Environmental and Molecular Mutagenesis 50:152–161 (2009)

Research Article

Structural Motifs Modulating the Carcinogenic Risk of Aromatic Amines

Romualdo Benigni,[1]* Andrew Worth,[2] Tatiana Netzeva,[2,3] Nina Jeliazkova,[4] Cecilia Bossa,[1] Andreas Gruska,[5] and Rainer Franke[5]

[1]*Environment and Health Department, Istituto Superiore di Sanità, Rome, Italy*
[2]*European Commission, Joint Research Centre, Institute for Health and Consumer Protection, Ispra, Italy*
[3]*European Chemicals Agency ECHA, Helsinki, Finland*
[4]*Ideaconsult Ltd., Sofia, Bulgaria*
[5]*Consulting in Drug Design GbR, Basdorf, Germany*

The structure alerts (SA) for carcinogenicity/mutagenicity are a repository of the science on chemical biological interactions; in addition, they have a crucial role in practical applications for risk assessment. In predictive toxicology, it is crucial that knowledge of SAs is accompanied by knowledge of the structural motifs that modulate their effects. Recently, we have compiled an updated list of SAs implemented in the expert system Toxtree 1.50 (open source, freely available). These SAs are aimed at discriminating between active and inactive chemicals, and include only modulating factors with a high probability of eliminating completely the effect of the SA. Here we have examined the factors that modulate carcinogenic potency: this is an additional piece of information that can have a role in fine-tuning a risk assessment. The case study selected is the carcinogenic potential of the aromatic amines in rats and mice. As the carcinogenic potency of these compounds is different in mice and rats (correlation coefficient = 0.546), there are both agreements and differences in the pattern of these motifs. Differences are observed mainly for the motifs that decrease the carcinogenic potency of aromatic amines. In mice, substitutions ortho and meta to the amino group tend to decrease the potency, as well as $-NO_2$ in any position. In rats, these motifs affect the potency to a more limited extent. On the other hand, increasing effects are quite similar in the two animals and are exerted mainly by additional rings, tricyclic systems, five-numbered rings, and N-heteroaromatic systems. Environ. Mol. Mutagen. 50:152–161, 2009.

Key words: structural alert; potency; predictive toxicology; pattern recognition; computer chemistry

INTRODUCTION

The structure–activity relationship (SAR) paradigm provides a wide range of tools that can be exploited in research on the toxicity of chemicals and in the generation of models aimed at predicting toxicity from the chemical structure, in the absence of experimental data. Such predictive models have an expanding role in the regulation of chemical risk [OECD, 2006; Benigni et al., 2007b; Worth et al., 2007].

The SAR approaches have different degrees of approximation/uncertainty and apply to different scopes [Franke, 1984; Kubinyi, 1993; Benigni, 2005]. On one hand, there are the fine-tuned quantitative SAR (QSAR) methods that apply to congeneric classes of chemicals and provide mathematical models based on physical chemical or structural parameters. These methods (with modifications) have been extended also to noncongeneric sets of chemicals; many of the most popular commercial systems belong to the latter category.

On the other side, there are coarse-grain approaches such as the structure alerts (SAs). These are reactive chemical groups, or molecular substructures that have been recognized to be able to elicit toxic effects: the

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

*Correspondence to: Romualdo Benigni, Environment and Health Department, Istituto Superiore di Sanità, Viale Regina Elena 299, 00161 Rome, Italy. E-mail: rbenigni@iss.it

Received 17 July 2008; provisionally accepted 24 September 2008; and in final form 13 November 2008

DOI 10.1002/em.20461

Published online 16 January 2009 in Wiley InterScience (www.interscience.wiley.com).

knowledge of the SAs is often accompanied by knowledge of their modulating factors, i.e., substructures that, when present together with a SA in a molecule, can enhance or inhibit the effect of the SA. In addition to increasing our understanding of the chemical biological interactions underlying chemical toxicity, the SAs have a crucial role in practical applications for risk assessment, including: (a) the description of sets of chemicals; (b) preliminary hazard characterization; (c) formation of categories, for example, for regulatory purposes; (d) generation of subsets of congeneric chemicals to be analyzed subsequently with QSAR methods; and (e) priority setting [Benigni et al., 2008a].

In the field of mutagenicity and carcinogenicity, SAs have a long history and several lists of SAs have been published or implemented into software programs for predictive toxicology [Benigni and Bossa, 2008]. Recently, in the framework of a collaboration between the European Chemicals Bureau and the Istituto Superiore di Sanità we reviewed the various evidence related to SAs for mutagenicity and carcinogenicity; based on this survey, a new compilation of SAs has been generated that compiles previous work in an optimized form for computer implementation. The revised compilation is now included in the open source Toxtree 1.50 software for predicting the toxicity of chemicals (freely available from the European Chemicals Bureau website: http://ecb.jrc.it/qsar/qsar-tools/index.php?c=TOXTREE) [Benigni et al., 2008a].

An important element for the efficiency of SAs in predictive toxicology is the knowledge of modulating factors. In the Toxtree 1.50 rules, the SAs are aimed at discriminating between active and inactive chemicals, and only modulating factors that have a high probability of completely eliminating the effect of SAs have been considered and implemented. In this article, we study the factors that modulate carcinogenic potency; this is an additional piece of information that, after having established the carcinogenic potential of a chemical, can have a role in risk assessment procedures. The case study selected for the present analysis is the aromatic amines. These chemicals have considerable environmental and industrial importance, and toxicological data have been reported for a large number of them [Lai et al., 1996; Sugimura, 1997; Vineis and Pirastu, 1997; Skog et al., 1998; Woo and Lai, 2001]. In previous work we have performed QSAR investigations on the activity and the potency of the aromatic amines, for both mutagenic (Salmonella) and carcinogenic (rat and mouse) endpoints [Benigni et al., 2000, 2007a; Franke et al., 2001]. In the present investigation the effect of structural motifs on the carcinogenic potency of aromatic amines in rats and mice are investigated by using the software program EVAL. EVAL is a topological pattern finder based on substructural descriptors that identify patterns (topological pharmacophores) typical of a biological activity of interest [Streich and Franke, 1988; Hübel and Franke, 1991a,b]. Thus, motifs that may increase or decrease the carcinogenic potency of an aromatic amine can be identified. These motifs can be used to fine-tune conclusions from previously characterized structural alerts.

