D^2_{Cross} = 4.913$; $p < 10^{-5}$;

$F = 28.958$; FIT = 0.869; $K = 0.707$

Eliminating the outlier produces an appreciable improvement in the statistical parameters as well as the percentages of classification (see Eq. (6) and Table 3). Another aspect deserving special attention is the degree of collinearity of the variables of the model, which can readily be diagnosed by analyzing the cross-correlation

Table 4
Intercorrelation among the descriptors selected for the initial AMES model (Eq. (4)).

	μ_1^{Dip2}	μ_6^{Dip2}	μ_5^{Hyd}	μ_1^{Gas}	$\mu_5^{Ab-\pi_2^H}$	$\mu_0\mu_{15}^{\text{Pol}}$	$\mu_1\mu_3^{\text{Hyd}}$
μ_1^{Dip2}	1.00	–	–	–	–	–	–
μ_6^{Dip2}	0.82	1.00	–	–	–	–	–
μ_5^{Hyd}	0.49	0.86	1.00	–	–	–	–
μ_1^{Gas}	-0.57	-0.63	-0.53	1.00	–	–	–
$\mu_5^{Ab-\pi_2^H}$	0.56	0.91	0.97	-0.60	1.00	–	–
$\mu_0\mu_{15}^{\text{Pol}}$	0.50	0.84	0.87	-0.56	0.91	1.00	–
$\mu_1\mu_3^{\text{Hyd}}$	0.18	0.34	0.57	-0.36	0.42	0.44	1.00

Table 5
Intercorrelation among the descriptors selected for the initial MCGM model (Eq. (5)).

	μ_2^{Dip}	μ_2^{Dip2}	μ_7^{Pol}	μ_1^{Gas}
μ_2^{Dip}	1.00	–	–	–
μ_2^{Dip2}	0.99	1.00	–	–
μ_7^{Pol}	0.89	0.89	1.00	–
μ_1^{Gas}	−0.71	−0.72	−0.49	1.00

matrix. Tables 4 and 5 show that, for both models, there are several descriptor variables highly correlated with each other. Rather than deleting any of these descriptors, it is of interest to examine the performance of orthogonal complements.

Following Randić's technique, we have determined orthogonal complements for all variables of the above non-orthogonalised models (Eqs. (5) and (6)), the following QSAR models were obtained:

$$\begin{aligned}\text{AMES} = {} & 2.188 + 0.230\Omega^3\mu_1^{Dip2} - 3.827\times10^{-3}\Omega^4\mu_6^{Dip2} \\ & + 1.133\times10^{-2}\Omega^6\mu_1\mu_3^{Hyd} + 4.050\Omega\mu_1^{Gas} \\ & - 5.036\times10^{-3}\Omega^5\mu_5^{Hyd} + 3.007\times10^{-3}\Omega^2\mu_5^{Ab-\pi_2^H} \\ & - 7.534\times10^{-12}\Omega^7\mu_0\mu_{15}^{Pol} \end{aligned} \quad (7)$$

$N = 175$ (82 positives, 93 negatives); $\lambda = 0.451$; $D^2 = 4.818$;

$\lambda_{Cross} = 0.448$; $D^2_{Cross} = 4.913$; $p < 10^{-5}$;

$F = 28.958$; $FIT = 0.869$; $K = 0.707$

$$\begin{aligned}\text{MCGM} = {} & 4.183 - 2.193\Omega^2\mu_2^{Dip} - 0.062\Omega\mu_2^{Dip2} \\ & - 1.088\times10^{-4}\Omega^3\mu_7^{Pol} + 5.271\Omega^4\mu_1^{Gas} \end{aligned} \quad (8)$$

$N = 39$ (26 positives, 13 negatives); $\lambda = 0.412$;

$D^2 = 6.343$; $\lambda_{Cross} = 0.406$; $D^2_{Cross} = 6.548$; $p < 10^{-5}$;

$F = 12.107$; $FIT = 0.881$; $K = 0.727$

The percentages of classifications of the derived orthogonalised models were found to be the same as the non-orthogonalised models (results not shown).

3.2. Comparison between the TOPS-MODE QSAR model and the TOXTREE software for predictions of AMES mutagenicity

The TOXTREE software is capable of making structure-based predictions for a number of toxicological endpoints but one of its modules aims at predicting carcinogenicity and mutagenicity. Mutagenic predictions from this expert software system are based in a revised list of structural alerts, taken from several literature sources, and in one QSAR model for α, β-unsaturated aldehydes (Benigni and Bossa, 2008; Benigni et al., 2008). The present analysis concerned comparing the predictions of our TOPS-MODE model against those of TOXTREE with respect to the AMES mutagenicity for this family of compounds. Thus, a chemical is considered to be predicted as positive (i.e., potentially mutagenic) either if it contains one genotoxic (DNA reactive) structural alert or if it belongs to the applicability domain of the relating QSAR model, otherwise it was identified as negative. For an easy visual comparison, the results are expressed as a Receive Operating Characteristics (ROC) graph (Fig. 2). A ROC graph reports true positive rate (sensitivity) on the *Y*-axis, and false positive rate (1-specificity) on the *X*-axis. In a ROC graph, perfect performance is located at the left upper corner, and random results lying on the diagonal line (Provost and Fawcett, 2001). Further details of the obtained results are collected in Table 6 (see also Table 1).

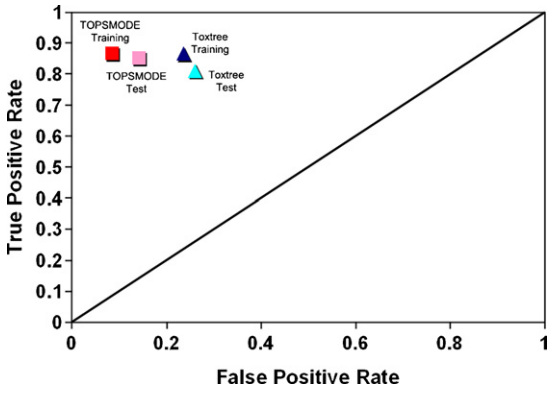


Fig. 2. Receiver operating characteristic graph for mutagenicity predictions of α, β-unsaturated carbonyl compounds given by TOXTREE and TOPS-MODE based on Ames data. The diagonal line corresponds to random responses whereas the top left corner to ideal performance.

