Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis

journal homepage: www.elsevier.com/locate/molmut
Community address: www.elsevier.com/locate/mutres

QSARs of aromatic amines: Identification of potent carcinogens

Rainer Franke^a, Andreas Gruska^a, Cecilia Bossa^b, Romualdo Benigni^{b,*}

^a Consulting in Drug Design GbR, Gartenweg 14, D-16348 Wandlitz OT Basdorf, Germany

^b Istituto Superiore di Sanita', Health and Environment Department, Viale Regina Elena 299, 00161 Rome, Italy

ARTICLE INFO

Article history:

Received 22 April 2010

Received in revised form 4 June 2010

Accepted 11 June 2010

Available online 23 June 2010

Keywords:

Cancer

Structure–activity

Modeling

Carcinogenic potency

Prediction

ABSTRACT

In previous investigations, we have developed Quantitative Structure–Activity Relationships (QSAR) models for a series of aromatic amines based on well defined physicochemical descriptors: these QSARs were aimed at: (a) describing the modulation of the carcinogenic potency among the active ones only; and (b) modeling the separation between carcinogens and non-carcinogens. In this analysis based on a larger range of chemicals, we checked and confirmed the validity and robustness of the previous models. Since the identification of high potency carcinogens (which pose the highest risk to human health) is particularly relevant to risk assessment, we also established a new QSAR model that points directly to aromatic amines likely to have high carcinogenic potency.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

Risk assessment is a key issue in environmental and medicinal chemistry. Among other biological end points, carcinogenicity is one of the most important toxicological issues when characterizing chemical compounds. As it is impossible to experimentally investigate all chemicals in the human environment, the prediction of carcinogenic activity of chemicals is of great practical importance. This is also true in the discovery process for new therapeutics. Together with the short-term tests for the pre-screening of carcinogenicity (such as the mutagenicity-based assays [1,2]), a very useful avenue for the early estimation of toxicity is the use of (Quantitative) Structure–Activity Relationships ((Q)SAR) models [3]. Because of the complex nature of the processes involved in chemical carcinogenesis this is an extremely difficult task. There are two possible strategies. One can try to create a black-box model from a great number of observations for diverse chemical structures and chemical descriptors using, for example, machine learning methods, or rule based approaches, with the assumption that the resulting model might be of general validity. We feel, however, that a general model for carcinogenicity is unlikely to exist as compounds from different chemical families act via very different mechanisms. A possible way to at least partly overcome this difficulty is to evaluate QSAR models for classes of chemicals with less compounds but greater homology and to use these models, after assigning the compounds to be evaluated to the correct

class, for predictions [4]. The latter approach is of special relevance in the cases – such as with the aromatic amines – where the large amount of available experimental determinations of toxicity [5] provides a sound basis for the modeling. It should be emphasized as well that the application of estimation methods to the toxicity of aromatic amines is an urgent need; in spite of the many data available on “old” amines, new amines continue to be synthesized (e.g., for hair dyes) and need to be regulated within new regulatory frameworks that limit or prohibit the use of animal testing (see, for example, the new regulation of chemicals called REACH, and the Cosmetics Directive of the European Union [6–8]).

In previous investigations, we have developed QSAR models for a series of aromatic amines [9,10] based on well defined physicochemical descriptors: these chemicals will be called series_old throughout this publication. These QSARs were aimed at (a) describing the modulation of the carcinogenic potency among the active ones only; and (b) modeling the separation between carcinogens and non-carcinogens. The two operations were performed through separate sets of equations, since it has been demonstrated that different patterns of chemical descriptors are needed to describe potency and yes/no activity, separately [10]. In a subsequent paper, a preliminary check for the ability of the models to predict the carcinogenic activity of new, external chemicals was performed [11].

More recently, another series of aromatic amines was analyzed by our substructural method EVAL [12] but not yet in terms of physicochemical parameters. This second series, which will be called series_new, has more data than the first one with some overlap between the two series. The specific aim of the more recent

* Corresponding author. Tel.: +39 06 49902579; fax: +39 06 49902999.

E-mail address: romualdo.benigni@iss.it (R. Benigni).

paper was the identification of substructural motifs that determine high carcinogenic potency. As a matter of fact, the identification of high potency carcinogens is particularly relevant to risk assessment, since these carcinogens pose the highest risk to human health.

In the present work, we wanted to learn whether the QSARs developed in [9,10] for mouse carcinogenicity for series_old also apply to series_new and vice versa. The robustness and repeatability of QSAR models is an essential element to judge about the value of a model, in addition to purely statistical ones. Another goal is to establish QSARs for directly identifying potent carcinogens through the use of physical chemical descriptors. The analysis is restricted to mouse carcinogenicity data, because mouse is the experimental model most widely used. To this end, QSAR analyses were performed for series_new using physicochemical descriptors in comparison with QSARs available for the old series. In addition, predictions were made from the QSARs derived from the old series for those compounds in the new series not also occurring in the old series and, vice versa, from the QSARs from the new series for those compounds of the old series not also occurring in the second one. This is very close to true predictions for new compounds.

2. Methods

Table 1 summarizes the structures of the compounds (mainly anilines, biphenylamines, naphthylamines, aminofluorenes, anthracenes, and derivatives) for which carcinogenic potency data were available. The carcinogenicity data were retrieved from the ISSCAN experimental carcinogenicity database, which is freely downloadable from the website of the Istituto Superiore di Sanita': <http://www.iss.it/ampp/dati/cont.php?id=233&lang=1&tipo=7>. Details on the ISSCAN database can be found in [13].

Compounds belonging to series_old and series_new are characterized by the indicator "Series" (= 1 for series_old compounds and = 2 for series_new compounds). Compounds appearing in only one of the two series are characterized by the indicators "only_old" and "only_new".

Carcinogenic potency in mice is defined as

$$\text{BRM} = \log \left(\frac{\text{MW}}{\text{TD}_{50}} \right) + 3$$

where TD_{50} is the dose producing tumors in 50% of the test animals in mg/kg body weight, and MW is the molecular weight. The scaling factor "3" is used to transform all the potency values into positive numbers, thus facilitating and making more intuitive the subsequent analyses.

For the inactive compounds TD_{50} was arbitrarily set equal to 10,000 so that, for these compounds

$$\text{BRM} = \log \left(\frac{\text{MW}}{10,000} \right) + 3$$

The resulting BRM values are only used for classification and not in regression analysis.

To describe chemical properties of the compounds, global and local parameters were used. Global electronic properties were characterized by EHOMO (energy of the highest occupied molecular orbital) and ELUMO (energy of the lowest empty molecular orbital) calculated by the PM3 semi-empirical Hamiltonian calculation method after optimizing structures (program system SYBYL, Tripos). Overall hydrophobicity is expressed in terms of $\log P$ computed from our in-house program.

In an attempt to get some insight into possible local effects ring substituents were characterized by hydrophobic, electronic, and steric substituent constants. To describe steric properties of substituent R at the functional amino group MR (molar refractivity; values scaled by 10^{-1}), and Charton's E_s values were used. To define ring positions the definitions described in [9,10] were used considering all compounds as substituted anilines as presented in Table 1. For the biphenylamines (see Fig. 1) substituents in the aniline part are characterized as in substituted anilines. In cases 1 and 2 the second part of the molecule (second phenyl ring plus substituents at this ring) is then treated as a para-substituent where the bridge X may be present or absent. Only MR values are available here. In case 3 the non-aniline part appears as ortho-substituent and is fully parameterized with MR. In the case of the naphthylamines (see Fig. 2) two situations are possible. They are treated as anilines substituted by the substructure $-C_4H_4-$, with an estimated MR of 0.8. This amount is equally distributed over the positions of substitution so that $\text{MR}_2 = \text{MR}_3 = 0.8$ in case 1 and $\text{MR}_3 = \text{MR}_4 = 0.8$ in case 2. If additional substituents occur the MR values are correspondingly corrected. For the aminofluorenes and derivatives, only steric effects (MR) could be parameterized fol-

Fig. 1. Treatment of substituents on biphenylamines.