METHODS

Structures of 107 aromatic amines and their carcinogenic potency in rats and mice were obtained from the ISSCAN database on chemical carcinogens (freely downloadable at www.iss.it/ampp/dati/cont.php?id=233&lang=1&tipo=7) [Benigni et al., 2008b]. Carcinogenic activity is expressed as

$$\text{mice} : \text{BRM} = \log(\text{MW}/\text{TD}_{50})_{\text{mouse}} + 3$$
$$\text{rats} : \text{BRR} = \log(\text{MW}/\text{TD}_{50})_{\text{rat}} + 3$$

(MW = molecular weight) in accordance with previous investigations [Benigni et al., 2000; Franke et al., 2001], where TD_{50} is the daily dose rate required to halve the probability of an experimental animal remaining free of tumors to the end of its standard lifespan. The TD_{50} values are derived from the Carcinogenic Potency Database project (http://potency.berkeley.edu/cpdb.html). The scaling factor ''+3'' is aimed at transforming all the potency values into positive numbers for the subsequent EVAL analysis.

Based on the frequency distribution of BRM and BRR (see Figs. 1 and 2) classes were defined as follows:

Class 1 (active compounds)
Rats: BRR > 3.2 ($TD_{50} \approx 134$)
Mice: BRM > 3.1 ($TD_{50} \approx 164$)
Class 2 (weakly active and inactive compounds)
Rats: BRR < 2.8 ($TD_{50} \approx 338$)
Mice: BRM < 2.9 ($TD_{50} \approx 268$)

Note that a region of uncertainty is allowed for between the class borders. Compounds falling into this region (three for mice and five for rats) are eliminated.* All structures and biological data are summarized in the Supporting Information Appendix.

Nine compounds could not be unambiguously coded for because their structure is too complex (more than six aromatic rings or condensed systems). These compounds could not be included into the analysis.

After the elimination of the compounds that do not possess a measured value for TD_{50} or which either fall into the uncertainty region between the classes or cannot be unambiguously coded, the number of compounds in Class 1 (active compounds) amounts to $n1 = 33$ for mice and $n1 = 43$ for rats, and in Class 2 (weakly active and inactive compounds) to $n2 = 49$ for mice and $n2 = 39$ for rats.

EVAL operates as follows:

1. A library of substructures is defined where the substructures are considered as potential centers of interaction with the biosystem. To this end, a two-dimensional artificial hypermolecule is constructed (see Fig. 3) by superposing all molecules of the series following predefined rules. The substructures are defined as patterns of atoms or atoms in regions, nodes, and edges of the hypermolecule, resulting in the list of descriptors (variables) summarized in Table I.†

2. Each compound is then superimposed over the hypermolecule and described in terms of the presence or absence of the substructures

*The eliminated compounds either bear a mixture of potency-enhancing and potency-decreasing fragments, or possess a unique structure. They do not add information.

†The letters assigned to the rings of the hypermolecule are arbitrary and do not follow a predefined system.

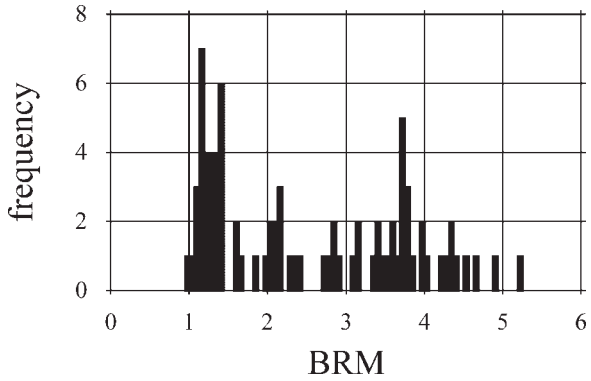

Fig. 1. Distribution of carcinogenic potency values in mice (see details in the text).

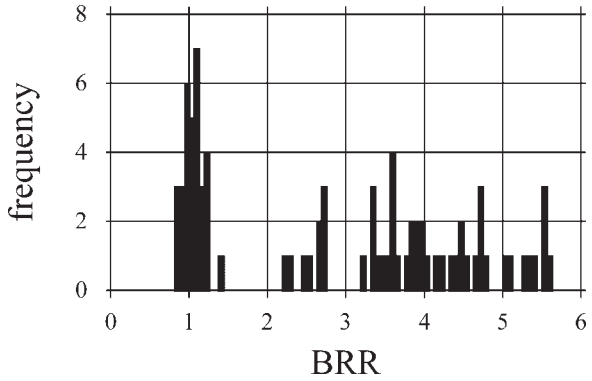

Fig. 2. Distribution of carcinogenic potency values in rats (see details in the text).

described in Point 1 by means of binary descriptor variables x_i which, for the *i*th substructure, take a value of 1 if this structure is present and a value of 0 otherwise. Each compound is then presented as a string containing the variables with a value of $x_i = 1$ which is a complete and unambiguous description of its structure.

3. All superimpositions are done by a strict protocol which follows from the definition of variables:

i. The functional amino group is always in position D_1 (arbitrary selection).
ii. If more than one amino group is present then the amino group in the ortho-position is selected to be the functional amino group.
iii. Six-membered rings in the reference pattern can accommodate five-membered rings.
iv. Other things being equal, molecules are so oriented in the reference pattern such that bulk characterized by molar refractivity is placed in the direction to the left (see Fig. 3) in the following sequence:
    Rings;
    Ortho-position; ortho-substituents in mono-ortho substituted compounds are placed in position C_1;
    Meta-position; meta-substituents in mono-meta substituted compounds are placed in position C_4;
    In the case of ortho, meta-disubstitution, nonpolar groups in the ortho-position such as alkyl are placed in position C_1. With the structures occurring in the present series this has the consequence that polar groups such as NH_2 or NO_2 are then directed towards position E_4.

This protocol is derived by the investigator after careful inspection of all structures. It is completely arbitrary.

4. An example of the whole procedure is presented in Figure 3 with compound 4 (ISS2_1_28) for which the following variables have a value of 1: Variable 1, 2, 6, 84, 11, 61, 125, 123, 130, 131, and 139 (see Table I).

5. The descriptor variables are combined into more complex expressions in a stepwise and interactive procedure using logical operations, mainly ''and,'' ''or'' (inclusive or), and ''not.'' In each step, one more variable is added.