This analysis shows that the present QSAR model (Eq. (7)) has higher specificity and accuracy than the TOXTREE software for the training set, though identical sensitivity. However, for the external test set, the percentage of chemicals correctly identified by the TOPS-MODE model is superior to the one attained by TOXTREE. These results imply that the performance of the TOPS-MODE model is better than that of the TOXTREE software, not only in the percentage of overall classifications, but also and more important, in terms of the SA modulating factors for the mutagenicity of this chemical class. This is clearly due to the ability of TOPS-MODE approach in modelling the mutagenic activity at a local scale, which further allows quantifying how each alert is modulated by several molecular environments (modulating factors). More details about this issue will be given in next subsection.

3.3. Identification of structural alerts

One advantage of the present approach for QSAR studies is that it can provide information explaining how structural features of molecules can account for the endpoint activities (Estrada, 2008). It is then possible to detect fragments that contribute positively or negatively to a particular target endpoint and their effects been interpreted in terms of physicochemical properties.

Specifically in our case, the contributions to the mutagenicity for each of the selected fragments (see Fig. 3) were extracted from the final orthogonal-descriptor models related to the two endpoints. Table 7 shows the particular numerical values of the contributions of such fragments. A careful look at these values allows us to find functional groups that either hamper the toxicity or enhance it. Further, it might lead us to design molecular structures that are less

Table 6
Results of the classification (%) of compounds in the training and external test sets, according to the TOPS-MODE model (Eq. (7)) and the TOXTREE software.

	TOPS-MODE	TOXTREE
Training		
Sensitivity	86.59	86.59
Specificity	91.40	76.34
Accuracy	89.14	81.14
Test		
Sensitivity	85.00	80.95
Specificity	85.71	73.91
Accuracy	85.37	77.27

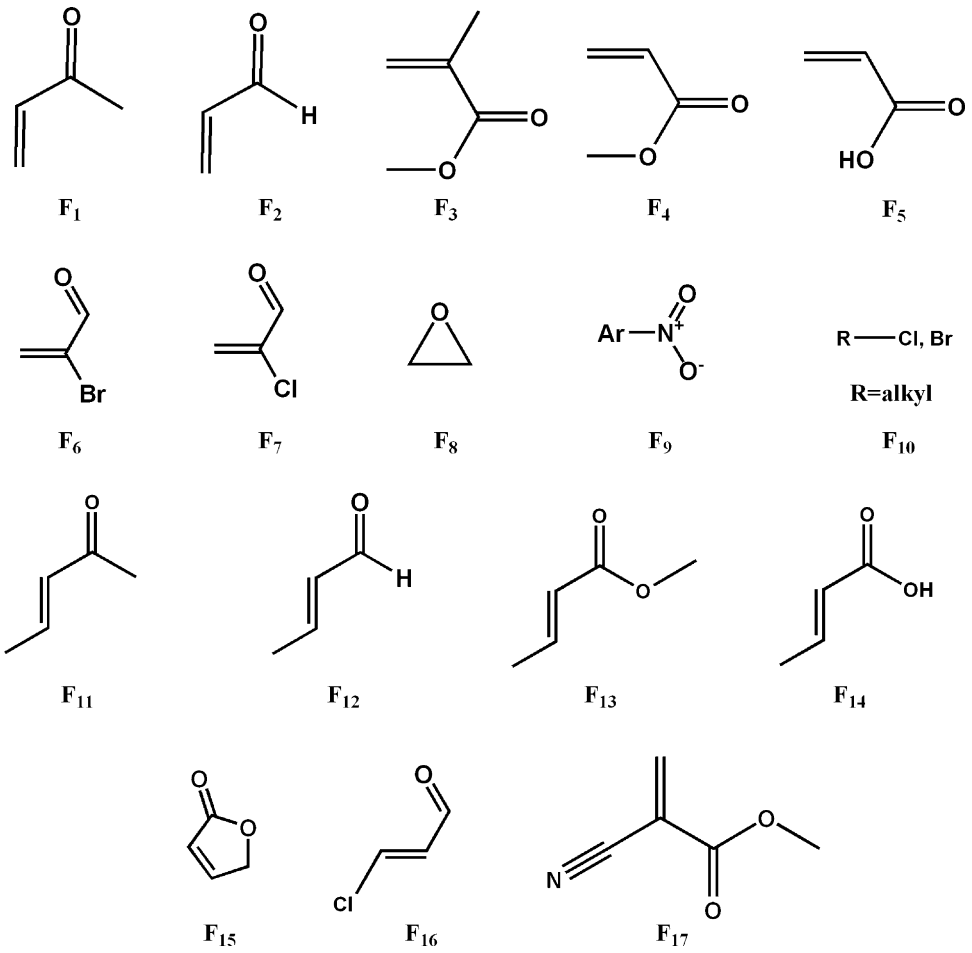


Fig. 3. Selected molecular fragments (substructures) for which their contributions to either the AMES or the MCGM mutagenicity were calculated according to the TOPS-MODE models obtained here (Eqs. (7) and eq:MLAorto).

toxic, to find new structural alerts or to a rapid screening among a long list of substances.

Regarding AMES mutagenicity, firstly, a comparison between fragments F_1 and F_2 shows that, the ketonic fragment (F_1) contributes less than the aldehyde fragment (F_2), the F_1 contribution being even negative. This in clear agreement with the analysis performed by Koleva et al. (2008), where the authors concluded that aldehydes are more reactive than ketones due to the size and electronic effects of their substituents. Besides, one can easily see in Fig. 1 that, the aldehyde group has a greater electron-withdrawing effect on the double bond than the ketonic group which increases its reactivity in the Michael addition mechanism.

Secondly, the presence of halogens in position α of the double bond adjacent to the carbonyl group (fragments F_6, F_7 and F_{16}) increase the mutagenicity of this family of compounds (Eder and Weinfurtner, 1994; Eder et al., 1990), due to the cross-linking potential with another DNA or protein nucleophilic centre (Van Beerendonk et al., 1992).

On the other hand, fragments F_8 to F_{10} relate to well recognized structural alerts for mutagenic AMES data, i.e.: epoxides (F_8), alkyl halides (F_{10}) and nitro aromatics (F_9). The first two functional groups are known alkylating agents while the latter one is mainly activated by means of nitroreduction and oxidative pathways involving several enzymes in different organisms (Purohit and Basu, 2000) to the N hydroxyl species, which are then transformed into reactive nitrogen esters or nitrenium ions and that in turn, may attack DNA forming adducts (Miller and Miller, 1983; Sasaki et al., 2002). In what concerns the positive contribution of the cyano acrylate group (fragment F_{17}), possibly it is due to an electron-withdrawing effect of the cyano moiety which increases the Michael addition reactivity of the double bond (Aptula and Roberts, 2006), as acrylates have a negative contribution (fragment F_4).

