Quantitative Structure−Activity Relationships of Mutagenic and Carcinogenic Aromatic Amines

Romualdo Benigni,*,† Alessandro Giuliani,† Rainer Franke,‡ and Andreas Gruska‡

Istituto Superiore di Sanitá, Laboratory of Comparative Toxicology and Ecotoxicology, Viale Regina Elena 299, I-00161 Rome, Italy, and Consulting in Drug Design GbR, Gartenstr. 14, D-16352 Basdorf, Germany

Received November 10, 1999

Contents

I.	Introduction	3697
II.	Toxicology of Aromatic Amines: Epidemiological Evidence	3698
III.	Mechanisms of Action	3699
IV.	Qualitative Notions on the Structure−Activity Relationships of Aromatic Amines	3700
V.	QSARs for Mutagenicity	3700
VI.	QSARs for Carcinogenicity	3703
VII.	Original Model for the Carcinogenic Potency of Nonheterocyclic Aromatic Amines in Rodents	3704
	1. Data and Methods	3704
	A. Carcinogenicity Data of Aromatic Amines	3704
	B. Chemical Structures and Chemical Parameters	3704
	2. Results	3707
	A. Modeling the Carcinogenic Potency in Mice	3707
	B. Modeling the Carcinogenic Potency in Rats	3708
	3. Discussion of the Carcinogenic Potency QSARs	3710
VIII.	The QSARs of Aromatic Amines in Perspective	3711
VIII.	Acknowledgments	3713
IX.	References	3713

I. Introduction

Among the various toxicity endpoints, chemical carcinogenicity is of primary interest because it drives much of the current regulatory actions on new and existing chemicals and its experimental determination involves time-consuming and expensive animal testing. However, only a relatively small percentage of the chemicals in commerce have currently undergone testing, so the support of structure−activity relationship (SAR) and quantitative structure−activity relationship (QSAR) approaches (as tools for both predictive toxicology and mechanism elucidation) in this field are of particular interest. In recent years, there has been strong pressure from society, in general, and from government agencies, in particular, to develop "general" prediction models in order to cope with the thousands of chemicals present in the environment, for which experimental data are not available and likely will never exist. Two recent comparative exercises on the prediction of chemical carcinogenicity using different methods or algorithms provided extremely important evidence on this subject.[1,2] It was demonstrated that the present level of SAR knowledge permits the identification of many potentially carcinogenic chemical functionalities. Thus, application of the SAR knowledge is already reliable for an efficient use in priority setting, as demonstrated by the successful prioritization performed by the U.S. National Toxicology Program, which found 70% carcinogens among the suspect chemicals, whereas only 10−20% of the chemicals selected on exposure/production considerations (hence without any bias in terms of biological activity) were carcinogenic.[3] However, a common weakness of the approaches was the difficulty in correctly predicting the noncarcinogens with alerting functionalities, i.e., the presence of a structurally alerting feature could be negated by other structural factors modulating potency or eliminating activity. Hence, although current prediction methods are reasonably successful at discerning major, structurally alerting classes of carcinogens, greater uncertainty is associated with the predictions for individual chemicals, because methods do not adequately discriminate activity within these classes.

A possible way to at least partially overcome this difficulty is to develop QSAR models for different classes of chemicals and to use the resulting models—after assigning the compounds to be evaluated to the correct class—for predictions. To be able to do that it is, of course, necessary to make such QSARs available. Whereas collections of QSARs of individual classes of toxic chemicals are largely available for some end points (e.g., aquatic toxicity[4]), QSARs for classes of carcinogens are quite limited and sparse. One of the reasons is that quantitative data on carcinogenic potency are largely missing. A remarkable exception is the class of aromatic amines. Obviously, the level of use and industrial importance have determined such large experimentation. In fact, aromatic amines are widespread chemicals with considerable industrial and environmental importance: for example, aromatic amine-derived dyes are synthetic organic colorants, widely used in the textile, paper, leather, plastics, cosmetics, drugs, and food

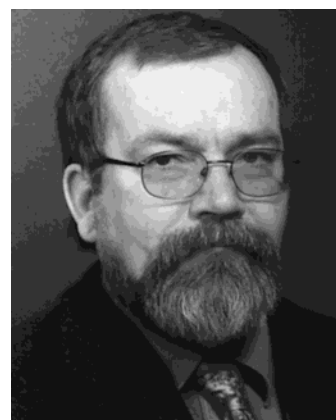
* To whom correspondence should be addressed. Telephone: +39 06 49902579. Fax: +39 06 49387139. E-mail: rbenigni@iss.it.
† Istituto Superiore di Sanitá.
‡ Consulting in Drug Design GbR.