6. Each combination of variables represents a pattern of substructures. For example (for a given compound): '' (x_1 or x_2) and x_3 and (x_4 or x_5)....'' The program identifies all compounds possessing a given substructure pattern.

7. As variables are added in each step, the substructure pattern becomes more complex and is, as a consequence, present in fewer compounds. Thus, as the pattern grows, compounds are eliminated in each step.

8. This process is interactively organized in such a way that the resulting pattern becomes true for as many compounds of the class of interest as possible and, in the ideal case, no compound of the other class. Such patterns are called topological pharmacophores. Different pharmacophores usually exist in parallel depending on, for example, the starting substructure. Different pharmacophores can also reflect different mechanisms of action.

9. A pharmacophore represents a rule of the general form: if substructure A and substructure B and substructures C or D, but not substructure E, appear in a compound, then this compound is to be assigned to the class for which this pharmacophore was developed.

10. The presence/absence of such pharmacophores in a compound allows estimates of its biological behavior.

In the present case the pharmacophores that were searched for are typical for Class 1 (contain motifs that increase carcinogenicity) or Class 2 (contain motifs that decrease or eliminate carcinogenicity) compounds. When presenting the results, reference is always made to the positions in the hypermolecule. As previously stated, nine compounds did not fit into the protocol of the superposition procedure and had to be eliminated.

For the sake of simplicity the variables occurring in the conjunctions are not given as numbers but translated into the corresponding substructures of the hypermolecule. Together with each conjunction, the number of compounds in Class 1 ($n1$) and Class 2 ($n2$) for which the respective conjunction is true are presented. Per definition, ring D and the functional amino group in position D_1 of the reference pattern are present in all compounds so that the corresponding variables need not occur in the conjunctions.

It should be emphasized that EVAL is a very flexible software program; thus it allows the human expert to contribute with her/his expertise to the derivation of chemically meaningful results. The stepwise interactive process on which EVAL is based together with a strategy to use several cuts through the data (e.g., by using different substructures as starting points for the development of conjunctions) minimizes the danger of artificial results even if the number of descriptors is high as compared with the number of observations. Disadvantages are that the definition of meaningful substructures depends very much on the skill and experience of the investigator. In addition, EVAL requires that the dynamic and three-dimensional events at a biological target reflect themselves sufficiently well in a two-dimensional description, at least as far as the topological pharmacophore level is concerned. Finally, features not adequately represented in the training series of compounds cannot be found.

RESULTS AND DISCUSSION

As detailed in the previous section, the software program EVAL first identifies the substructures present in the set of molecules under study. Conceptually, the substructures correspond to the variables of a fragment-based QSAR analysis. In a second phase, combinations of substructures are selected in an interactive way. These combinations, or patterns, are called conjunctions. For each

TABLE I. EVAL Descriptors

Variable	Substructure	Variable	Substructure	Variable	Substructure	Variable	Substructure
1	Dummy	42	C_3 N		**G**	123	L_3 Cl
2	D_1 NH_2	43	C_4 N	84	G_1 CH or C	124	L_3 Me
	Rings	44	C_4 NH	85	G_1 N		**M**
3	Ring A	45	C_4 NH_2	86	G_1 O	125	M_1 NH_2
4	Ring B	46	C_4 NO_2	87	G_1 S		**O**
5	Ring C	47	C_4 Cl	88	G_1 NH	126	O_1 NH
6	Ring D	48	C_4 NO_2	89	G_1 NH_2		**P**
7	Ring E	49	C_4 NHCOMe	90	G_1 OMe	127	P_4 NH
8	Ring F	50	C_4 Me	91	G_1 O	128	P_1 OMe
9	Ring G	51	C_4 F	92	G_1 OH		**General**
10	Ring H	52	C_4 SO_3H	93	G_1 Cl	129	Not H in C_1
11	Ring I	53	C_4 COOH	94	G_1 Br	130	Not H in E_1
12	Ring J	54	C_4 Et	95	G_1 Me	131	Not H in (C_1 oder E_1)
13	Ring K		**D**	96	G_1 NO_2	132	Not H in (C_1 und E_1)
14	Ring L	55	Not defined	97	G_1 F	133	Not H in C_4
15	Ring M	56	D_2 N	98	G_1 NHC_2H_4OH	134	Not H in E_4
16	Ring N	57	D_3 N	99	G_2 CH	135	Not H in (C_4 oder E_4)
17	Ring O	58	D_4 N	100	G_2 N	136	Not H in (C_4 und E_4)
18	Only A		**E**	101	G_2 NH	137	NO_2 present
19	A 5-Ring	59	E_1 Me	102	G_2 O	138	2 NO_2 present
20	D 5-Ring	60	E_1 OMe	103	G_2 NH_2	139	2 NH_2 present
21	G 5-Ring	61	E_1 Cl		**H**		**Additional**
22	H 5-Ring	62	E_1 Br	104	H_1 OH	140	>2 NH_2 present
23	J 5-Ring	63	E_1 C=O	105	H_1 Et	141	C_1 OEt
	A	64	E_1 NH_2	106	H_2 OH	142	E_4 NHCOMe
24	A_2 O	65	E_1 OEt	107	H_2 O	143	C_1 O
25	A_6 NH_2	66	E_1 NO_2	108	H_2 N	144	I_3 Me
	B	67	E_1 OH	109	H_2 Et	145	M_1 Cl
26	B_1 OH	68	E_1 F		**I**	146	Q_3 NH_2
27	B_2 N	69	E_2 OH	110	I_1 N or NH	147	Not defined
28	B_3 O	70	E_4 F	111	I_2 Cl	148	G_1 SO_2
29	B_5 NO_2	71	E_4 N	112	I_2 NH_2	149	R_2 N(O)OH
	C	72	E_4 O	113	I_2 OMe	150	D_2 S
30	C_1 Me	73	E_4 NH_2	114	I_3 SO_3H	151	D_2 O
31	C_1 OMe	74	E_4 NH	115	I_3 Cl	152	F 5_ring
32	C_1 OH	75	E_4 NO_2	116	I_3 OMe	153	B_6 N(O)OH
33	C_1 Cl	76	E_4 Cl		**J**	154	J_1 S
34	C_1 F	77	E_4 Me	117	J_1 N	155	I_2 $NHNH_2$
35	C_1 N	78	E_4 CO		**K**	156	F_2 N
36	C_1 NH_2	79	E_4 COOH	118	K_1 NH_2	157	H_2 Me
37	C_1 NO_2	80	E_4 NO_2	119	K_3 CH_3	158	J_2 Me
38	C_1 COOH		**F**		**L**	159	E 5-ring
39	C_2 N	81	F_1 O	120	L_1 NH_2	160	E_1 N
40	C_2 S	82	F_1 NH_2	121	L_1 F	161	G_1 CH_2OEt
41	Not defined	83	F_3 NH_2	122	L_1 Cl		

For the identification of rings and positions, see Figure 1.

conjunction, the number of active ($n1$) and inactive or weakly active ($n2$) compounds is determined: based on the relative proportion of $n1$ to $n2$, the influence exerted by the substructure patterns (conjunctions) on the carcinogenic potency is assessed.