So, without doubt, the main mechanism of action for this family of compounds is the Michael type addition mechanism since our

Table 7
Contributions of the different structural fragments to the AMES and MCGM mutagenicity according to the TOPS-MODE models obtained here.

Fragment	Ames contribution	MCGM contribution
F_1	−0.112	0.196
F_2	1.673	2.075
F_3	−0.824	1.465
F_4	−0.775	2.981
F_5	−1.494	2.769
F_6	2.225	–
F_7	2.234	2.784
F_8	0.091	–
F_9	3.332	–
F_{10}	1.364	–
F_{11}	−0.143	−1.932
F_{12}	1.470	−0.060
F_{13}	−0.464	2.119
F_{14}	−1.449	0.648
F_{15}	1.710	3.866
F_{16}	2.126	–
F_{17}	0.371	–

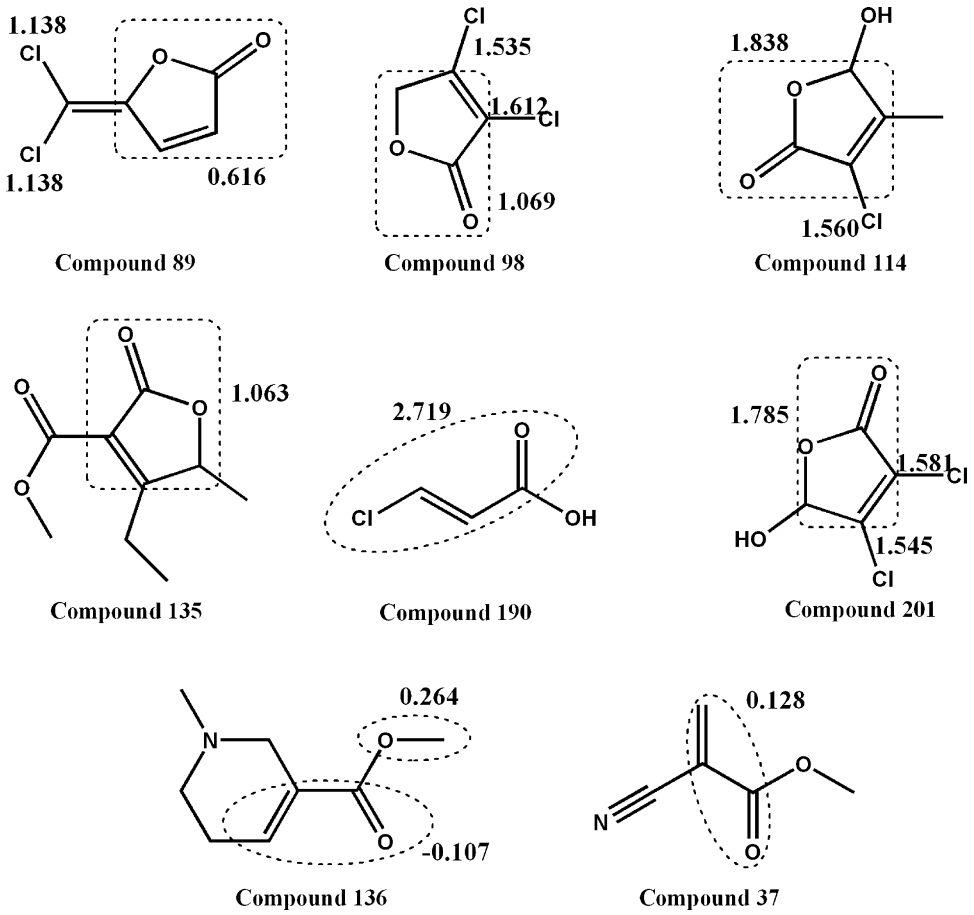


Fig. 4. Structure representation of some TOXTREE false negative compounds correctly predicted by the TOPS-MODE model together with their bond contributions (Eq. (7)).

analysis revealed that substituents in the α or β-carbon atoms have a strong influence in mutagenicity just as for Michael acceptors. Although it is known that the Michael acceptors are soft electrophiles, it does not mean that they are unreactive toward hard nucleophiles like DNA (Aptula and Roberts, 2006).

Another important feature that we can draw from both of our models (Eqs. (7) and (8)) is the mutagenicity of the furan-2(5H)-one ring (fragment F_{15}). This moiety together with halogenated α, β-unsaturated carbonyl compounds (fragments F_6, F_7 and F_{16}) are present in known mutagenic substances such as 3-chloro-4-(dichloromethyl) 5-hydroxy 2(5H) furanone (compound **213**) and 3-chloro-4(chloromethyl)-5-hydroxy-2(5H)-furanone (compound **82**) (McDonald and Komulainen, 2005), including in others not as well known, like compounds **89**, **98**, **114**, **135**, **190** and **201**. Here it should be emphasised that, the TOXTREE software does not recognize any structural alerts in the latter, classifying thus them as false negatives. Fig. 4 displays some TOXTREE false negatives for the Ames test mutagenicity, which were correctly predicted by our model together with the computed TOPS MODE fragment contributions.

A close inspection of Fig. 4, shows that fragments F_7, F_{15} and F_{16} have positive contributions to the mutagenic activity and so, this may be due to the presence of the furan 2(5H) one ring or to chlorine at the double bond adjacent to the carbonyl group, as it is known that the presence of halogens or halogenated alkyl groups at the furan double bond increases mutagenic potency (Lalonde et al., 1991; McDonald and Komulainen, 2005). In relation to compound **89**, the presence of allylic chlorines contributing positively also has a role. Notice however that our TOPS MODE model predicts a false positive, namely butenolide (compound **188**), which also contains fragment F_{15}. Thus, the modulating factors for this substructure have to be studied further.

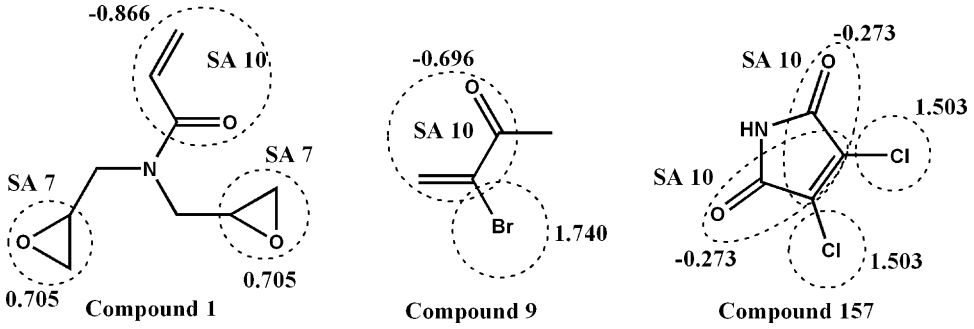


Fig. 5. Structure representation of some compounds of the AMES training set along with their TOPS-MODE bond contributions.