Fig. 2. Treatment of substituents on naphthylamines.

lowing the scheme presented in Fig. 3. Anthracenes, finally, are treated as shown in Fig. 4.

In some compounds of series_new the functional amino group is attached to a 5-membered ring. For these compounds positions cannot be defined so that they were not included into the present analysis. They are, however, included into Table 1 because they are needed for QSAR analyses based on substructural parameters to be presented in a subsequent paper.

All chemical parameters appearing in the QSARs as well as carcinogenic potencies are summarized in Table 1.

There are differences between compounds with one and with two amino groups which are accounted for by the indicator variables $I(\text{monoNH}_2)$ (= 1, if only one amino group is present) and $I(\text{diNH}_2)$ (= 1, if more than one amino group is present).

The methods used in the evaluation of QSARs are multiple regression analysis, and discriminant analysis. To check the predictive power of the old QSARs (QSARs_old) for series_new and of the new QSARs (QSARs_new) for series_old, the compounds in series_new not also appearing in series_old and the compounds in series_old not also appearing in series_new were used as test sets. The compounds belonging to these test sets are indicated in Table 1. For QSARs_new, additional test sets were stochastically selected from the compounds of series_new which are also indicated in Table 1: test_1 for regression analysis and test_2 and test_3 for discriminant analysis.

3. Results

3.1. QSARs for the modulation of the potency of carcinogenic aromatic amines

From the active compounds of the new series the following equation is obtained for BRM (the biphenyls #146, 155, 168 and

Table 1
Chemical structures, CAS numbers, carcinogenic activity, and parameters.

No.	CAS	Structure			
		Ring	Bridge	AnX	R
1	54827-17-7	An		2,6-Me ₂ ,4-Ph-3,5-Me ₂ ,4-NH ₂	H ₂
2	28322-02-3	F		2,3-Me ₂	H, COMe
3	134-32-7	N		2-C ₄ H ₄ -3	H ₂
4	91-59-8	N		3-C ₄ H ₄ -4	H ₂
5	92-87-5	Bi		4-Ph-4-NH ₂	H ₂
6	53-96-3	F		3,4-Me ₂	H, COMe
7	101-21-3	An		3-Cl	H, COOiPr
8	101-14-4	Bi	CH ₂	2-Cl,4-Ph-3-Cl,4-NH ₂	H ₂
9	636-21-5	An		2-Me	H ₂
10	2465-27-2	Bi	C=NH ₂	4-C(=NH)-Ph-4-N(Me) ₂	Me ₂
11	2185-92-4	Bi		2-Ph	H ₂
12	609-20-1	An		2,6-Cl ₂ ,4-NH ₂	H ₂
13	2784-94-3	An		2-NO ₂ ,4-N(C ₂ H ₄ OH) ₂	H, Me
14	13552-44-8	Bi	CH ₂	4-CH ₂ -Ph-4-NH ₂	H ₂
15	150-68-5	An		4-Cl	H, CONMe ₂
16	101-80-4	Bi	O	4-O-Ph-4-NH ₂	H ₂
17	87-62-7	An		2,6-Me ₂	H ₂
18	33229-34-4	An		2-NO ₂ ,4-N(C ₂ H ₄ OH) ₂	H, C ₂ H ₄ OH
19	15481-70-6	An		2-Me,3-NH ₂	H ₂
20	20265-97-8	An		4-OMe	H ₂
21	118-92-3	An		2-COOH	H ₂
22	140-49-8	An		4-COCH ₂ Cl	H, COMe
23	61702-44-1	An		2-Cl,4-NH ₂	H ₂
24	54150-69-5	An		2,4-OMe ₂	H ₂
25	315-18-4	An		2,6-Me ₂ ,4-OCONHMe	Me ₂
26	1465-25-4	N		2-C ₄ H ₄ -3	H, C ₂ H ₄ NH ₂
27	619-17-0	An		2-COOH,5-NO ₂	H ₂
28	99-56-9	An		2-NH ₂ ,4-NO ₂	H ₂
29	624-18-0	An		4-NH ₂	H ₂
30	101-54-2	Bi	NH	4-NH-Ph-4-NH ₂	H ₂
31	103-85-5	An		H	H, CSNH ₂
32	6369-59-1	An		2-Me,4-NH ₂	H ₂
33	17026-81-2	An		2-OEt,5-NHCOMe	H ₂
34	132-32-1	F		3-Me,4-NHEt	H ₂
35	119-34-6	An		3-NO ₂ ,4-OH	H ₂
36	142-04-1	An		H	H ₂
37	134-29-0	An		2-OMe	H ₂
38	20265-96-7	An		4-Cl	H ₂
39	5131-60-2	An		2-Cl,5-NH ₂	H ₂
40	95-83-0	An		2-NH ₂ ,4-Cl	H ₂
41	95-74-9	An		2-Cl,4-Me	H ₂
42	3165-93-3	An		2-Me,4-Cl	H ₂
43	95-79-4	An		2-Me,5-Cl	H ₂
44	102-50-1	An		2-Me,4-OMe	H ₂
45	120-71-8	An		2-OMe,5-Me	H ₂
46	80-08-0	Bi	SO ₂	4-SO ₂ -Ph-4-NH ₂	H ₂
47	615-05-4	An		2-OMe,5-NH ₂	H ₂
48	95-80-7	An		2-Me,5-NH ₂	H ₂
49	2164-17-2	An		3-CF ₃	H, CONMe ₂
50	101-61-1	Bi	CH ₂	4-CH ₂ -Ph-4-N(Me) ₂	Me ₂
51	90-94-8	Bi	CO	4-CO-Ph-4-N(Me) ₂	Me ₂
52	2243-62-1	N		2-C ₃ H ₃ C(NH ₂)-3	H ₂
53	1777-84-0	An		3-NO ₂ ,4-OEt	H, COMe
54	94-52-0	An		2-OMe,5-NO ₂	H ₂
55	5307-14-2	An		2-NO ₂ ,4-NH ₂	H ₂
56	99-55-8	An		2-Me,5-NO ₂	H ₂
57	139-65-1	Bi	S	4-S-Ph-4-NH ₂	H ₂
58	1582-09-8	An		2,6-(NO ₂) ₂ ,4-CF ₃	(nPr) ₂
59	137-17-7	An		2,4,5-Me ₃	H ₂
60	92-67-1	Bi		4-Ph	H ₂
61	121-88-0	An		2-OH,4-NO ₂	H ₂
62	135-88-6	N		3-C ₄ H ₄ -4	H, Ph
63	137-09-7	An		2-OH,5-NH ₂	H ₂
64	100-01-6	An		4-NO ₂	H ₂
65	4075-79-0	Bi		4-Ph	H, COMe
66	324-93-6	Bi		4-Ph-4-F	H ₂
67	398-32-3	Bi		4-Ph-4-F	H, COMe
68	363-17-7	F		3,4-Me ₂	H, COCF ₃
69	101-79-1	Bi	O	4-O-Ph-4-Cl	H ₂
70	63019-65-8	F		2,3-Me ₂	(COMe) ₂
71	91-94-1	Bi		2-Cl,4-Ph-3-Cl,4-NH ₂	H ₂
72	77-46-3	Bi	SO ₂	4-SO ₂ -Ph-4-NHCOMe	H, COMe
73	62-44-2	An		4-OEt	H, COMe
74	937-25-7	An		4-F	Me,NO

Table 1 (Continued)