Results for Mouse Carcinogenicity

Motifs That Decrease the Carcinogenic Potency

Substitution in the ortho-position frequently, but not always, decreases or eliminates carcinogenic potency. According to the superposition protocol, the ortho-position corresponds to position C_1 in the hypermolecule (Fig. 3). A one-variable conjunction (1) indicating ortho-substitution already eliminates 20 of the 33 active compounds, but it is still true for 32 out of the 49 Class 2 analogs:

$$\begin{matrix}\text{Not H in position C\_1} \\ n1 = 13, \quad n2 = 32\end{matrix} \tag{1}$$

Conjunction (2) shows ortho-substituents that almost always lead to very low carcinogenic potency or to inac-

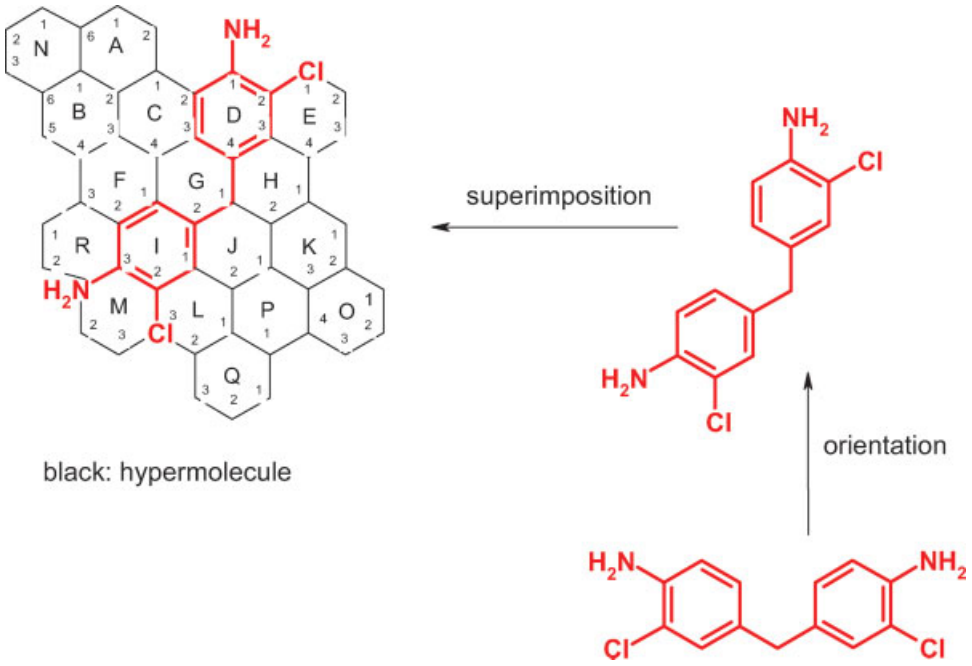

Fig. 3. Reference pattern onto which the individual chemicals are mapped to be processed by EVAL (see details in the text). Superposition of compound ISS1_1_28 is shown as an example.

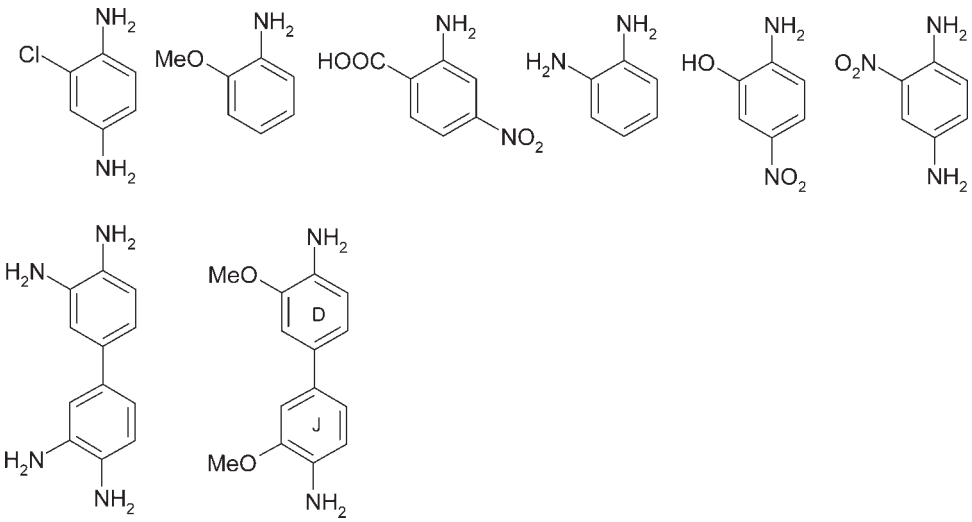

Fig. 4. Examples of compounds in Class 2 (mouse) according to conjunction (2).

tivity, with 21 inactive and only one active compound covered; the most frequently occurring substituents of this type are o-OMe and o-Cl (see also Fig. 4):

$$\begin{gathered}(\text{OMe or Cl or OH or } NH_2 \text{ or } NO_2 \text{ or COOH}) \\ \text{present in C\_1} \\ n1 = 2, \quad n2 = 21 \end{gathered} \tag{2}$$

The substituent Me in the ortho-position is compatible with both high potency as well as low or no activity depending on the coexisting substructures; o-Me is not a typical potency decreasing motif. The substituents F or N (as ring atom) can increase the carcinogenic risk. Thus, in agreement with results of previous discriminant and Hansch analysis [Benigni et al., 2000; Franke et al., 2001], small ortho substituents may increase carcinogenic potency, while larger substituents tend to have the opposite effect.