Another type of examples of TOXTREE false negatives, correctly predicted by our model, should be pointed out. For instance, our model recognises that the presence of a methyl group in compound **136** (see Fig. 4) yields a positive contribution (0.264) to its mutagenicity. For this compound, the mechanism of mutagenic action in bacteria is probably the same as the one in rats where it has been shown that the loss of the methyl group (Boyland and Nery, 1969) may bind with nucleic acid and protein (Nery, 1971). With regard to compound **37** (Fig. 4), its mutagenic mode of action is still not known (Richard, 2001), but the presence of fragment F_{17} (referred to above) can be responsible for the mutagenicity of its α, β-unsaturated carbonyl moiety.

Fig. 5 depicts three mutagenic substances for which the TOXTREE software identifies structural alerts, but also the TOPS-MODE is able of discriminating their differences in terms of fragment contributions.

For instance, the values of the fragment contributions in compound **1** show that the presence of α, β-unsaturated carbonyl moiety is not responsible for its mutagenicity, but instead the presence of the epoxide because similar substances without oxirane moiety such as N,N-diethylacrylamide and N,Ndimethylacrylamide are not mutagenic to Ames test (Hashimoto and Tanii, 1985).

As to compounds **9** and **157**, the structural alert detected by TOXTREE corresponds to the α, β-unsaturated carbonyl moiety while our model detected, as shown in Fig. 5, a negative contribution for this substructure, and relates their activity possibly due to the presence of allyls halogens corresponding to fragments F_6, F_7 and F_{16}, which could act as cross-linking agents (Van Beerendonk et al., 1992; Lynch and Crovetti, 1972; Smith, 1987).

Moreover, TOPS-MODE classifies correctly the majority of false positives obtained by the TOXTREE software, further detecting negative bond contributions for the structural alerts that allow, in a quantitative way, to properly interpret the possible cause of their non-mutagenicity (Table 8).

As for such compounds, there is no mechanism that explains their non-mutagenicity in *Salmonella typhimurium* strains with or without metabolic activation; we will try to set up hypothesis based on the contributions' results and on the bio-transformations produced by other organisms.

For example, in compound **2**, despite from having several hydroxyl groups which makes it more hydrophilic and therefore less mutagenic, the epoxide group displays no mutagenicity most likely due to a metabolic reduction of this group as seen in gastrointestinal microbes (Hedman and Pettersson, 1997).

The non-mutagenicity of compounds **3** and **5** are probably due to the large size of these molecules, but nevertheless TOPS-MODE identifies negative bond contributions for the TOXTREE structural alerts.

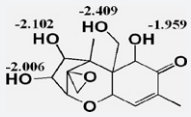
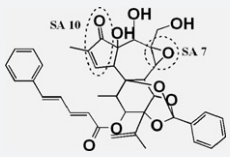
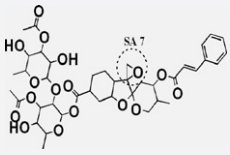
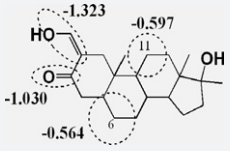
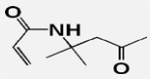
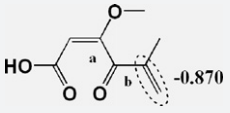
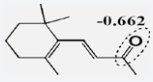
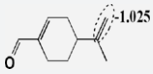
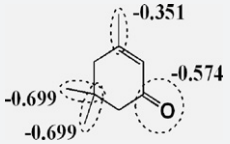
As to compound **42**, it is known that it is metabolized in humans mainly by several mechanisms, i.e.: reduction and subsequent oxidation of the hydroxymethylene group, a hydroxylation in position 6 and a carbonyl reduction (Fragkakia et al., 2009). These changes are consistent with the negative contributions obtained, and based on this a similar hydroxylation in position 11 could thus happen.

The presence of an acryl amide group, as seen previously for N,N-diglycidylacrylamide (compound **1**), is not responsible for the mutagenicity in the Ames test, which can be further observed by the values obtained for the SA of compound **72**.

For the following compound (**144**), a detoxification mechanism in mice through its conjugation with glutathione in the double bond has been presented (Chan et al., 1982, 1984), which has a large negative contribution (see Table 8), and so it could act similarly in *Salmonella typhimurium*.

The lack of mutagenicity for compound **149**, as seen in rabbits (Lalko et al., 2007), may be due to a hydroxylation of the carbonyl group (very negative bond contribution).

Table 8
Structure representation of some TOXTREE false positive compounds correctly predicted by the TOPS-MODE model together with their bond contributions (Eq. (7)).

Compound no.	Chemical representation	SA10 [a]	SA7 [a]
2		−0.969	−0.34
3		−1.381	−0.019
5		–	−0.032
42		0.054	–
72		−0.609	–
144		(a) 0.298 (b) −1.682	–
149		−0.764	–
154		1.593	–
155		−0.375	–

[a] SA10 and SA7 are codes of the structural alerts defined by Benigni and Bossa (2008) corresponded to α, β-unsaturated carbony and oxirane moieties, respectively.

Compound **154**, although it is an aldehyde, is not mutagenic in the Ames test. This compound has a detoxification metabolism in *Euglena gracilis Z* (Noma et al., 1991) due to the oxidation of aldehyde and reduction of the double bond. Perhaps the latter is the one that predominates in Salmonella typhimurium because of the negative contribution that such double bond has in this compound. Compound **154** also has an aliphatic cycle around the double bond which affords an electron-donating effect that decreases its Michael addition reactivity (Aptula and Roberts, 2006).

The mechanism of detoxification in rats and rabbits of compound **155** is by methyl carboxylation or the reduction of the carbonyl group (Dutertre-Catella et al., 1978), and as can be seen both bonds have negative contributions, but maybe in Salmonella typhimurium, it is more important the metabolic pathway by a possible hydroxylation of methyl in position 5 as it has been observed too in *Aspergillus niger* (Joe et al., 1989).