No.	CAS	Structure			
		Ring	Bridge	AnX	R
75	614-00-6	An		H	Me,NO
76	615-28-1	An		2-NH ₂	H ₂
77	7411-49-6	Bi		2-NH ₂ ,4-Ph-3,4-(NH ₂) ₂	H ₂
78	63886-77-1	An		2,4,5,6-F ₄ ,3-NH ₂	H ₂
79	6334-11-8	An		2,4,6-Me ₃	H ₂
80	121-69-7	An		H	Me ₂
81	2871-01-4	An		3-NO ₂ ,4NHC ₂ H ₄ OH	H ₂
82	540-23-8	An		4-Me	H ₂
83	99-57-0	An		2-OH,5-NO ₂	H ₂
84	634-93-5	An		2,4,6-Cl ₃	H ₂
85	638-03-9	An		3-Me	H ₂
86	20325-40-0	Bi		2-OMe,4-Ph-3-OMe,4-NH ₂	H ₂
87	59820-43-8	An		2-OC ₂ H ₄ OH,4-NO ₂	H, C ₂ H ₄ OH
88	612-82-8	Bi		2-Me,4-Ph-3Me,4-NH ₂	H ₂
89	133-90-4	An		2,5-Cl ₂ ,3-COOH	H ₂
90	838-88-0	Bi	CH ₂	2-Me,4-CH ₂ -Ph-3-Me,4-NH ₂	H ₂
91	91-59-8	N		3-C ₄ H ₄ -4	H ₂
92	61-82-5	5			H ₂
93	636-21-5	An		2-Me	H ₂
94	2185-92-4	Bi		2-Ph	H ₂
95	609-20-1	An		2,6-Cl ₂ ,4-NH ₂	H ₂
96	108-78-1	An		2,4,6-N ₃ ,3,5-(NH ₂) ₂	H ₂
97	13552-44-8	Bi	CH ₂	4-CH ₂ -Ph-4-NH ₂	H ₂
98	101-80-4	Bi	O	4-O-Ph-4-NH ₂	H ₂
99	15481-70-6	An		2-Me,3-NH ₂	H ₂
100	20265-97-8	An		4-OMe	H ₂
101	118-92-3	An		2-COOH	H ₂
102	61702-44-1	An		2-Cl,4-NH ₂	H ₂
103	54150-69-5	An		2,4-OMe ₂	H ₂
104	619-17-0	An		2-COOH,5-NO ₂	H ₂
105	99-56-9	An		2-NH ₂ ,4-NO ₂	H ₂
106	624-18-0	An		4-NH ₂	H ₂
107	101-54-2	Bi	NH	4-NH-Ph-4-NH ₂	H ₂
108	6369-59-1	An		2-Me,4-NH ₂	H ₂
109	117-79-3	Ant		3,4-(CHO) ₂	H ₂
110	17026-81-2	An		2-OEt,5-NHCOMe	H ₂
111	132-32-1	F		3-Me,4-NHEt	H ₂
112	82-28-0	Ant		2,3-(CHO) ₂ ,6-Me	H ₂
113	119-34-6	An		3-NO ₂ ,4-OH	H ₂
114	121-66-4	5			H ₂
115	142-04-1	An		H	H ₂
116	134-29-2	An		2-OMe	H ₂
117	20265-96-7	An		4-Cl	H ₂
118	5131-60-2	An		2-Cl,5-NH ₂	H ₂
119	95-83-0	An		2-NH ₂ ,4-Cl	H ₂
120	95-74-9	An		2-Cl,4-Me	H ₂
121	3165-93-3	An		2-Me,4-Cl	H ₂
122	95-79-4	An		2-Me,5-Cl	H ₂
123	87-29-6	An		2-COOCH ₂ CH=CHPh	H ₂
124	102-50-1	An		2-Me,4-OMe	H ₂
125	120-71-8	An		2-OMe,5-Me	H ₂
126	80-08-0	Bi	SO ₂	4-SO ₂ -Ph-4-NH ₂	H ₂
127	615-05-4	An		2-OMe,5-NH ₂	H ₂
128	95-80-7	An		2-Me,5-NH ₂	H ₂
129	1937-37-7	Bi	N=N	3-NH ₂ ,4-N=NPh-4-Ph-4-N=N-Xi	H ₂
130	2602-46-2	An		2-CH(OH)CH ₂ C(SO ₃)CH-3,5-SO ₃	H ₂
131	2475-45-8	Ant		2,3-(CHO) ₂ ,4-NH ₂	H ₂
132	99-59-2	An		2-OMe,5-NO ₂	H ₂
133	5307-14-2	An		2-NO ₂ ,4-NH ₂	H ₂
134	99-55-8	An		2-Me,5-NO ₂	H ₂
135	136-40-3	Bi	N=N	2-N,3-NH ₂ ,6-N=NPh	H ₂
136	952-23-8	Ant		3-NHMe,4-Me	H ₂
137	127-69-5	An		4-SO ₂ NHR _i	H ₂
138	139-65-1	An		4-S-Ph-4-NH ₂	H ₂
139	137-17-7	An		2,4,5-Me ₃	H ₂
140	92-67-1	Bi		4-Ph	H ₂
141	121-88-0	An		2-OH,4-NO ₂	H ₂
142	569-61-9	An	CH(Ph-4-NH ₂ ⁺)	p-CH(Ph-4-NH ₂) ₂	H ₂
143	396-01-0	N		2,4-(N)2,3-NCHCN-4	H ₂
144	100-01-6	An		4-NO ₂	H ₂
145	67730-10-3	F		2-N,5,6-(NMe) ₂	H ₂
146	97-56-3	An	N=N	2-Me,4-N=NPh-2-Me	H ₂
147	3693-22-9	F		3-Me,4-OMe	H ₂
148	76180-96-6	5			H ₂
149	68006-83-7	F		2-N,3-NHMe,4,6-Me ₂	H ₂

Table 1 (Continued)