Conjunction 3 shows that the occurrence of the NO_2 group in aromatic amines diminishes or destroys carcino-

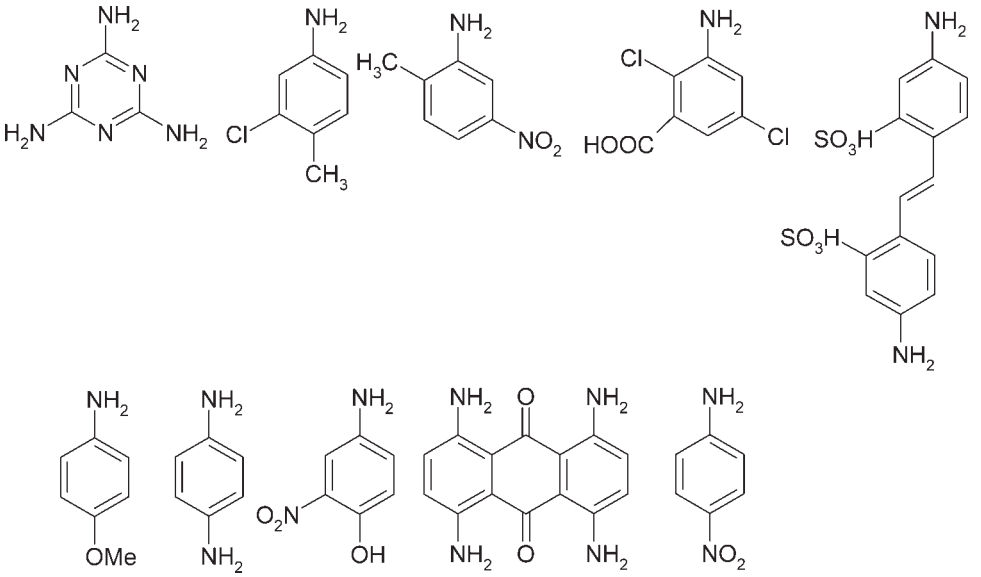

Fig. 5. Examples of compounds in Class 2 (mouse) according to conjunction (5) (first row) and conjunction (6) (second row).

genic potency, unless this effect is counterbalanced by the presence of a strongly activity enhancing substructure such as ring F (see below); the negation ''not ring F'' eliminates two Class 1 compounds:

$$\begin{gathered}\text{NO}_2 \text{ present in any position and not ring F} \\ n1 = 0, \quad n2 = 12\end{gathered} \tag{3}$$

Conjunction 4 shows that substituents in positions C_4 and E_4 of the reference pattern, which correspond to the meta-positions in anilines, are compatible with both high and low potency:

$$\begin{gathered}\text{Not H in C\_4 or E\_4} \\ n1 = 10, \quad n2 = 20\end{gathered} \tag{4}$$

Although conjunction (4) describes a group of compounds in which the inactive ones dominate, there are still 10 active analogs present. Substituents in the meta-positions with the potential to decrease the carcinogenic activity are summarized in conjunction (5) (see Fig. 5; acidic groups: COOH or SO_3H):

$$\begin{gathered}(\text{NH}_2 \text{ in C\_4 or E\_4}) \text{ or } (\text{Cl in C\_4 or E\_4}) \text{ or } (\text{NO}_2 \text{ in} \\ \text{C\_4 or E\_4}) \text{ or } (\text{acidic group in C\_4 or E\_4}) \text{ present} \\ n1 = 3, \quad n2 = 19\end{gathered} \tag{5}$$

In the node G_1 which corresponds to the para-position of anilines a number of substituents (not: second phenyl ring) decrease activity according to conjunction (6) (see Fig. 5):

$$\begin{gathered}(\text{NH}_2 \text{ or OMe or OH or NO}_2 \text{ or NHC}_2\text{H}_4\text{OH in G\_1}) \\ \text{and not Me in E\_4} \\ n1 = 0, \quad n2 = 15\end{gathered} \tag{6}$$

Note that the features G_1 OH and G_1 NHC_2H_4OH, respectively, are supported by only one observation. The other substructures can be regarded as activity decreasing motifs.

With 43 out of 49 Class 2 compounds, the above conjunctions describe the majority of the Class 2 compounds and, thus, represent the most typical structural motifs that render aromatic amines weakly active or inactive in the structural space spanned by the training series. It must be noted, however, that these motifs partially overlap and occur together. With EVAL it is not possible, in such cases, to decide which of two motifs is the more important or whether both are needed to decrease the activity.

Motifs That Increase the Carcinogenic Potency

In keeping with results of previous Hansch analyses [Benigni et al., 2000] emphasizing the role of hydrophobicity for the carcinogenic potency of aromatic amines, additional rings frequently show an activity supporting effect. In a first step the effect of an additional ring in compounds of the biphenyl type with the second ring in the para-position of ring D, which corresponds to ring J in the hypermolecule, were investigated:

$$\begin{gathered}\text{Ring J and not (OMe or NH2 in C\_1) and not Et in C\_4} \\ n1 = 8, \quad n2 = 0\end{gathered} \tag{7}$$

Obviously, ring J is a potency-enhancing feature. Its effect can be destroyed by the occurrence of potency

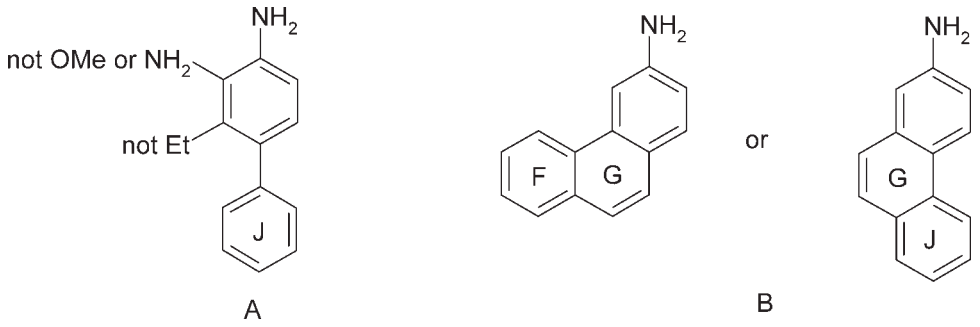

Fig. 6. Structures according to conjunctions (7) (**A**) and (8) (**B**): Active compounds in mice. Lettering corresponds to the hypermolecule.

decreasing motifs such as OMe or NH_2 in C_1 (see Point 3i in the description of EVAL in the ''Methods'' section) or Et in C_4, which are eliminated in conjunction (7) by the corresponding negations (see Fig. 6).