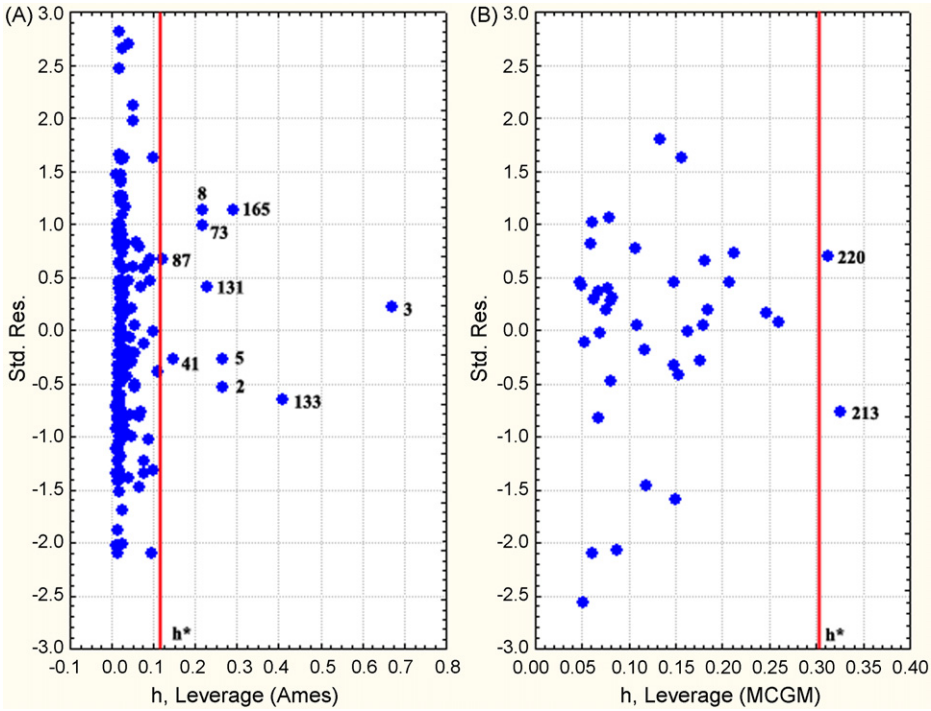


Fig. 6. Williams plot based on Eq. (7) (A) and (8) (B), i.e., plot of standardised residuals versus leverage values with a warning leverage of $h* = 0.120$ and $h* = 0.307$, respectively.

For the other endpoint studied, MCGM, there are no structural alerts identified in the literature. For this endpoint, the presence of the α, β-unsaturated carbonyl moiety produces mutagenicity either ketonic, aldehydic, acrylic or methacrylic (fragments F_1 to F_5 in Fig. 3 and Table 7). One can see also an increase in the contribution of fragment F_4 compared to fragment F_3. Electro-donating substituents such as methyl groups in position α reduce the reactivity of this moiety by Michael type mechanism (Aptula and Roberts, 2006). These findings, among other comparative analysis of the mutagenic potency of various acrylate and methyl acrylate derivatives, lead to the hypothesis that acrylate were more active mutagens than methylacrylates (Dearfield et al., 1989).

Moreover, when comparing fragments F_1, F_2, F_4 and F_5 to F_{11}, F_{12}, F_{13} and F_{14}, respectively, a similar conclusion for this endpoint to that obtained by our group in a previous work (Pérez-Garrido et al., in press) is attained. That is to say, that alkyl substituents in position β at the double bond adjacent to the carbonyl group decrease the mutagenic character of the substance by reducing the positive charge at the terminal carbon (Aptula and Roberts, 2006), and the latter is the preferred site of nucleophilic attack (Feron et al., 1991; Dearfield et al., 1991) by a Michael type addition mechanism to the sulfhydryl of glutathione (GSH) or by an enzymatic reaction catalyzed by GSH transferase (Ciaccio et al., 1998; Schultz et al., 2005). But also, GSH when depleted down to $< 20\%$ (Glaab et al., 2001) is a prerequisite for α, β-unsaturated carbonyl-mediated generation of ROS (Radical Oxigen Species) and might initiate lipid peroxidation and other processes, leading to enhanced cytotoxic/genotoxic cell damage (Janzowski et al., 2003). Hence, the presence of a terminal double bound without electron donating substituent makes these compounds more mutagenic. Based on this, we can say that, following the Michael addition mechanism, the presence of electron-withdrawing substituents in the double bound (i.e. fragment F_7) increase the mutagenicity of the substance (Aptula and Roberts, 2006; Schultz et al., 2005).

Compared with the results of the Ames test, the reactivity as Michael acceptors is more pronounced in MCGM, judging by the higher variation of its contribution values. As mentioned above, this is most likely because Michael acceptors are soft electrophiles and as such reactivity is higher towards soft nucleophiles like GSH, then such appears to be the main mechanism step producing DNA damage in mammalian cells for these substances (Glaab et al., 2001; Janzowski et al., 2003).

3.4. Applicability domain

It would be very interesting to have a predictive model for the vast majority of chemicals, especially for those who have not been tested yet and thus, with unknown mutagenicity, in particular taking into account that the European Union is launching the REACH standard. Since this is usually not possible, one should define the applicability domain of the QSAR model, that is, the range within which it bears a new compound. For that purpose, we built a Williams plot using the leverage values calculated for each compound. As seen in Fig. 6, 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^* ($= 0.120$ and $= 0.307$ for AMES and MCGM, respectively), save for compounds 2, 3, 5, 8, 42, 73, 87, 131, 133 and 165 (AMES) and for compounds 213 and 220 (MCGM). Even so, the latter should not be considered outliers but influential chemicals (Eriksson et al., 2003).

Nevertheless, all evaluations pertaining to the external set were performed by taking into account the applicability domain of our QSAR model. So, if a chemical belonging to the test set had a leverage value greater than h^*, we consider that this means that the prediction is the result of substantial extrapolation and therefore may be unreliable (Netzeva et al., 2005).

4. Conclusions

Herein, we have examined the ability of the TOPS-MODE approach to provide discriminant models for probing the mutagenicity of the α, β-unsaturated carbonyl compounds over two endpoints: the Ames and mammalian cell mutation gene tests.