10.1021/cr9901079 CCC: $35.00 © 2000 American Chemical Society
Published on Web 08/17/2000

Romualdo Begnigni received his education in chemistry at the University of Rome "La Sapienza". He then joined the Istituto Superiore di Sanitá (Italian National Institute of Health), where he got a permanent position in 1977 and remained except for a sabbatical at New York University in 1988. He worked experimentally in the field of molecular biology and environmental chemical mutagenesis. In the 1980s he turned his attention to the statistical analysis and modeling of toxicological data and to the study of the relationships between the structure of organic compounds and their toxicological properties (mainly mutagenesis and carcinogenesis).

Alessandro Giuliani received his degree in biological sciences at the University of Rome "La Sapienza" in 1982. Until 1977 he worked at the Sigma-Tau Laboratories; then he moved to the Istituto Superiore di Sanitá, where now he is Senior Scientist in the unit directed by Dr. Benigni. His research interests focus on the exploration of soft modeling approaches to quantitate biological experimental results. To this aim, he applied multidimensional techniques to a variety of fields ranging from classical QSAR to physiology, behavioral sciences, cardiology, and signal analysis. In recent years he has collaborated on the development and application of a novel signal analysis technique, recurrence quantification analysis (RQA), to physiology, molecular dynamics, and complex systems physics. Recently the technique was demonstrated to be promising in developing quantitative sequence/activity studies in proteins.

Rainer Franke received his Dr. rer. nat degree in physical chemistry from Technical University Dresden and his habilitation from Martin-Luther University Halle/Wittenberg. After working in a district hospital in Dresden and at the Humboldt University in Berlin in the field of biochemistry, he joined the Academy of Sciences of the GDR in Berlin doing research in medicinal chemistry. Since 1991 he has been Director of Consulting in Drug Design GbR. His primary interest is in the relationship between the structure of organic compounds and their biological properties.



Andreas Gruska studied chemistry at the Ernst-Moritz Arndt University in Greifswald. After working with R. Franke at the Academy of Sciences of the GDR in the drug design field for four years, he joined Chemiekombinat Bitterfeld doing pesticide design. In 1982 he joined the Parmacological Institute of the Ernst-Moritz Arndt University in Greifswald. Since 1992 he has been Vice-Director of Consulting in Drug Design GbR in Basdorf. His research interests include QSAR and computer-assisted drug design.

industries. Moreover, several types of aromatic amines are generated during cooking.[5-7]

This paper, first, illustrates the toxicological problems posed by the aromatic amines, together with evidence on their mechanisms of action. A short overview of the qualitative structure–activity notions derived from the above evidence is also provided. Then we review in detail the available QSARs for the experimental results on the mutagenic and carcinogenic properties of the amines. Given the paucity of QSARs for carcinogenicity, we developed ad hoc for this paper QSAR models for the carcinogenic potency of nonheterocyclic aromatic amines. Finally, the various QSARs are put into perspective.

II. Toxicology of Aromatic Amines: Epidemiological Evidence

The aromatic amines are one of the chemical classes in which the structural and molecular basis of carcinogenicity is most clearly understood.[5] This class of molecules offers the unique possibility of covering all the investigation levels, ranging from physicochemical properties to epidemiological findings in human populations, with rational explanations.

Exposure to aromatic amines occurs in different industrial and agricultural activities as well as in tobacco smoking. Aromatic amines have been used as antioxidants in the production of rubber and in cutting oils, as intermediates in azo dye manufacturing, and as pesticides. They are a common contaminant in several working environments, including the chemical and mechanical industries, and arylamines-based dyes are widely used in textile industry and in cosmetics.[8] The wide use of aromatic amines

together with the presence of relatively specific, very high exposures permitted the development of epidemiological knowledge unparalleled for other chemical classes.