Another typical feature for high potency is a tricyclic ring system always containing ring G plus the rings F or J [conjunction (8); see Fig. 6]:

$$\begin{gathered}\text{Ring G and (ring F or ring J) present} \\ n1 = 7, \quad n2 = 0\end{gathered} \tag{8}$$

The presence of five-membered rings in either ring D or ring G seems to be typical of active structures as follows from conjunction (9) (negation of the activity destroying NO_2 in C_4 group eliminates one inactive compound):

$$\begin{gathered}\text{Ring D or ring G } 5-\text{membered and not } NO_2 \text{ in C\_4} \\ n1 = 11, \quad n2 = 0\end{gathered} \tag{9}$$

N in ring D replacing C in the ortho- or meta-positions of ring D (but not in the para-position) is also an activity increasing motif (compounds with the activity decreasing features NH_2, NO_2, or Et in C_4 eliminated):

$$\begin{gathered}\text{N present in ortho- or meta-positions of ring D} \\ \text{but not (NH2 or NO2 or Et) in C\_4} \\ n1 = 11, \quad n2 = 0\end{gathered} \tag{10}$$

The group of compounds described by conjunction (10) contains a subgroup of nine compounds which also possess the activity promoting features ''ring D or ring G five-membered ring'' described by conjunction (9).

Obviously, there is a large overlap of the groups of compounds described by the above conjunctions. Figure 7 summarizes some compounds that fit both conjunction (9) and (10). Nine active compounds have both a five-membered ring and nitrogen in ring D (not in the para-position), two active compounds have a five-membered ring but not N in ring D, and two active compounds have N in ring D but no five-membered ring. Therefore, it is difficult to decide which feature has more weight for the increase of carcinogenic potency: N in ring D or the presence of a five-membered ring in D or G. Also, the possible effect of coexisting features is not quite clear. With the information available from the training series the following conclusions seem to be reasonable from a practical point of view: the presence of a five-membered ring in D or G together with N in ring D (not in para-position D_4) is a strong indication of an increase in carcinogenic potency, and the presence of a five-membered ring in D or G alone without N in ring D or no five-membered ring in D or G with N in ring D (not in the para-position) still make an increase in carcinogenic potency very likely.

Results for Rat Carcinogenicity

Motifs Which Decrease the Carcinogenic Potency

Conjunction (11) shows that ortho-substitution does not decrease rat carcinogenicity as frequently as seen with mouse carcinogenicity [compare with conjunction (1)]:

$$\begin{gathered}\text{Not H in C\_1} \\ n1 = 19, \quad n2 = 24\end{gathered} \tag{11}$$

As shown in conjunction (12), if compounds possessing an additional ring (activity enhancing feature, see below) are eliminated, only seven Class 1 analogs remain:

$$\begin{gathered}\text{not H in C\_1 and no additional ring} \\ n1 = 7, \qquad n2 = 23\end{gathered} \tag{12}$$

A closer inspection of the compounds involved and the development of additional conjunctions leads to the conclusion that o-Me, o-OH, o-Cl, and o-NH_2 have the potential to decrease carcinogenic activity in rats.

Conjunction (13) represents a large group of mostly inactive anilines substituted in the meta-positions (positions C_4 and E_4 in the hypermolecule). Two active compounds are eliminated by negating the presence of the substituents o-OMe and o-F, which seem to increase potency (see below; compounds of this type have already been presented in Fig. 5):

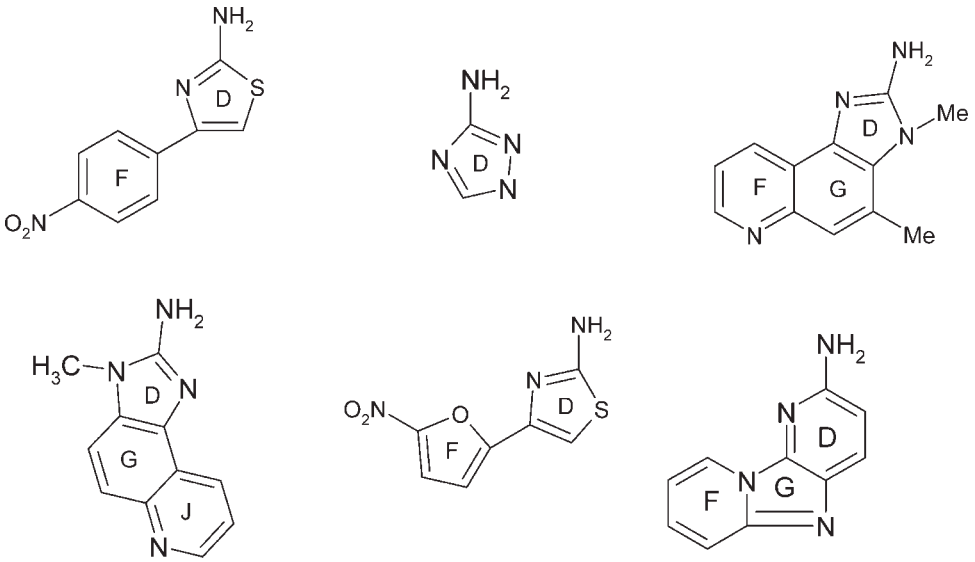

Fig. 7. Examples of compounds that fit both conjunctions (9) and (10) (active in mice). Lettering corresponds to the hypermolecule.

$$\begin{gathered}(NH_2 \text{ in C\_4 or E\_4}) \text{ or } (\text{Cl in C\_4 or E\_4}) \text{ or } (NO_2 \text{ in} \\ (\text{C\_4 or E\_4}) \text{ or } (\text{COOH in C\_4 or E\_4}) \text{ or } (\text{Me or Et in} \\ \text{C\_4}) \text{ or } (SO_3H \text{ in C\_4}) \text{ but not } (\text{OMe or F in C\_1}) \\ n1 = 2, \quad n2 = 14 \end{gathered} \tag{13}$$

Thus, NH_2, Cl, NO_2, and COOH in both C_4 and E_4, and Me, Et, and SO_3H in C_4 are motifs that decrease potency.