With regard to the QSAR modelling, the combination of LDA in conjunction with the TOPS-MODE structure representation was found to produce final classification models with high sensitivity, specificity and accuracy. The predictive power of such QSAR models was proved to even exceed that of state-of-art expert systems, such as the TOXTREE software, for this family of compounds. Furthermore, due to the ability of TOPS-MODE to express the activity at a local level, we could obtain a series of structural alerts for each compound and both endpoints under study. Among such alerts, the halogenated α, β-unsaturated carbonyls and the 2-furanone ring should be studied further in what concerns their Ames mutagenicity. As regards the mammalian cell gene mutation end-point, the presence of a terminal double bound with electron-withdrawing or without electron-donating substituents turns the compounds more mutagenic probably because they act through a Michael type addition mechanism. For both endpoints, we note that the predominant mechanism is Michael type addition by forming adducts either with DNA or with GSH. Moreover, by carefully analysing the fragment contributions obtained with the TOPS MODE approach, we were able to propose possible mutagenic mechanisms for a number of false negatives and false positives compounds settled on by the expert system TOXTREE in relation to the Ames data. In addition, the TOPS-MODE approach was able to quantify the influence of several molecular environments (modulating factors) to the structural alerts that describe the mutagenicity of the α, β-unsaturated carbonyl moiety. Overall, that structural information and the QSAR models per se can definitely aid in future improvements of software or experts systems based on SAs.

Conflict of interest

None.

Acknowledgements

The authors acknowledge to MODESLAB 1.0 software owners for delivering a free copy of such program. AMH acknowledges the *Portuguese Fundação para a Ciência e a Tecnologia* (FCT - Lisboa) (SFRH/BD/22692/2005) for financial support. APG acknowledges to Dr. Antonio Pérez-Garrido for his valuable collaboration in the completion of this work.

Appendix A. Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.tox.2009.11.023.

References

Aptula, A.O., Roberts, D.W., 2006. Mechanistic applicability domains for non-animal-based toxicological end points: general principles and application to reactive toxicity. Chem. Res. Toxicol. 19, 1097–1105.

Belsito, D., Bickers, D., Bruze, M., Calow, P., Greim, H., Hanifin, J.M., Rogers, A.E., Saurat, J.H., Sipes, I.G., Tagami, H., 2007. A toxicologic and dermatologic assessment of related esters and alcohols of cinnamic acid and cinnamyl alcohol when used as fragrance ingredients. Food Chem. Toxicol. 45, S1–S23.

Benigni, R., Bossa, C., 2008. Structure alerts for carcinogenicity, and the Salmonella assay system: a novel insight through the chemical relational databases technology. Mutat. Res. 659, 248–261.

Benigni, R., Bossa, C., Jeliazkova, N.G., Netzeva, T.I., Worth, A.P. (Eds.), 2008. The Benigni/Bossa rulebase for mutagenicity and carcinogenicity-a module of Toxtree. EUR 23241 EN, EUR-Scientific and Technical Report Series, Office for the Official Publications of the European Communities, Luxenbourg.

Benigni, R., Conti, L., Crebelli, R., Rodomonte, A., Vari, M.R., 2005. Simple and α, β-unsaturated aldehydes: Correct prediction of genotoxicity activity through structure–activity relationship models. Environ. Mol. Mutagen. 46, 268–280.

Benigni, R., Passerini, L., Gallo, G., Giorgi, F., Cotta-Ramusino, M., 1998. QSAR models for discriminating between mutagenic and nonmutagenic aromatic and heteroaromatic amines. Environ. Mol. Mutagen. 32, 75–83.

Boelens, M.H., Gemert, L.J., 1987. Organoleptic properties of aliphatic aldehydes. Perfumer Flavorist 12, 31–43.

Boyland, E., Nery, R., 1969. Mercapturic acid formation during metabolism of arecoline and arecaidine in the rat. Biochem. J. 113, 123–130.

Chan, P.K., Hayes, W.A., Siraj, M.Y., 1982. Excretion of conjugated metabolites of the mycotoxin penicillic acid in male mice. Toxicol. Appl. Pharm. 66, 259–326.

Chan, P.K., Hayes, W.A., Siraj, M.Y., Meydrech, E.F., 1984. Pharmacokinetics of the mycotoxin penicillic acid in male mice: Absorption, distribution, excretion, and kinetics. Toxicol. Appl. Pharm. 73, 195–203.

Ciaccio, P.J., Gicquel, E., O'Neill, P.J., Scribner, H.E., Vandenberghe, Y.L., 1998. Investigation of the positive response of ethyl acrylate in the mouse lymphoma genotoxicity assay. Toxicol. Sci. 46, 324–332.

Cohen, J., 1960. A coefficient of agreement for nominal scales. J. Educ. Psychol. Meas. 20, 37–46.

Dearfield, K.L., Harrington-Brock, K., Doerr, C.L., Rabinowitz, J.R., Moore, M.M., 1991. Genotoxicity in mouse lymphoma cells of chemicals capable of michael addition. Mutagenesis 6, 519–525.

Dearfield, K.L., Millis, C.S., Harrington-Brock, K., Doerr, C.L., Moore, M.M., 1989. Analysis of genotoxicity of nine acrylatermethacrylate compounds in l5178y mouse lymphoma cells. Mutagenesis 4, 381–393.

Duchowicz, P.R., Castro, E.A., Fernndez, F.M., 2006. Alternative algorithm for the search of an optimal set of descriptors in QSAR-QSPR studies. MATCH Commun. Math. Comput. Chem. 55, 179–192.

Dutertre-Catella, H., Nguyen, P.L., Dang Quoc, Q., Truhaut, R., 1978. Transformations métaboliques de la trimethyl-3,5,5-cyclohexene-2, one-1 (isophorone). Toxicol. Eur. Res. 1, 209–216.

Eder, E., Hoffman, C., Bastian, H., Deininger, C., Scheckenbach, S., 1990. Molecular mechanisms of dna damage initiated by α, β-unsaturated carbonyl compounds as criteria for genotoxicity and mutagenicity. Environ. Health Perspect. 88, 99–106.

Eder, E., Weinfurtner, E., 1994. Mutagenic and carcinogenic risk of oxigen containing chlorinated c-3 hydrocarbons: Putative secondary products of c-3 chlorohydrocarbons and chlorination of water. Chemosphere 29, 2455–2466.

Eriksson, L., Jaworska, J., Worth, A.P., Cronin, M.T.D., McDowell, R.M., 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.

Estrada, E., 1995. Edge adjacency relationships and a novel topological index related to molecular volume. J. Chem. Inf. Comput. Sci. 35, 31–33.

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.

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.

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.

Estrada, E., Molina, E., 2006. Automatic extraction of structural alerts for predicting chromosome aberrations of organic compounds. J. Mol. Graph. Model. 25, 275–288.

Estrada, E., Molina, E., Uriarte, E., 2001. Quantitative structure–toxicity relationships using tops-mode. 2. Neurotoxicity of a non-congeneric series of solvents. SAR QSAR Environ. Res. 12, 445–459.