Bladder cancer in men is the most studied tumor type: a large number of studies (see Vineis and Pirastu[8] for a comprehensive review) relating professional exposure to arylamines (both complex mixtures and single chemical agents) and bladder cancer have been published. The odds ratios (the ratio between the tumors observed in exposed population and the tumors observed in a carefully matched control population) for arylamine exposure go from around 2 (i.e., a 2-fold increase in the probability of developing bladder cancer) for very mild exposures up to around 100 for extremely high exposures.[8–11]

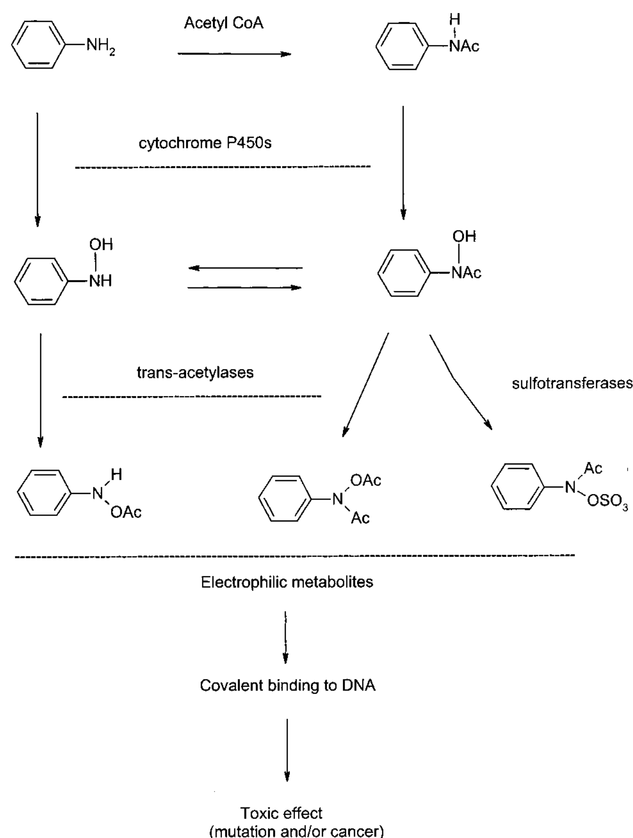
Most of the above studies refer to exposures to mixtures of aromatic amines. For 2-naphthylamine, *o*-toluidine, benzidine, and 4-aminobiphenyl, it has been possible to select cohorts of individuals experiencing exposure that can be reasonably considered as single-agent exposure,[8] thus providing formal demonstration of the carcinogenic potential of these agents for humans. In the case of 4-aminobiphenyl, there are molecular epidemiology studies[8,12] that were able to identify a specific DNA adduct identified as a derivative of 4-aminobiphenyl. This same DNA adduct was present in exfoliated bladder cells of smokers;[13] the presence and concentration of DNA adducts was correlated strongly with 4-aminobiphenyl–hemoglobin adducts. The 4-aminobiphenyl–hemoglobin adducts in both smokers and nonsmokers is modulated by the *N*-acetylation phenotype: irrespective of the smoking status of the subjects, the genetically based slow-acetylator phenotype was associated with high concentrations of the adduct.[8]

The evidence regarding the carcinogenic potential of aromatic amines in animals was available before formal epidemiologic studies were conducted: in this sense, arylamines are one of the best examples of the predictivity of animal experiments for human risk.[14] The evidence in experimental animals has been crucial in the classification of some aromatic amines for their carcinogenicity to humans. Benzidine-based dyes and MOCA (4,4′-methylene bis-2-chloroaniline) were classified by the International Agency for Research on Cancer (IARC) as probable carcinogens based on the strong evidence in animals before epidemiological evidence was available.[8]

III. Mechanisms of Action

The aromatic amines have to be metabolized to reactive electrophiles in order to exert their carcinogenic potential. Scheme 1 provides a simplified picture of the main metabolic steps. For aromatic amines and amides, this typically involves an initial *N*-oxidation to *N*-hydroxyarylamines and *N*-hydroxyarylamides, which in rat liver is mediated primarily by cytochrome P-450 isozyme *c* (BNF–B) and *d* (ISF-G).[15,16] The initial activation of nitroaromatic hydrocarbons is likewise through the formation of an *N*-hydroxyarylamine, a reduction catalyzed by both microsomal and cytosolic enzymes.[5,16] Microsomal nitroreduction also appears to depend on cytochrome P-450 complex, in particular rat liver isozymes *c, d* and *b* (PB–B) and *e* (PB-D). Cytosolic nitroreductase activity is associated with a number of enzymes, including DT-diaphorase, xanthine oxidase, aldehyde oxidase, and alcohol dehydrogenase.[16] In addition to the reactions of nitrogen oxidation and reduction (main activation pathways), certain aromatic amines and nitroaromatic hydrocarbons are converted into electrophilic derivatives through ring-oxidation pathways. *N*-Hydroxyarylamines, iminoquinones, and epoxide derivatives are directly electrophilic metabolites, while *N*-hydroxyarylamides require esterification before becoming capable of reacting with DNA.[17]