No.	CAS	Structure					
		Ring	Bridge	AnX	R		
150	2104-09-8	5			H ₂		
151	324-93-6	Bi		4-Ph-4-F	H ₂		
152	75104-43-7	F		2-N,3,4,6-Me3,5-NHMe	H ₂		
153	38514-71-5	5			H ₂		
154	26049-71-8	An		p-Ri ₂	H ₂		
155	101-79-1	Bi	O	4-O-Ph-4-Cl	H ₂		
156	77094-11-2	5			H ₂		
157	67730-11-4	F		2-N,5,6-(NMe) ₂	H ₂		
158	53-95-2	F		3,4-Me ₂	COMe,OH		
159	615-28-1	An		2-NH ₂	H ₂		
160	7411-49-6	Bi		2-NH ₂ ,4-Ph-3,4-(NH ₂) ₂	H ₂		
161	63886-77-1	An		2,4,5,6-F4,3-NH ₂	H ₂		
162	6334-11-8	An		2,4,6-Me ₃	H ₂		
163	2871-01-4	An		3-NO ₂ ,4NHC ₂ H ₄ OH	H ₂		
164	540-23-8	An		4-Me	H ₂		
165	99-57-0	An		2-OH,5-NO ₂	H ₂		
166	634-93-5	An		2,4,6-Cl ₃	H ₂		
167	638-03-9	An		3-Me	H ₂		
168	20325-40-0	Bi		2-OMe,4-Ph-3-OMe,4-NH ₂	H ₂		
169	7336-20-1	Bi	CH=CH	3-SO ₃ ,4-CH=CHPh-2-SO ₃ ,4-NH ₂	H ₂		
170	612-82-8	Bi		2-Me,4-Ph-3Me,4-NH ₂	H ₂		
171	133-90-4	An		2,5-Cl ₂ ,3-COOH	H ₂		
172	1918-02-1	An		2,3,6-Cl ₃ ,4-N,5-COOH	H ₂		
173	26148-68-5	F		2-N,3-NHMe,4-Me	H ₂		
174	28754-68-9	5			H ₂		
175	91-76-9	Bi		2,4,6-N3,3-NH ₂ ,5-Ph	H ₂		
176	99-30-9	An		2,6-Cl ₂ ,4-NO ₂	H ₂		
177	5803-51-0	Bi	CH=CH	4-CH=CHPh-2,5-OMe ₂	H ₂		
No.	Carcinogenic activity		Parameters				
	BRM _{obs}	BRM _{eq1}	log P	EHOMO	ELUMO	MR2	MR6
1			4.03	-8.088	0.044	0.56	0.56
2			2.61	-9.015	-0.633	0.56	0.1
3			2.27	-8.262	-0.382	0.8	0.1
4	3.59		2.27	-8.372	-0.367	0.1	0.1
5	3.97		2.16	-8.234	0.001	0.1	0.1
6	4.47		2.61	-8.470	-0.399	0	0.1
7			2.79	-9.162	-0.154	0.1	0.1
8			3.6	-8.559	0.026	0.6	0.1
9	2.11		1.73	-8.553	0.406	0.6	0.1
10	3.63		3.02	-8.348	-0.029	0.1	0.1
11	2.18		2.95	-8.563	0.101	2.54	0.1
12	2.34		1.52	-8.261	-0.089	0.6	0.6
13	3.47		0.34	-8.721	-1074	0.74	0.1
14	3.79		2.56	-8.437	0.314	0.1	0.1
15			1.64	-8.777	-0.119	0.1	0.1
16	3.78		1.91	-8.325	0.163	0.1	0.1
17			2.2	-8.491	0.427	0.56	0.56
18			-0.11	-9.220	-0.899	0.74	0.1
19			0.95	-8.361	0.433	0.56	0.1
20			1.01	-8.310	0.363	0.1	0.1
21			0.96	-8.828	-0.455	0.69	0.1
22			0.8	-9.325	-0.81	0.1	0.1
23			1	-8.163	0.149	0.6	0.1
24			0.76	-8.308	0.26	0.79	0.1
25			2.25	-8.938	0.089	0.56	0.56
26			1.69	-8.528	-0.413	0.8	0.1
27			0.92	-9.429	-1594	0.69	0.1
28			0.43	-9.050	-1026	0.54	0.1
29			0.48	-8.072	0.411	0.1	0.1
30			2.38	-8.046	0.119	0.1	0.1
31			1.86	-8.699	-0.735	0.1	0.1
32			0.95	-8.071	0.403	0.56	0.1
33	1.97		0.2	-8.717	-0.166	1.25	0.1
34	3.74		2.39	-8.013	-0.25	0.1	0.1
35			0.93	-9.039	-0.754	0.1	0.1
36			1.26	-8.609	0.415	0.1	0.1
37	2.11		1.01	-8.571	0.283	0.79	0.1
38	3.15		1.78	-8.591	0.105	0.1	0.1
39	2.06		1	-8.380	0.176	0.6	0.1
40	2.03		1	-8.382	0.111	0.54	0.1
41			2.25	-8.541	0.144	0.6	0.1
42						0.56	0.1
43			2.25			0.56	0.1

Table 1 (Continued)

No.	Carcinogenic activity		Parameters				
	BRM _{obs}	BRM _{eq1}	log <i>P</i>	EHOMO	ELUMO	MR2	MR6
44			1.48	−8.519	0.261	0.56	0.1
45	3.4		1.48	−8.544	0.253	0.79	0.1
46			1.31	−9.026	−0.394	0.1	0.1
47	2.18		0.23	−8.165	0.382	0.79	0.1
48	3.66		0.95	−8.290	0.435	0.56	0.1
49			2	−9.318	−0.523	0.1	0.1
50	3.09		3.71	−8.296	0.343	0.1	0.1
51	3.5		2.85	−8.602	−0.278	0.1	0.1
52	2.99		1.48	−7.989	−0.354	0.8	0.1
53	1.99		0.94	−9.544	−1.189	0.1	0.1
54	2.66		0.96	−9.200	−1.126	0.79	0.1
55			0.43	−8.571	−0.991	0.74	0.1
56			1.68			0.56	0.1
57	3.81		2.25	−8.589	−0.163	0.1	0.1
58	3.01		4.25	−10.264	−1.997	0.74	0.74
59	4.34		2.67	−8.361	0.4	0.56	0.1
60	4.91		2.95	−8.469	−0.064	0.1	0.1
61			0.93	−9.142	−1.068	0.28	0.1
62			4.16	−8.293	−0.393	0.1	0.1
63	2.94		0.2	−8.094	0.343	0.28	0.1
64			1.22	−9.434	−1.012	0.1	0.1
65			2.58	−8.668	−0.223	0.1	0.1
66	5.22		3.09	−8.564	−0.304	0.1	0.1
67			2.72	−8.968	−0.645	0.1	0.1
68			3.73	−9.171	−0.889	0.1	0.1
69	2.8		3.21	−8.784	−0.016	0.1	0.1
70			2.61	−9.062	−0.57		
71			3.2	−8.341	−0.326	0.6	0.1
72			0.57	−9.157	−0.567	0.1	0.1
73	1.92		0.99	−8.697	0.109	0.1	0.1
74			1.83	−9.764	−0.417	0.1	0.1
75			1.69	−9.265	−0.183	0.1	0.1
76	2.17		0.48	−8.332	0.4	0.54	0.1
77	2.74		0.6	−8.101	−0.009	0.54	0.1
78	3.32		1.04	−9.063	−0.742	0.09	0.09
79	3.74		2.67	−8.366	0.438	0.56	0.56
80			1.84	−8.445	0.454	0.1	0.1
81			0.2	−9.227	−0.67	0.1	0.1
82	3.11		1.73	−8.470	0.393	0.1	0.1
83			0.93	−9.075	−1.026	0.28	0.1
84	2.88		2.82	−8.718	−0.301	0.6	0.6
85	1.87		1.73	−8.562	0.414	0.1	0.1
86			1.66	−8.222	−0.16	0.79	0.1
87			0.29				0.1
88	3.87		3.1	−8.134	0.002	0.56	0.1
89	1.6		2	−8.988	−0.634	0.6	0.1
90			3.5	−8.356	0.348	0.56	0.1
91	3.59		2.27	−8.225	−0.175	0.1	0.1
92	3.52		−0.54	−9.029	0.668		
93	2.11		1.73	−8.435	0.601	0.56	0.1
94	2.18		2.95	−8.423	0.116	2.54	0.1
95	2.34		1.52	−8.195	−0.004	0.6	0.6
96	1.1		−0.73	−9.364	0.743	−0.24	−0.24
97	3.79		2.56	−8.309	0.55	0.1	0.1
98	3.78		1.91	−8.102	0.351	0.1	0.1
99	1.09		0.95	−8.245	0.685	0.56	0.1
100	1.09		1.01	−8.204	0.517	0.1	0.1
101	1.14		0.96	−8.661	−0.24	0.69	0.1
102	1.15		1	−8.053	0.296	0.6	0.1
103	1.19		0.76	−8.087	0.495	0.79	0.1
104	1.26		0.92	−9.310	−1.484	0.69	0.1
105	1.18		0.43	−8.363	−0.606	0.54	0.1
106	1.03		0.48	−7.915	0.641	0.1	0.1
107	1.27		2.38	−7.914	0.295	0.1	0.1
108	1.34		0.95	−7.861	0.61	0.56	0.1
109	2.27		1.66	−8.951	−1.254	0.1	0.1
110	1.97		0.2	−8.338	0.242	1.25	0.1
111	3.74		2.39	−7.902	−0.098	0.1	0.1
112	3.13		2.13	−8.710	−1.206	0.69	0.56
113	1.19		0.93	−8.960	−1.120	0.1	0.1
114	1.16		0.76	−9.454	−1.325		
115	0.97		1.26	−8.521	0.64	0.1	0.1
116	2.11		1.01	−8.325	0.544	0.79	0.1
117	3.15		1.78	−8.314	0.472	0.1	0.1
118	2.06		1	−8.310	0.431	0.6	0.1

Table 1 (Continued)