Decreased potency also tends to occur with certain small substituents in position G_1 of the hypermolecule under the condition that only ring D is present (all compounds are para-substituted anilines), as is shown by conjunction (14) (for compounds of this type, see Fig. 5):

$$\begin{gathered}(NH_2 \text{ or OMe or OH or Cl or } NO_2 \text{ present in G\_1}) \text{ and} \\ \text{no additional rings} \\ n1 = 1, \quad n2 = 13 \end{gathered} \tag{14}$$

Two additional compounds, which are not covered by conjunction (14) with Br or F in G_1, are active, and p-Me appears in active as well as in inactive compounds depending on what features are present in other positions.

Motifs That Increase the Carcinogenic Potency

Ortho-substitution can also promote activity depending on the ortho-substituent and the substructural features appearing in other regions of the molecules. For example, OMe in the ortho-position occurs in four active and one inactive compound. The inactivity of the latter may be due to the presence of the strongly potency decreasing feature ''OMe in G_1'' (see above). However, two of the four active compounds also possess additional rings that will be shown to be strongly activity increasing motifs. This makes it difficult to decide whether or not o-OMe is a prominent activity enhancing motif. What can be safely concluded is that the substituent o-OMe is not incompatible with high carcinogenic potency in rats. The same is true for o-F. With o-Me the situation is even more complicated. Compounds with o-Me can show high carcinogenicity when substructures that enhance potency are present in the other positions. On the other hand, if o-Me is accompanied by strongly activity decreasing substructures the compounds will have low potency.

Conjunction (15) is an adjunction of ring variables and covers all compounds with more than one ring:

$$\begin{gathered}\text{Ring A or B or C or E or F or G or H or I or J or K} \\ \text{or L or N} \\ n1 = 33, \quad n2 = 6 \end{gathered} \tag{15}$$

Potency decreasing features occur in four out of the six inactive compounds. It is obvious, therefore, that within the structural space spanned by the compounds considered, addition of rings almost always leads to high potency as was already the case with mouse carcinogenicity. The following conjunctions provide a more detailed picture.

Conjunction (16) presents nine active compounds with ring J after elimination of two analogs with the inactivity features (see above) NH_2 in C_4 or E_4 (see Fig. 8):

$$\begin{gathered}\text{Ring J and not } (NH_2 \text{ in C\_1 or E\_4}) \\ n1 = 9, \quad n2 = 0 \end{gathered} \tag{16}$$

Thus, ring J clearly is a motif that increases potency. The compounds described by conjunction (16) include biphen-

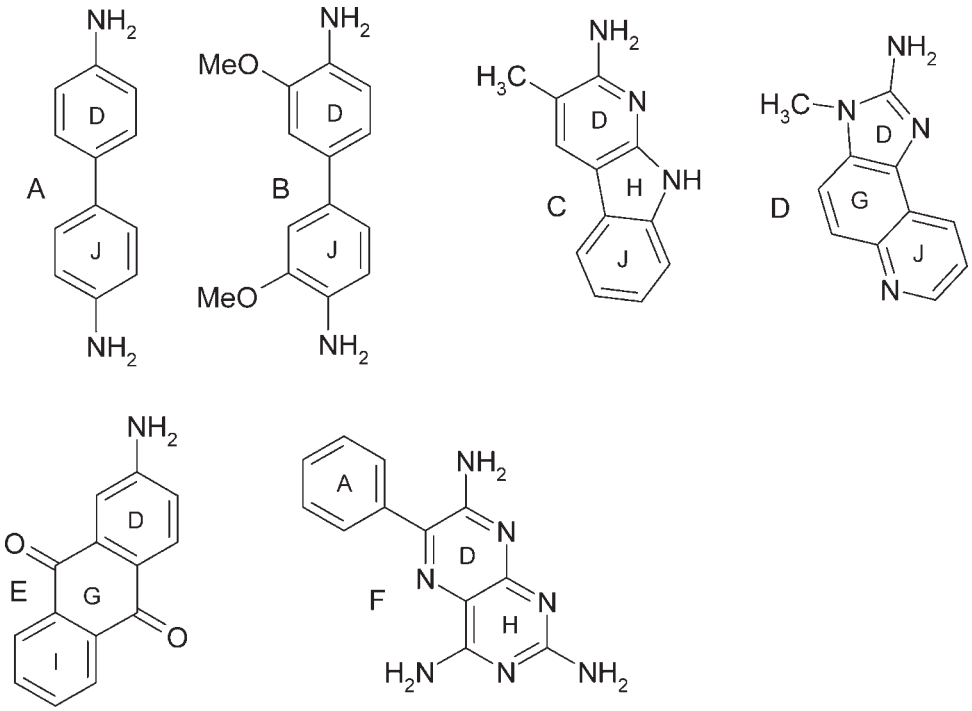

Fig. 8. Examples of compounds described by conjunctions (16) and (17). Compounds A and B fit into (16), C and D into (16) and (17), and E and F into (17). Lettering as in the hypermolecule.

yls and tricyclic systems. A bridge of the type C, CH, N, O, S, or SO_2 between the two phenyl rings (compounds of the type Ph-bridge-Ph-NH_2) is also compatible with high potency.

Another typical feature for high potency is a tricyclic system containing the rings G or H (note that ring D is always present); these rings are present in five of the nine compounds covered by conjunction (16) (see Fig. 8):

$$\begin{gathered}\text{Ring G or ring H but not } NH_2 \text{ in M\_1} \\ n1 = 11, \quad n2 = 0\end{gathered} \tag{17}$$

As shown in conjunction (18), many of the active compounds contain five-membered rings in D or G:

$$\begin{gathered}\text{Rings D or G 5-membered} \\ n1 = 10, \quad n2 = 0\end{gathered} \tag{18}$$

Conjunction (18) corresponds to conjunction (9) for mice carcinogenicity. Considering the nitrogens in ring D leads to conjunction (19):

$$\begin{gathered}(\text{N in position C\_2 or C\_3 or D\_2 or D\_3}) \text{ and} \\ \text{not } NH_2 \text{ in E\_4} \\ n1 = 11, \quad n2 = 0\end{gathered} \tag{19}$$

Conjunction (19) corresponds to conjunction (10) for mice carcinogenicity. As was the case for mice carcinogenicity, a large overlap between the groups of compounds described by the above conjunctions occurs (compare with Fig. 7). The conclusions are the same as those already discussed for mouse carcinogenicity. The important message is that additional rings (within the frame of the hypermolecule and with the possible exception of rings A and E) as well as N in certain ring positions, are activity enhancing motifs and lead to high rat carcinogenicity, unless their effect is counterbalanced by accompanying activity decreasing features.