Scheme 1. Pattern of Metabolic Activation Pathways of Aromatic Amines. The Scheme Sketches the Most Representative Pathways (see details in the text)



A case in point of the crucial role played by metabolism in determining biological activity of aromatic amines is the case of 1-naphthylamine. This chemical was originally considered to be a human bladder carcinogen: the results of subsequent epidemiological studies coupled with the failure to demonstrate a carcinogenic response in animal models indicated that this is not the case.[18,19] This lack of carcinogenicity appears to be due to the failure of 1-naphthylamine to be metabolized to a reactive electrophile.[20] Although 1-naphthylamine has not been found to be carcinogenic, its *N*-oxidized derivative, *N*-hydroxy-1-naphthylamine, is strongly tumorigenic.[18,21,22] *N*-Hydroxy-1-naphthylamine readily

binds to DNA, and the reaction results in the formation of a major DNA adduct through reaction of the aryl nitrogen and ortho carbon atoms at O^6 of deoxyguanosine.[23] Evidence has also been presented that a minor adduct is formed by N-substitution at C8 of deoxyguanosine.[24] This last reaction is typical for *N*-hydroxyarylamines, whereas reaction with the O^6 position, which is normally associated with S_N1-type reactions, seems to be unique to *N*-hydroxy-1-naphthylamine. The reaction of *N*-hydroxyarylamines with DNA is proposed to proceed through a protonated nitrenium ion pair;[17] thus, the relative stability of this reaction intermediate appears to be a crucial point in determining the biological activity of aromatic amines.

The DNA adducts generated in animals are similar to those found in vitro and have a very variable persistence in tissues for the different aromatic amines. This difference in persistence may result from the fact that different structural distortions of the DNA are recognized with different efficiency by the DNA repair enzymes that operate the excision of the adduct.[16]

The polymorphism and differential distribution of the enzymes responsible for the metabolic activation of aromatic amines has a crucial role in determining the organ specificity observed with these substances. For instance, if *N*-acetylation precedes *N*-oxidation, the concentration of *N*-hydroxyarylamine available for transport to the bladder decreases.[16] Thus, individuals with a rapid acetylator phenotype should be at a lower risk for bladder cancer from exposure to aromatic amines, which is what has been actually observed.[25] Likewise, the inability of dogs to *N*-acetylate aromatic amines is consistent with their susceptibility to bladder tumors. While acetylation appears to afford protection from bladder tumor induction, the opposite may be true for other tissues. Thus, a higher incidence of colon cancer has been found in low-risk individuals with a rapid acetylator phenotype.[16]

There is evidence[26] in rats that the expression of acetyltransferase in tissues of the central nervous, gastrointestinal, urinary, and reproductive systems is highly regulated, as it is in other organs commonly associated with aromatic amine carcinogenicity. The subtleties and specificities of such complex and highly organ-specific toxification/detoxification balance produce a high variability in the target organs of aromatic amines that in fact exert their carcinogenic potential at many different sites.

IV. Qualitative Notions on the Structure–Activity Relationships of Aromatic Amines

The large amount of data on animal carcinogenesis allowed for the sketching of some basic SAR requirements for the carcinogenesis induced by aromatic amines. These qualitative rules are clearly summarized by Lai et al.[5] The basic requirement is the presence of an aromatic ring system (a single ring or more than one ring forming a conjugated system, fused or nonfused) and the amine/amine-generating group(s). Amine-generating groups (due to metabolic interconversion) are typically the hydroxylamino, nitro, and nitroso groups. In some cases, replacement of an amino group with a dimethylamino group does not result in a significant loss of the carcinogenic activity of aromatic amine compounds since metabolic *N*-demethylation readily occurs in vivo. Other important structural features are (1) the number and nature of aromatic rings, (2) the nature and position of the amine/amine generating groups, (3) the nature number and position of other ring substituents, and (4) the size, shape, and polarity of the molecules. Interestingly, many of the structural features that are important for the carcinogenicity also have important influences on their bioactivation mechanisms.