No.	Carcinogenic activity		Parameters				
	BRM _{obs}	BRM _{eq1}	log <i>P</i>	EHOMO	ELUMO	MR2	MR6
119	2.03		1	−8.334	0.262	0.54	0.1
120	1.15		2.25	−8.555	0.285	0.1	0.1
121	3.74		2.25	−8.501	0.299	0.56	0.1
122	2.86		2.25	−8.626	0.234	0.56	0.1
123	1.99		3.3	−8.644	−0.198	4.71	0.1
124	1.14		1.48	−8.148	0.489	0.56	0.1
125	3.4		1.48	−8.267	0.529	0.79	0.1
126	1.39		1.31	−8.888	−0.254	0.1	0.1
127	2.18		0.23	−8.206	0.549	0.79	0.1
128	3.66		0.95	−8.122	0.707	0.56	0.1
129	1.87					0.1	
130	1.92						
131	1.43		−0.69	−7.958	−0.934	0.69	0.1
132	1.66		0.96	−8.706	−0.805	0.79	0.1
133	2.4		0.43	−8.446	−0.801	0.74	0.1
134	2.74		1.68	−9.166	−0.944	0.56	0.1
135	3.48		2.87	−8.022	−0.493	−0.24	3.13
136	1.32		1.97	−8.121	−0.839	0.1	0.1
137	1.43		1.28	−9.095	−0.654	0.1	0.1
138	3.81		2.25	−7.451	0.161	0.1	0.1
139	4.34		2.67	−8.233	0.576	0.56	0.1
140	4.91		2.95	−8.274	0.046	0.1	0.1
141	1.19		0.93	−9.048	−0.915	0.28	0.1
142	3.75		1.88			0.1	0.1
143	3.62		1.37	−8.517	−0.466	−0.24	−0.24
144	1.14		1.22	−9.160	−0.705	0.1	0.1
145	4.06		1.68	−7.986	−0.426	−0.24	0.1
146	1.35		4.5	−8.321	−0.591	0.56	0.1
147	4.64		1.98	−8.116	−0.272	0.1	0.1
148	4		1.38	−8.310	−0.42		
149	3.95		2.42	−8.172	−0.168	−0.24	0.56
150	4.35		2.44	−8.805	−1219		
151	5.22		3.09	−8.347	−0.192	0.1	0.1
152	3.72		2.31	−8.148	−0.21	−0.24	0.56
153	4.43		1.33	−9.121	−1377		
154	4.26		1.8	−8.230	−0.36	0.1	0.1
155	2.8		3.21	−8.589	0.096	0.1	0.1
156	4.24		2.31	−8.336	−0.427		
157	4.56		1.84	−8.115	−0.492	−0.24	0.1
158	4.58		3.1	−8.525	−0.341	0.1	0.1
159	2.17		0.48	−8.195	0.607	0.54	0.1
160	2.87		0.6	−7.849	0.097	0.54	0.1
161	3.32		1.04	−8.788	−0.511	0.09	0.09
162	3.74		2.67	−8.216	0.63	0.56	0.56
163	1.29		0.2	−8.873	−0.576	0.1	0.1
164	3.11		1.73	−8.356	0.616	0.1	0.1
165	1.19		0.93	−9.469	−1078	0.28	0.1
166	2.88		2.82	−8.612	−0.18	0.6	0.6
167	1.87		1.73	−8.477	0.606	0.1	0.1
168	1.39		1.66	−7.900	0.025	0.79	0.1
169	1.57					0.1	0.1
170	3.87		3.1	−7.881	0.182	0.56	0.1
171	1.6		2	−9.065	−0.646	0.6	0.1
172	1.38		1.97	−9.556	−0.989	0.6	0.6
173	3.57		1.95	−8.243	−0.184	−0.24	0.1
174	3.3		1.3	−9.555	−1418		
175	1.27		1.44	−9.400	−0.291	−0.24	−0.24
176	1.32		2.25	−9.362	−1096	0.6	0.6
177	3.43		2.97	−8.011	−0.078	0.1	0.1
No.	Series						
			Only old	Only new	test.1	test.2	test.3
1	Old	1					
2	Old	1					
3	Old	1					
4	Old						
5	Old	1					
6	Old	1					
7	Old	1					
8	Old	1					
9	Old						
10	Old	1					
11	Old						
12	Old						

Table 1 (Continued)

No.	Series					
		Only old	Only new	test_1	test_2	test_3
13	Old	1				
14	Old					
15	Old	1				
16	Old					
17	Old	1				
18	Old	1				
19	Old					
20	Old					
21	Old					
22	Old	1				
23	Old					
24	Old					
25	Old	1				
26	Old	1				
27	Old					
28	Old					
29	Old					
30	Old					
31	Old	1				
32	Old					
33	Old					
34	Old					
35	Old					
36	Old					
37	Old	1				
38	Old					
39	Old					
40	Old					
41	Old					
42	Old					
43	Old					
44	Old					
45	Old					
46	Old					
47	Old					
48	Old					
49	Old	1				
50	Old	1				
51	Old	1				
52	Old	1				
53	Old	1				
54	Old	1				
55	Old					
56	Old					
57	Old					
58	Old	1				
59	Old					
60	Old					
61	Old					
62	Old	1				
63	Old	1				
64	Old					
65	Old	1				
66	Old					
67	Old	1				
68	Old	1				
69	Old					
70	Old	1				
71	Old	1				
72	Old	1				
73	Old	1				
74	Old	1				
75	Old	1				
76	Old					
77	Old					
78	Old					
79	Old					
80	Old	1				
81	Old					
82	Old					
83	Old					
84	Old					
85	Old					
86	Old					
87	Old	1				

Table 1 (Continued)

No.	Series					
		Only old	Only new	test.1	test.2	test.3
88	Old					
89	Old					
90	Old	1				
91	New					1
92	New		1			
93	New			1		
94	New					
95	New				1	
96	New		1			1
97	New				1	
98	New					1
99	New					
100	New					
101	New					
102	New					
103	New					1
104	New				1	
105	New					
106	New					
107	New					
108	New					
109	New		1		1	
110	New					
111	New					1
112	New		1			
113	New					1
114	New		1			
115	New					
116	New		1			1
117	New			1		
118	New				1	
119	New				1	
120	New					
121	New			1		
122	New					
123	New		1			
124	New					1
125	New					
126	New					1
127	New			1		
128	New					
129	New		1			
130	New		1			
131	New		1			
132	New		1			
133	New					
134	New					
135	New		1			
136	New		1			1
137	New		1			1
138	New					1
139	New				1	
140	New					1
141	New			1		
142	New		1			
143	New		1	1		
144	New					
145	New		1			
146	New		1			
147	New		1		1	
148	New		1			
149	New		1			
150	New		1			
151	New					1
152	New		1		1	
153	New		1			
154	New		1			
155	New					1
156	New		1			
157	New		1		1	
158	New		1	1		
159	New				1	
160	New					
161	New				1	
162	New					

Table 1 (Continued)

No.	Series					
		Only old	Only new	test.1	test.2	test.3
163	New			1		
164	New					
165	New				1	
166	New				1	
167	New					
168	New					
169	New		1			
170	New					
171	New					
172	New		1	1		
173	New		1			
174	New		1			
175	New		1			
176	New		1			1
177	New		1			

For the definition of the various parameters, see Section 2. *Ring*: An, aniline; Ant, anthracene; Bi, biphenyl; F, fluorene; N, naphthalene; 5, 5-membered ring. *Bridge*, bridge between the rings in biphenyls if such a bridge occurs. *AnX*, ring substituent with an aniline moiety as core according to the conventions presented in Figs. 1–4; *R*, substituents at the nitrogen of the functional amino group.

Fig. 3. Treatment of substituents on aminofluorenes.

177 not included as these compounds behave as outliers):

$$\begin{aligned} \text{BRM} = & 1.01(\pm 0.33) \log P \times I(\text{monoNH}_2) + 0.45(\pm 0.27) \log P \\ & \times I(\text{diNH}_2) + 1.74(\pm 0.53) \text{EHOMO} - 0.42(\pm 0.36) \text{ELUMO} \\ & - 0.56(\pm 0.19) \sum \text{MR}_{2,6} + 0.89(\pm 0.84) I(\text{diNH}_2) + 16.13(\pm 4.71) \\ n = 52 \quad r = 0.889 \quad r^2 = 0.791 \quad s = 0.497 \quad F = 28.4 \quad q^2 = 0.724 \end{aligned} \quad (1)$$

The statistical fit and the stability of Eq. (1) in cross-validation are quite acceptable for the type of biological data considered. Fitted values from Eq. (1) are included into Table 1, and a plot of fitted versus observed activities is shown in Fig. 5.