CONCLUDING REMARKS

The results from EVAL aid in the evaluation of motifs that influence the carcinogenic risk of aromatic amines in mice and rats. As the carcinogenic potency of compounds is different in mice and rats (they correlate with only $r = 0.546$ in the compounds considered) there are both agreements and differences in the pattern of these motifs. Differences are observed mainly for the motifs that decrease the carcinogenic potency of the aromatic amines. In mice, substitutions in ortho and meta to the amino group tend to decrease the potency, as does the presence of a $-NO_2$ substituent in any position. In rats, these motifs affect the potency to a more limited extent (e.g., -NO_2 has a decreasing effect only in the meta position). On the contrary, increasing effects are quite similar in the two experimental animals and are exerted mainly by additional

rings, tricyclic systems, five-numbered rings, and N-heteroaromatic systems.

If more than one modulating motif is present in a compound, their effects can mutually enhance or compensate each other. Thus, the motifs must not be considered in an isolated context, and the carcinogenic potential of a given structure should be assessed considering all motifs and alerts. In addition, the relatively small training series of aromatic amines only allows limited conclusions. The list of risk modifying motifs, therefore, is incomplete and has to be continuously modified and extended as more experimental results become available. With these limitations in mind, the motifs evaluated by EVAL can still be of high practical value when prioritization for biological testing is required. In particular they can aid in the fine-tuning of results obtained from a more coarsely meshed primary filter, like the SA. In addition, estimation of the potency of active compounds can be useful in risk assessment procedures aimed at establishing levels, or thresholds, of exposure.

A key point to be kept in mind is that a downregulating motif will decrease potency but not necessarily lead to inactivity. Evidence from QSAR analyses of aromatic amines, and of mutagenic/carcinogenic chemicals in general, suggests that the structural factors and chemical properties that differentiate between active and inactive compounds are different from those that modulate the gradation of potency of the active compounds. In addition to several studies previously performed in our laboratory [Benigni et al., 2007a], a recent analysis showed that the probability for aromatic amines to be carcinogenic, as calculated by QSAR models, is not correlated with their potency (our unpublished results). The above results, taken together, indicate that the assessment of the carcinogenic risk of the amines through theoretical models should be performed in two steps: first, discrimination between positives and negatives; and second, estimation of the potency of the actives alone. The results of the present study are aimed at contributing to the second phase of the process.

REFERENCES

Benigni R. 2005. Structure-activity relationship studies of chemical mutagens and carcinogens: Mechanistic investigations and prediction approaches. Chem Revs 105:1767–1800.

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, Giuliani A, Franke R, Gruska A. 2000. Quantitative structure-activity relationships of mutagenic and carcinogenic aromatic amines. Chem Revs 100:3697–3714.

Benigni R, Bossa C, Netzeva TI, Rodomonte A, Tsakovska I. 2007a. Mechanistic QSAR of aromatic amines: New models for discriminating between homocyclic mutagens and nonmutagens, and validation of models for carcinogens. Environ Mol Mutagen 48:754–771.

Benigni R, Netzeva TI, Benfenati E, Bossa C, Franke R, Helma C, Hulzebos E, Marchant CA, Richard AM, Woo YT, Yang C. 2007b. 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 Revs 25:53–97.

Benigni R, Bossa C, Jeliazkova NG, Netzeva TI, Worth AP. 2008a. The Benigni/Bossa rulebase for mutagenicity and carcinogenicity—A module of Toxtree. EUR—Scientific and technical report series, EUR 23241 EN.Office for the Official Publications of the European Communities, Luxenbourg. Available at http://ecb.jrc.it/documents/QSAR/EUR_23241_EN.pdf.

Benigni R, Bossa C, Richard AM, Yang C. 2008b. A novel approach: Chemical relational databases, and the role of the ISSCAN database on assessing chemical carcinogenicity. Ann Inst Super Sanità 44:48–56.

Franke R. 1984. Theoretical Drug Design Methods. Amsterdam: Elsevier.

Franke R, Gruska A, Giuliani A, Benigni R. 2001. Prediction of rodent carcinogenicity of aromatic amines: A quantitative structure-activity relationships model. Carcinogenesis 22:1561–1571.

Hübel S, Franke R. 1991a. An interactive tool to evaluate two-dimensional pharmacophores. Pharmacochem Libr 16:37–42.

Hübel S, Franke R. 1991b. EVAL—A new tool to evaluate topological pharmacophores. In: Gmehling J, editor. Software Development in Chemistry 5. Berlin: Springer-Verlag. pp 85–90.

Kubinyi H. 1993. 3D QSAR in Drug Design: Theory, Methods and Applications. Leiden: ESCOM.

Lai DY, Woo YT, Argus MF, Arcos JC. 1996.Cancer risk reduction through mechanism-based molecular design of chemicals. In: DeVito SC, Garrett RL, editors. Designing Safer Chemicals. Green Chemistry for Pollution Prevention. Washington, DC: American Chemical Society. pp 1–13.

OECD. 2006. Report on the regulatory uses and applications in OECD member countries of (quantitative) structure-activity relationship [(Q)SAR] models in the assessment of new and existing chemicals. 58. OECD series on testing and assessment.Organisation of Economic Cooperation and Development, Paris.

Skog KI, Johansson MA, Jagerstad MI. 1998. Carcinogenic heterocyclic amines in model system and cooked foods: A review on formation, occurrence and intake. Food Chem Toxicol 36:879–896.

Streich WJ, Franke R. 1988. The use of mathematical logics in drug design. Math Comput Model 11:647–651.

Sugimura T. 1997. Overview of carcinogenic heterocyclic amines. Mutat Res 376:211–219.

Vineis P, Pirastu R. 1997. Aromatic amines and cancer. Cancer Causes Control 8:346–355.

Woo YT, Lai DY. 2001.Aromatic amino and nitro-amino compounds and their halogenated derivatives. In: Bingham E, Cohrssen B,Powell CH, editors. Patty's Toxicology. New York: Wiley. pp 969–1105.

Worth AP, Bassan A, de Brujin J, Gallegos Saliner A, Netzeva TI, Pavan M, Patlewicz G, Tsakovska I, Eisenreich S. 2007. The role of the European chemicals bureau in promoting the regulatory use of (Q)SAR methods. SAR QSAR Environ Res 18:111–125.

Accepted by—
R. Snyder