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.

Estrada, E., Patlewicz, G., Chamberlain, M., Basketter, D., Larbey, S., 2003a. Computer-aided knowledge generation for understanding skin sensitization mechanisms: the tops-mode approach. Chem. Res. Toxicol. 16, 1226–1235.

Estrada, E., Uriarte, E., 2001. Quantitative structure–toxicity relationships using tops-mode. 1. Nitrobenzene toxicity to tetrahymena pyriformis. SAR QSAR Environ. Res. 12, 309–324.

Estrada, E., Uriarte, E., Gutierrez, Y., González, H., 2003b. 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.

Feron, V.J., Til, H.P., de Vrijer, F., Woutersen, R.A., Cassee, F.R., van Bladeren, P.J., 1991. Aldehydes: ocurrence, carcinogenic potential, mechanism of action and risk assessment. Mutat. Res. 259, 363–385.

Fragkakia, A.G., Angelis, Y.S., Tsantili-Kakoulidouc, A., Koupparisb, M., Georgakopoulo, C., 2009. Schemes of metabolic patterns of anabolic androgenic steroids for the estimation of metabolites of designer steroids in human urine. J. Steroid Biochem. Mol. Biol. 115, 44–61.

Frank, J., 2002. STATISTICA, 8th ed. Statsoft, Inc.

García-Lorenzo, A., Tojo, E., Teijeira, M., Rodríguez-Berrocal, F.J., González, M.P., Martínez-Zorzano, V.S., 2008. Cytotoxicity of selected imidazolium-derived ionic liquids in the human caco-2 cell line. sub-structural toxicological interpretation through a QSAR study. Green Chem. 10, 508–516.

Glaab, V., Collins, A.R., Eisenbrand, G., Janzowski, C., 2001. DNA damaging potential and glutathione depletion of 2-cyclohexene-1-one inmammalian cells, compared to food relevant 2-alkenals. Mutat. Res. 497, 185–197.

González, M.P., Dias, L.C., Helguera, A.M., 2004a. A topological sub-structural approach to the mutagenic activity in dental monomers. 2. Cycloaliphatic epoxides. Polymer 15, 5353–5359.

González, M.P., Helguera, A.M., Cabrera, M.A., 2005a. Quantitative structure–activity relationship to predict toxicological properties of benzene derivative compounds. Bioorg. Med. Chem. 13, 1775–1781.