The number and nature of aromatic rings modulates the carcinogenic potential of aromatic amines via the modulation of the leaving potential of the acyloxy anion that is the rate-limiting step of the bioactivation process. The force of conjugation, facilitating the departure of acyloxy anion, increases from phenyl toward higher aryl groups. This is consistent with the findings that aniline (single phenyl ring) is a weaker carcinogen than benzidine or β-naphthylamine (two phenyl rings) and more likely with the presence of the term “number of rings” in the QSARs of the aromatic amines.[27] Even the nature and position of the amine or of the amine-generating group influences the carcinogenic potential at the level of bioactivation step: for example, for dialkylamino groups with bulky or long alkyl substitution, *N*-dealkylation does not readily occur to allow further bioactivation. Replacement of the dimethylamino group of 4-dimethylaminoazobenzene by a diethylamino or a higher dialkylamino group has been shown to lead to a marked attenuation of its carcinogenicity[28] and mutagenicity.[29]

Ring substituents other than amino or amino-generating groups have been reported to modulate aromatic amines carcinogenicity mainly by steric effects: the larger the substituents (especially in the ortho position), the less potent the chemical.[5] On the contrary, the substitution of a chloro group or a methyl/methoxy group ortho to the amino group often enhances activity.[30,31]

V. QSARs for Mutagenicity

Because of the shortcomings of the rodent carcinogenicity bioassay (long times, high price, sacrifice of large numbers of animals), the aromatic amines have repeatedly been tested in short-term mutagenicity assay, notably with the *Salmonella typhimurium* (Ames test) bacterial assay.[32,33] This assay is a reliable tool for qualitatively predicting rodent carcinogenicity (hence for extrapolation to humans), since chemicals which are positive in the Ames test have a high probability of also being rodent carcinogens (80% for the general “universe” of chemicals, with differences from class to class). It should be added that the reverse is not true: unfortunately, a negative Ames test does not provide useful informa-

tion, since it has been shown that an Ames-test negative chemical has about the same probability of being a carcinogen or a noncarcinogen.[34−37] The large number of experiments on aromatic amines performed with the Ames test has provided a large database of mutagenicity results that have been studied with QSAR approaches by several authors. Two reviews have appeared on such QSARs.[38,39] The following is a presentation of the individual QSAR studies.

Trieff et al.[40] studied the Salmonella mutagenicity of 19 aromatic amines tested in the strains TA98 (frame-shift mutations) and TA100 (base-pair mutations), with the addition of S9 metabolizing fraction from Aroclor 1254-induced rat liver. Separate QSAR models were found for the two strains by multiple linear regression.

$$\log \text{BR-TA98} = -1.639 (\pm 0.399) + 0.816 (\pm 0.127) \log P - 0.752 (\pm 0.174) I_1 + 0.377 (\pm 0.174) I_2$$

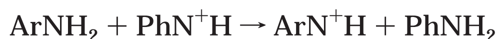
$$s = 0.78 \quad n = 19 \quad r^2 = 0.78 \qquad (1)$$

$$\log \text{BR-TA100} = -1.559 (\pm 0.282) + 0.784 (\pm 0.090) \log P - 0.735 (\pm 0.123) I_1 + 0.496 (\pm 0.123) I_2$$

$$s = 0.80 \quad n = 19 \quad r^2 = 0.88 \qquad (2)$$

The bacterial mutagenic potency was defined as BR = 1 + NR/nmol, where NR is the net revertant number. The revertants are the cells that underwent mutation. The indicator variable I_1 was 1.0 if the amine or acetamido group was proximal (adjacent) to the juncture (i.e., the carbon atom connecting the substituted ring with the rest of the molecule). I_2 related to whether the amine group was free ($I_2 = 1$) or acetylated ($I_2 = 0$). Equations 1 and 2 are quite similar and show that mutagenicity increased with lipophilicity. On the other hand, mutagenicity was reduced when the amine or acetamido group was in ortho to the juncture, because of steric hindrance in its biotransformation. Mutagenic potency was also decreased by the acetylation of the amino group, probably because the acetyl group needs to be first split off prior to oxidation of the amine group.