Most important for carcinogenic potency is hydrophobicity. The coefficients of the log *P*-terms in Eq. (1) show that the influence of hydrophobicity is stronger in compounds with one amino group as compared to di-amines. These terms together with the orbital energies and the $\sum \text{MR}_{2,6}$ term reflect the key factors for the gradation of carcinogenic potency in mice. Potency increases with increasing hydrophobicity, increasing values of the energy of the highest occupied molecular orbital and decreasing values of the lowest empty molecular orbital. Bulk in the positions adjacent to the functional

Fig. 4. Treatment of substituents on anthracenes.

Fig. 5. Observed carcinogenic potency versus fitted values from Eq. (1).

amino group is unfavorable and leads to a decrease of potency. The $I(\text{diNH}_2)$ -term in Eq. (1) is of only marginal importance indicating that, other things being equal, compounds with more than one amino group tend to be intrinsically somewhat more active than mono-amines.

A test set of 9 active compounds was stochastically selected from series.new (test.1, indicated in Table 1). After elimination of these compounds not only remains Eq. (1) stable ($r^2 = 0.769$, $r = 0.877$, $q^2 = 0.691$) but also provides a very good prediction of the test set presented in Table 2 with $r_{\text{pred}}^2 = 0.874$ ($r = 0.935$).

Eq. (1) fully agrees with the equation obtained for the series.old in [4] which reads:

$$\begin{aligned} \text{BRM} = & 0.88(\pm 0.27) \log P + I(\text{monoNH}_2) + 0.29(\pm 0.20) \log P \\ & \times I(\text{diNH}_2) + 1.38(\pm 0.76) \text{EHOMO} - 1.28(\pm 0.54) \text{ELUMO} \\ & - 1.06(\pm 0.34) \sum \text{MR}_{2,6} - 1.10(\pm 0.80) \text{MR}_3 \\ & - 0.20(\pm 0.16) E_S(R) + 0.75(\pm 0.75) I(\text{diNH}_2) + 11.16(\pm 6.68) \\ n = 37 \quad r = 0.907 \quad r^2 = 0.823 \quad s = 0.381 \quad F = 16.3 \quad P = 100\% \end{aligned} \quad (2)$$

The only differences between Eqs. (1) and (2) are that the terms with MR_3 and $E_S(R)$ occurring in Eq. (2) do not occur in Eq. (1). The MR_3 term in Eq. (2) is of only marginal importance, and an $E_S(R)$ cannot be expected in Eq. (1) as only one of the 56 active series.new compounds bears a substituent at the functional amino group.

There are 12 compounds in series.old which do not occur in series.new and thus provide a real test set. Table 3 summarizes the carcinogenic activity of these compounds predicted from Eq. (1). With $r_{\text{pred}}^2 = 0.69$ ($r = 0.83$) Eq. (1) shows a very satisfactory predictive power considering the fact that the 12 compounds considered have the property of being randomly selected.

Of series.new, 19 compounds are not part of series.old. Their carcinogenic activity should be predictable from Eq. (2). However, the structural diversity within the active compounds of series.old is smaller than that in series.new. For example, anthraquinones, anilines substituted in all positions, or heterocyclic compounds

Table 2
Carcinogenic activity of the compounds in test set test.1 predicted from Eq. (1).

No	Observed	Predicted	Residual
93	2.77	2.11	0.66
117	3.15	3.35	−0.20
121	3.13	3.74	−0.61
127	2.18	2.06	0.06
141	1.81	1.19	0.62
143	3.20	3.62	−0.42
158	4.20	4.58	−0.38
163	1.57	1.29	0.28
172	1.11	1.38	−0.27
		$r_{\text{pred}}^2 = 0.87$	$r_{\text{pred}} = 0.94$

No = number in Table 1.

Table 3

Carcinogenic activity of those compounds in series.old not occurring in series.new predicted from Eq. (1).

No	Observed	Predicted	Residual
5	3.97	3.65	0.32
6	4.47	4.02	0.45
10	3.63	3.77	−0.14
13	3.47	2.69	0.78
37	2.11	2.48	−0.37
51	3.50	3.52	−0.02
52	2.99	3.57	−0.58
53	1.99	2.02	0.03
54	2.66	2.15	0.51
58	3.01	3.06	−0.05
63	2.94	3.12	−0.18
73	1.92	2.61	−0.69
		$r_{\text{pred}}^2 = 0.69$	$r_{\text{pred}} = 0.83$

No = number in Table 1.

Bold indicates compounds with largest deviations from predicted BRM.

do not occur in series.old but are present in series.new. In other words, the applicability domain of Eq. (2) referring to the old compounds will not cover all series.new compounds. In a first step, a prediction was made for all relevant series 2 compounds except #146 which is a known outlier in series 2 (see above), #175 which has two amino groups attached to a triazine ring, and #109 which is an anthraquinone. Even after eliminating these compounds the prediction still is only poor with $r_{\text{pred}}^2 = 0.29$ ($r = 0.62$). This is almost entirely due to compounds #132 and 176 which both have a NO_2 group in para-position to the functional amino group. A strong cross-conjugation is to be expected here which does not occur in any of the series.old compounds and is, thus, not reflected by Eq. (2). The impact of cross-conjugation can be accounted for by introducing an indicator variable, I_{NO_2} , in the correlation between observed and predicted values which leads to a pronounced improvement with $r_{\text{pred}}^2 = 0.73$ ($r = 0.85$). The predicted values after correction with I_{NO_2} are summarized in Table 4. Thus, as Eq. (1) for series.old, Eq. (2) shows a very satisfactory predictive power for the compounds of series.new if the limits of its applicability domain are taken into account. There is only one compound whose prediction deviates more than one unit from the observed value (#177). It should be pointed out that, again, the compounds for which potencies are predicted are not a designed test set but rather have the property of being randomly selected.

Table 4

Carcinogenic activity of those compounds in series.new not occurring in series.old predicted from Eq. (2). The prediction was performed after correction for cross-conjugation in compounds #132 and 176 (see text).

No	Observed	Predicted	Residual
112	3.13	3.71	−0.58
116	2.11	3.00	−0.89
123	1.99	2.28	−0.29
132	1.66	1.43	0.23
135	3.48	2.91	0.57
143	3.62	3.75	−0.13
145	4.06	3.92	0.14
147	4.64	4.04	0.60
149	3.95	3.55	0.40
152	3.72	3.80	−0.08
154	4.26	4.04	0.21
157	4.56	3.90	0.66
158	4.58	3.91	0.67
173	3.57	3.67	−0.10
176	1.32	1.55	−0.23
177	3.43	4.60	−1.17
		$r_{\text{pred}}^2 = 0.73$	$r_{\text{pred}} = 0.85$

No = number in Table 1.

Table 5

Values of BRM predicted from Eq. (1) for the inactive compounds in series_new.

No	BRM predicted from Eq. (1)
96	0.38
99	2.46
100	2.57
101	1.71
102	2.95
103	2.13
104	1.05
105	2.57
106	3.10
107	4.09
108	3.15
113	1.86
115	2.22
120	3.31
124	2.89
126	2.15
131	2.82
136	4.02
137	1.94
144	1.62

Table 5 summarizes BRM values of non-carcinogenic aromatic amines from series_new predicted from Eq. (1). All compounds appear as weak carcinogens. Thus, Eq. (1) allows to recognize strong carcinogens and to estimate the gradation of potency within active compounds but cannot separate weak carcinogens from inactive compounds. The same is true for Eq. (2) as shown in [9,10]. This is not an uncommon situation as the properties connected with the gradation of potency need not be identical with those discriminating between active and inactive compounds. What is satisfactory, however, is that the non-carcinogens are placed in the region of very low to low potency with the only exception of compounds #107 and 136 which show high values for BRM. The situation is much better than in the case of mutagenic potency of aromatic amines where the predicted potency for most of the non-mutagenic amines spans the entire range of values up to very high potency [14]. In order to separate active from inactive compounds classification methods have to be applied.