González, M.P., Helguera, A.M., Collado, I.G., 2006. A topological substructural molecular design to predict soil sorption coefficients for pesticides. Mol. Divers. 10, 109–118.
González, M.P., Helguera, A.M., Molina, R.R., García, J.F., 2004b. A topological substructural approach of the mutagenic activity in dental monomers. 1. Aromatic epoxides. Polymer 45, 2773–2779.
González, M.P., Teran, M.C.T., Fall, Y., Dias, L.C., Helguera, A.M., 2005b. A topological sub-structural approach to the mutagenic activity in dental monomers. 3. Heterogeneous set of compounds. Polymer 46, 2783–2790.
Gramatica, P., 2007. Principles of QSAR models validation: internal and external. QSAR Comb. Sci., 1–9.
Gutierrez, Y., Estrada, E., 2002. Modes lab, version 1.0.
Hashimoto, K., Tanii, H., 1985. Mutagenicity of acrylamide and its analogues in salmonella typhimurium. Mutat. Res. 158, 129–133.
Hedman, R., Pettersson, H., 1997. Transformation of nivalenol by gastrointestinal microbes. Arch. Anim. Nutr. 50, 321–329.
Helguera, A.M., Cabrera, M.A., Combes, R.D., González, M.P., 2006. Quantitative structure activity relationship for the computational prediction of nitrocompounds carcinogenicity. Toxicology 220, 51–62.
Helguera, A.M., Cabrera, M.A., González, M.P., Molina, R.R., González, H.D., 2005. A topological sub-structural approach applied to the computational. prediction of rodent carcinogenicity. Bioorg. Med. Chem. 13, 2477–2488.
Helguera, A.M., Cordeiro, M.N.D.S., Cabrera, M.A., Combes, R.D., González, M.P., 2008a. 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.
Helguera, A.M., González, M.P., Briones, J.B., 2004. Tops-mode approach to predict mutagenicity in dental monomers. Polymer 45, 2045–2050.
Helguera, A.M., González, M.P., Cordeiro, M.N.D.S., Cabrera, M.A., 2007. Quantitative structure carcinogenicity relationship for detecting structural alerts in nitroso-compounds. Toxicol. Appl. Pharmacol. 221, 189–202.
Helguera, A.M., González, M.P., Cordeiro, M.N.D.S., Cabrera, M.A., 2008b. 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.
Holder, A.J., Ye, L., 2009. Quantum mechanical quantitative structure–activity relationships to avoid mutagenicity. Dent. Mater. 25, 20–25.
Janzowski, C., Glaab, V., Mueller, C., Straesser, U., Kamp, H.G., Eisenbrand, G., 2003. α, β-unsaturated carbonyl compounds: induction of oxidative dna damage in mammalian cells. Mutagenesis 18, 465–470.
Joe, Y.A., Goo, Y.-M., Lee, Y.Y., 1989. Microbiological oxidation of isophorone to 4-hydroxyisophorone and chemical transformation of 4-hydroxyisophorone to 2,3,5-trimethyl-p-benzoquinone. Arch. Pharm. Res. 12, 73–78.
Kazius, J., McGuire, R., Bursi, R., 2005. Derivation and validation of toxicophores for mutagenicity prediction. J. Med. Chem. 48, 312–320.
Klein, D., Randić, M., Babic, D., Lucic, B., Nikolic, S., Trinajstic, N., 1997. Hierarchical orthogonalization of descriptors. Int. J. Quantum Chem. 63, 215–222.
Koleva, Y.K., Madden, J.C., Cronin, M.T.D., 2008. Formation of categories from structure–activity relationships to allow read-across for risk assessment: toxicity of α, β-unsaturated carbonyl compounds. Chem. Res. Toxicol. 21, 2300–2312.
Kubinyi, H., 1994a. Variable selection in QSAR studies. 1. An evolutionary algorithm. Quant. Struct. Act. Relat. 13, 285.
Kubinyi, H., 1994b. Variable selection in QSAR studies. 2. A highly efficient combination of systematic search and evolution. Quant. Struct. Act. Relat. 13, 393.
Lalko, J., Lapczynski, A., McGinty, D., Bhatia, S., Letizia, C.S., Api, A.M., 2007. Fragrance material review on beta-ionone. Food Chem. Toxicol. 45, S241–S247.
Lalonde, R.T., Cook, G.P., Perakyla, H., Bu, L., 1991. Structure–activity-relationships of bacterial mutagens related to 3-chloro-4-(dichloromethyl)-5-hydroxy-2(5h)-furanone an emphasis on the effect of stepwise removal of chlorine from the dichloromethyl group. Chem. Res. Toxicol. 4, 540–545.
Landis, J.R., Koch, G.G., 1977. The measurement of observer agreement for categorical data. Biometrics 33, 159–174.
Lucic, B., Nikolic, S., Trinajstic, N., Juric, D., 1995. The structure–property models can be improved using the orthogonalized descriptors. J. Chem. Inf. Comput. Sci. 35, 532–538.
Lynch, D.M., Crovetti, A.J., 1972. Reactions of dichloromaleimides with alcohols, phenols, and thiols. J. Heterocycl. Chem. 9, 1027–1032.
McDonald, T.A., Komulainen, H., 2005. Carcinogenicity of the chlorination disinfection by-product mx. J. Environ. Sci. Health, Part C 23, 163–214.
Miller, I.A., Miller, E.C., 1983. Some historical aspects of n-aryl carcinogens and their metabolic activation. Environ. Health Perspect. 49, 3–12.
Nery, R., 1971. The metabolic interconversion of arecoline and arecoline-1-oxide in the rat. Biochem. J. 122, 503–508.
Netzeva, T.I., Worth, A.P., Aldenberg, T., Benigni, R., Cronin, M.T.D., Gramatica, P., Jaworska, J.S., Kahn, S., Klopman, P., Marchant, C.A., Myatt, G., Nikolova-Jeliazkova, N., Patlewicz, G.Y., Perkins, R., Roberts, D.W., Schultz, T.W., Stanton, D.T., van de Sandt, J.J.H., 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.
Noma, Y., Takahashi, H., Asakawa, Y., 1991. Biotransformation of terpene aldehydes by euglena gracilis z. Phyrochemiary 30, 1147–1151.
OECD, 2007. Guidance document on the validation of (quantitative) structure–activity relationships [(Q)SAR] models, 49, OECD series on testing and assessment. Tech. Rep., Organisation for Economic Co-operation and Development, Paris, France.
Pérez-Garrido, A., González, M.P., Escudero, A.G., 2008. Halogenated derivatives QSAR model using spectral moments to predict haloacetic acids (haa) mutagenicity. Bioorg. Med. Chem. 16, 5720–5732.
Pérez-Garrido, A., Helguera, A.M., Girón-Rodríguez, F., Cordeiro, M.N.D.S., in press. QSAR models to predict mutagenicity of acrylates, methacrylates and α, β-unsaturated carbonyl compounds. Dent. Mater.
Provost, F., Fawcett, T., 2001. Robust classification for imprecise environments. Mach. Learn. 42, 3.
Purohit, V., Basu, A.K., 2000. Mutagenicity of nitroaromatic compounds. Chem. Res. Toxicol. 13, 673–692.
Randić, M., 1991a. Correlation of enthalpy of octanes with orthogonal connectivity indices. J. Mol. Struct. (Theochem.) 233, 45–59.
Randić, M., 1991b. Orthogonal molecular descriptors. N. J. Chem. 15, 517–525.
Randić, M., 1991c. Resolution of ambiguities in structure–property studies by use of orthogonal descriptors. J. Chem. Inf. Comput. Sci. 31, 311–320.
Richard, C., 2001. Concise international chemical assessment document 36: methyl cyanoacrylate and ethyl cyanoacrylate (draft). Tech. Rep., United Nations Environment Programme, International Labour Organization, and the World Health Organization, Geneva.
Sasaki, J.C., Feller, R.S., Colvin, M.E., 2002. Metabolic oxidation of carcinogenic arylamines by p450 monooxygenases: theoretical support for one-electron transfer mechanism. Mutat. Res. 506/507, 79–89.
Schultz, T.W., Yarbrough, J.W., Johnson, E.L., 2005. Structure–activity relationships for reactivity of carbonyl-containing compounds with glutathione. SAR QSAR Environ. Res. 16, 313–322.
Smith, T.L., 1987. Method for detecting molecules containing amine or thiol groups. U.S. Pat. 4 680 272.
Sosted, H., Basketter, D.A., Estrada, E., Johansen, J.D., Patlewicz, G.Y., 2004. Ranking of hair dye substances according to predicted sensitization potency: quantitative structure–activity relationships. Contact Dermat. 51, 241–254.
Van Beerendonk, G.J.M., Nivard, M.J.M., Vogel, E.W., Nelson, S.D., Meerman, J.H.N., 1992. Formation of thymidine adducts and cross-linking potencial of 2-bromoacrolein, a reactive metabolite of tris(2,3-dibromopropyl)phosphate. Mutagenesis 7, 19–24.
van Noort, R., Brown, D., Causton, B.E., Combe, E.C., Fletcher, A.M., Lloyd, C.H., McCabe, J.F., Piddock, V., Sherriff, M., Strang, R., Waters, N.E., Watts, D.C., Williams, K., 1990. Dental materials: 1989 literature review. J. Dent. 18, 327–352.
Vighi, M., Gramatica, P., Consolaro, F., Todeschini, R., 2001. Qsar and chemometrics approaches for setting water quality objetives for dangerous chemicals. Ecotoxicol. Environ. Saf. 49, 206–220.
Weininger, D., 1988. Smiles, a chemical languaje and information system. 1. introduction to methodology and encoding rules. J. Chem. Inf. Comput. Sci. 28, 31–36.
Yang, C., Hasselgren, C.H., Boyer, S., Arvidson, K., Aveston, S., Dierkes, P., Benigni, R., Benz, R.D., Contrera, J., Kruhlak, N.L., Matthews, E.J., Han, X., Jaworska, J., Kemper, R.A., Rathman, J.F., Richard, A.M., 2008. Understanding genetic toxicity through data mining: The process of building knowledge by integrating multiple genetic toxicity databases. Toxicol. Mech. Methods 18 (2–3), 277–295.
Yourtee, D., Holder, A.J., Smith, R., Morrill, J.A., Kostoryz, E., Brockmann, W., Glaros, A., Chappelow, C., Eick, D., 2001. Quantum mechanical quantitative structure activity relationships to avoid mutagenicity in dental monomers. J. Biomater. Sci. Polym. Ed. 12, 89–105.