3.2. QSARs for identifying potent carcinogens

This analysis is aimed at separating potent carcinogens from inactive and weakly positive aromatic amines. To separate carcinogenic from non-carcinogenic aromatic amines – independently from the degree of carcinogenic potency – discriminant analysis was already successfully applied to the compounds of the old series [10]. In the present analysis, compounds were divided into two classes after omitting the two severe outliers #146 and 168 (see above) according to

class 1 (active carcinogens): $BRM > 2.5$

class 2 (inactive compounds and very weak carcinogens): $BRM < 2.5$

The following simple discriminant function affords a statistically highly significant separation of the two classes:

$$w = -1.20 \log P + 0.96 MR_2 - 1.17 EHOMO \quad (3)$$

$$w_{(\text{mean, class1})} = -1.08 \quad n_1 = 33$$

$$w_{(\text{mean, class2})} = 0.87 \quad n_2 = 41$$

Eq. (3) correctly reclassifies 86.5% of the compounds (class 1: 81.8%, misclassified compounds #112, 125, 128, 134, 160 and 161; class 2: 90.2%, misclassified compounds: #107, 120, 136 and 167)

Table 6

Class membership of those compounds in series_old not occurring in series_new predicted from Eq. (3).

No	Observed		Class membership predicted from (3)
	BRM	Class	
5	3.97	1	1
6	4.47	1	1
10	3.63	1	1
13	3.47	1	2
37	2.11	2	2
51	3.50	1	1
52	2.99	1	2
53	1.99	2	2
54	2.66	1	2
58	3.01	1	1
63	2.94	1	2
73	1.92	2	2

No = number in Table 1.

Bold indicates compounds incorrectly predicted.

with a very good result in cross-validation identical with the simple reclassification (all compounds: 86.5%; class 1: 81.8%; class 2: 90.2%). Taking into account the signs of the class means of the discriminant variable w it follows from Eq. (3) that the chance of a compound to be an active carcinogen increases with hydrophobicity and EHOMO and decreases with the size of substituents in position 2; most important is the hydrophobic effect. This is in full agreement with Eq. (1). However, not all of the variables appearing in Eq. (1) also occur in Eq. (3); in addition the hydrophobic effect need no longer to be differentiated between mono- and di-amines. These differences indicate that for the decision active/inactive on the one and for the gradation of potency within the active compounds on the other hand in fact different (but overlapping) combinations of properties are important.

The discriminant function (Eq. (3)) cannot directly be compared with the results of discriminant analysis in [10] since the classification used there is based on a different type of biological measurement (positives *versus* negatives in the previous paper; potent positives *versus* negatives and weak carcinogens in this paper). In addition, structural variations are different with, for example, many N-substituted compounds in the inactive class of series_old while only one such compound exists in the new series. However, the compounds summarized in Table 3 can again be used as a test set with the result presented in Table 6. The class membership of 8 out of 12 compounds (66.7%) is correctly predicted with 5 out of 9 (55.5%) for the active class and 3 out of 3 (100%) for the inactive class. Even when considering the stochastic nature of the test set this is not a very convincing result. However, three of the four wrongly assigned active compounds (#52, 54, and 63) are close to inactivity with BRM values less than 0.5 BRM units above the class border of $BRM = 2.5$, and the same compounds also show the largest deviations of predicted BRM in Table 3. It thus might be that the activity of these compounds shows a relatively large error of measurement. In order to further check the predictive capability of Eq. (3) two further test sets stochastically selected from series_new were used (test_2 and test_3). The compounds belonging to these series are indicated in Table 1.

When eliminating the test sets the discriminant function remains completely stable with correct reclassification rates of 85.0% and 86.2% (same values in cross-validation) for test_2 and test_3, respectively. The classification by Eq. (3) is summarized in Tables 7 and 8. For stochastically selected test sets the results are extremely good. In test_2, 13 out of 14 compounds (92.9%) are correctly predicted (class 1: 6 out of 7 compounds (85.7%) correct; class 2: 7 out of 7 (100%) compounds correct). 14 out of 16 compounds (87.5%) are correctly predicted for test_3 with 7 out of 7 correct (100%) for class 1 and 7 out of 9 correct (77.8%) for class 2.

Table 7

Class membership of the compounds in test.2 predicted from Eq. (3).

No	Observed		Class membership predicted from (3)
	BRM	Class	
95	2.34	2	2
97	3.79	1	1
104	1.26	2	2
109	2.27	2	2
118	2.06	2	2
119	2.03	2	2
139	4.34	1	1
147	4.64	1	1
152	3.72	1	1
157	4.56	1	1
159	2.17	2	2
161	3.32	1	2
165	1.19	2	2
166	2.88	1	1

No = number in Table 1.

Table 8

Class membership of the compounds in test.3 predicted from Eq. (3).

No	Observed		Class membership predicted from (3)
	BRM	Class	
91	3.59	1	1
96	1.10	2	2
98	3.78	1	1
103	1.19	2	2
111	3.74	1	1
113	1.19	2	2
116	2.11	2	2
124	1.14	2	1
126	1.39	2	2
136	1.32	2	1
137	1.43	2	2
138	3.81	1	1
140	4.91	1	1
151	5.22	1	1
155	2.80	1	1
176	1.32	2	2

No = number in Table 1.

These results indicate a high predictive power of the discriminant function Eq. (3).

4. Discussion and conclusions

The results are fully compatible with and confirm our earlier QSAR results for aromatic amines derived from a smaller sample [9,10]. Significant Hansch equations and discriminant functions stable in cross-validation are obtained. The present results emphasize the validity of performing comparative QSAR analyses in well defined series of carcinogenic compounds.

The key factor in aromatic amines carcinogenesis is an increase of potency with hydrophobicity. The influence of hydrophobicity on the gradation of carcinogenic potency within active compounds is stronger for compounds with one amino group in comparison with compounds with more than one amino group. This effect has explicitly to be taken into account by allowing for different regression coefficients of $\log P$ for mono- and di-amines in Eqs. (1) and (2). For the sole separation of active from inactive compounds this is not necessary [10], whereas a simple $\log P$ term in the present discriminant functions is sufficient which affords the largest contribution for separating potent carcinogens from weak and non-carcinogens. In addition to hydrophobicity, electronic factors also play a role. For the gradation of potency for active compounds potency increases with increasing energy of the highest

occupied (EHOMO) and with decreasing energy of the lowest empty molecular orbital (ELUMO). In the case of separation of active from inactive compounds, EHOMO alone is sufficient for class separation. In addition BRM strongly decreases with bulk in the positions adjacent to the functional amino group (positions 2 and 6). For the separation of active from inactive compounds only the bulk in position 2 is important. In contrast to our previous results, an effect of substituents at the nitrogen could not be found in the present case. The reason simply is that there is only one compound with a substituted amino group in the series considered.

In mechanistic terms, hydrophobicity is an essential parameter that rules absorption and transport of chemicals as well as their interactions with receptors, such as those of the metabolism [15,16]. HOMO parameterizes the ability of chemicals to undergo oxidation; as a matter of fact, the main metabolic pathway of the aromatic amines includes oxidation to hydroxylated compounds. The difference between HOMO and LUMO can be seen as a measure of chemical reactivity [17]. Regarding steric effects, the presence of bulky substituents on the ring can hinder the ability of the amine nitrogen to undergo metabolic activation.

The results for the test sets show that the QSARs have high predictive power within their applicability domain. For the prediction of BRM for the test set test.1 from Eq. (3) $r_{\text{pred}}^2 = 0.874$ is obtained and, for compounds from the old series considered in [9,10] not occurring in the present training series $r_{\text{pred}}^2 = 0.730$ is obtained. For the discriminant functions the prediction of class membership is correct for 92.2% of the compounds in test set test.2 and 85.7% on test set test.3.

The present QSARs can be compared with QSARs for the mutagenic potency of aromatic amines [17] because bioactivation through similar mechanisms is involved in both cases (see also [11]). There is good agreement. Debnath et al. [17] also found that hydrophobicity is the main determinant of mutagenic potency: potency increases linearly with $\log P$. It was also found that this type of relationship is only valid for compounds with $\log P > 1$ and does not hold for more hydrophilic analogs. It could very well be that behind this finding a similar effect is hidden as found for BRM: potency of mono-amines shows a much stronger dependence on $\log P$ than for di-amines. The following is at least in keeping with such an explanation: (i) the majority of compounds in [17] for which the QSARs hold consists of mono-amines, (ii) practically all of the “too hydrophilic” compounds are di-amines. For the latter an inverse relationship with $\log P$ was suggested which, however, is not well supported by the data. Electronic effects are of secondary importance: potency increases with EHOMO and decreases with ELUMO. Another QSAR available for the carcinogenic potency of aromatic amines [18] is based on the geometrical structure of compounds and MO indices using neural network modeling. A direct comparison with the present results is not possible because of the black-box nature of the neural network results. The success of this approach was limited as only a weak correlation between descriptors and carcinogenic potency was found.

According to [18], substitution of a chloro group or methyl or methoxy group ortho to the amino group is considered to often enhance potency. This statement needs a comment. The present QSARs show that ortho-substituents can operate through at least three effects: direct steric, electronic, and hydrophobic. It thus depends on the nature of an ortho-substituent as well as on structural motifs in the rest of the molecule whether it will decrease or increase potency (see also [12]). The general picture is that carcinogenic potency decreases with bulk in the ortho-position [19]. This is in keeping with the negative terms with the molar refractivity observed in the Hansch equations and discriminant functions.

In conclusion, predictions of carcinogenic potency can now be made in two steps: (i) simple yes/no prediction of activity with

the help of the discriminant function provided in [10], and (ii) for compounds predicted to be carcinogenic, prediction of the degree of carcinogenic potency from Eq. (1) or Eq. (2). In addition the potent carcinogens, most relevant for risk assessment, can be identified directly through Eq. (3).

Conflict of interest

No conflicts of interest.

Acknowledgements

Work partially granted by:

- Project EU FP6 no. 037017 OSIRIS 'Optimized Strategies for Risk Assessment of Industrial Chemicals through Integration of Non-Test and Test Information'.
- Project EU FP7 no. 200787 OpenTox "An Open Source Predictive Toxicology Framework".

References

- [1] E. Zeiger, Identification of rodent carcinogens and noncarcinogens using genetic toxicity tests: premises, promises, and performance, *Regulat. Pharmacol. Toxicol.* 28 (1998) 85–95.
- [2] R. Benigni, C. Bossa, O. Tcheremenskaia, A. Giuliani, Alternatives to the carcinogenicity bioassay: *in silico* methods, and the *in vitro* and *in vivo* mutagenicity assays, *Exp. Opin. Drug Metab. Toxicol.* 6 (2010) 1–11.
- [3] R. Benigni, T.I. Netzeva, E. Benfenati, C. Bossa, R. Franke, C. Helma, E. Hulzebos, C.A. Marchant, A.M. Richard, Y.T. Woo, C. Yang, The expanding role of predictive toxicology: an update on the (Q)SAR models for mutagens and carcinogens, *J. Environ. Sci. Health. C: Environ. Carcinog. Ecotoxicol. Rev.* 25 (2007) 53–97.
- [4] R. Benigni, Structure-activity relationship studies of chemical mutagens and carcinogens: mechanistic investigations and prediction approaches, *Chem. Rev.* 105 (2005) 1767–1800.
- [5] Y.T. Woo, D.Y. Lai, Aromatic amino and nitro-amino compounds and their halogenated derivatives, in: E. Bingham, B. Cohrssen, C.H. Powell (Eds.), *Patty's Toxicology*, 5th ed., John Wiley and Sons, Inc., New York, 2001, pp. 969–1105.
- [6] F. Pedersen, J. de Bruijn, S.J. Munn, K. Van Leeuwen, Assessment of additional testing needs under REACH, Effects of (Q)SARs, risk based testing and voluntary industry initiatives, JRC report EUR 20863 EN, 2003, Ispra, EUR.
- [7] Anonymous, Animal tests inescapable, *Nature* 453 (2008) 563–564.
- [8] W. Lilienblum, W. Dekant, H. Foth, T. Gebel, J.G. Hengstler, R. Kahl, P.J. Kramer, H. Schweinfurth, K.M. Wollin, Alternative methods to safety studies in experimental animals: role in the risk assessment of chemicals under the new European Chemicals Legislation (REACH), *Arch. Toxicol.* 82 (2008) 211–236.
- [9] R. Benigni, A. Giuliani, R. Franke, A. Gruska, Quantitative structure-activity relationships of mutagenic and carcinogenic aromatic amines, *Chem. Rev.* 100 (2000) 3697–3714.
- [10] R. Franke, A. Gruska, A. Giuliani, R. Benigni, Prediction of rodent carcinogenicity of aromatic amines: a quantitative structure-activity relationships model, *Carcinogenesis* 22 (2001) 1561–1571.
- [11] R. Benigni, C. Bossa, T.I. Netzeva, A. Rodomonte, I. Tsakovska, Mechanistic QSAR of aromatic amines: new models for discriminating between homocyclic mutagens and nonmutagens, and validation of models for carcinogens, *Environ. Mol. Mutagen.* 48 (2007) 754–771.
- [12] R. Benigni, A.P. Worth, T.I. Netzeva, N.G. Jeliakova, C. Bossa, A. Gruska, R. Franke, Structural motifs modulating the carcinogenic risk of aromatic amines, *Environ. Mol. Mutagen.* 50 (2009) 152–161.
- [13] R. Benigni, C. Bossa, A.M. Richard, C. Yang, A novel approach: chemical relational databases, and the role of the ISSCAN database on assessing chemical carcinogenicity, *Ann. Ist. Super. Sanità* 44 (2008) 48–56.
- [14] R. Benigni, C. Andreoli, A. Giuliani, QSAR models for both mutagenic potency and activity: application to nitroarenes and aromatic amines, *Environ. Mol. Mutagen.* 24 (1994) 208–219.
- [15] C. Hansch, A. Leo, *Exploring QSAR. 1. Fundamentals and applications in chemistry and biology*, American Chemical Society, Washington, DC, 1995.
- [16] C. Hansch, L. Zhang, Quantitative structure-activity relationships of cytochrome P-450, *Drug Metab. Rev.* 25 (1993) 1–48.
- [17] A.K. Debnath, G. Debnath, A.J. Shusterman, C. Hansch, A QSAR investigation of the role of hydrophobicity in regulating mutagenicity in the Ames test. 1. Mutagenicity of aromatic and heteroaromatic amines in *Salmonella typhimurium* TA98 and TA100, *Environ. Mol. Mutagen.* 19 (1992) 37–52.
- [18] M. Vracko, A study of structure-carcinogenic potency relationship with artificial neural networks. The using of descriptors related to geometrical and electronic structures, *J. Chem. Inf. Comput. Sci.* 37 (1997) 1037–1043.
- [19] D.Y. Lai, Y.T. Woo, M.F. Argus, J.C. Arcos, Cancer risk reduction through mechanism-based molecular design of chemicals, in: S.C. DeVito, R.L. Garrett (Eds.), *Designing Safer Chemicals. Green Chemistry for Pollution Prevention*, American Chemical Society, Washington, DC, 1996, pp. 1–13.