Ford and Griffin[41] related the mutagenicity of a variety of heteroaromatic amines present in cooked foods with the stabilities of the corresponding nitrenium ions (see Scheme 1). The stability of the nitrenium ions was measured by the calculated energy ($\Delta\Delta H$) of the process



$$ArNH_2 + PhN^+H \rightarrow ArN^+H + PhNH_2$$

$\Delta\Delta H$ was calculated using the semiempirical AM1 molecular orbital procedure. It appeared that the mutagenic potencies (m) in three Salmonella strains (TA98, TA100, and TA1538) correlated with the $\Delta\Delta H$ values according to the following equations.

$$\log(m)\ \text{TA98} = -0.181 (\pm 0.043) \Delta\Delta H + 0.227 (\pm 0.2792)$$

$$s = 0.966 \quad r^2 = 0.593 \quad n = 14 \qquad (3)$$

$$\log(m)\ \text{TA100} = -0.147 (\pm 0.024) \Delta\Delta H - 0.1619 (\pm 0.450)$$

$$s = 0.540 \quad r^2 = 0.770 \quad n = 13\text{f} \qquad (4)$$

$$\log(m)\ \text{TA1538} = -0.2417 (\pm 0.0353) \Delta\Delta H - 0.801 (\pm 0.765)$$

$$s = 0.245 \quad r^2 = 0.922 \quad n = 6 \qquad (5)$$

Ford and Herman[42] studied the relative energetics ($\Delta\Delta H$) of arylamine *N*-hydroxylation and N−O heterolysis ($ArNH_2 \rightarrow ArNHOH \rightarrow ArNh^+$) for condensed systems of two, three, and four rings using semiempirical AM1 molecular orbital theory. Limited correlations between the energetics of nitrenium ion formation and experimental TA98 and TA100 mutagenicities were found.

An important contribution to the QSAR modeling of aromatic and heteroaromatic amines mutagenicity was provided by Debnath et al.,[43] who collected data on a wide number of chemicals with largely different basic structures (e.g., aniline, biphenyl, anthracene, pyrene, quinoline, carbazole, etc). The experimental data referred to Salmonella TA98 and TA100 strains, with S9 metabolic activation. The mutagenic potency is expressed as log(revertants/nmol). The AM1 molecular orbital energies are given in electronvolts. The mutagenic potency in TA98 + S9 was modeled by

$$\log \text{TA98} = 1.08 (\pm 0.26) \log P + 1.28 (\pm 0.64)\ \text{HOMO} - 0.73 (\pm 0.41)\ \text{LUMO} + 1.46 (\pm 0.56) I_L + 7.20 (\pm 5.4)$$

$$n = 88 \quad r = 0.898 \quad s = 0.860 \qquad (6)$$

where HOMO is the energy of the highest occupied molecular orbital, LUMO is the energy of the lowest unoccupied molecular orbital, and I_L is an indicator variable that assumes a value of 1 for compounds with three or more fused rings. The electronic terms HOMO and LUMO, though statistically significant, accounted for only 4% of variance, whereas log P alone accounted for almost 50%. The most hydrophilic amines ($n = 11$) could not be treated by eq 6 and were modeled by a separate equation containing only log P, thus suggesting that these amines may act by a different mechanism. The mutagenic potency in the Salmonella strain TA100 + S9 was expressed by

$$\log \text{TA100} = 0.92 (\pm 0.23) \log P + 1.17 (\pm 0.83)\ \text{HOMO} - 1.18 (\pm 0.44)\ \text{LUMO} + 7.35 (\pm 6.9)$$

$$n = 67 \quad r = 0.877 \quad s = 0.708 \qquad (7)$$

Also in this case, a different equation was necessary for the most hydrophilic amines ($n = 6$). Overall, the principal factor affecting the relative mutagenic-

ity of the aminoarenes was their hydrophobicity. Mutagenicity increased with increasing HOMO values: this positive correlation seems reasonable since compounds with higher HOMO values are easier to oxidize and should be readily bioactivated. For the negative correlation with LUMO, on the other hand, no simple explanation could be offered by the authors. A remarkable difference between the models for the two Salmonella strains was that the TA100 QSAR lacked the I_L term present in the TA98 model. It was hypothesized that larger amines are more capable of inducing frame shift mutations (TA98 is specific for frame shift mutations, whereas TA100 is specific for base pair substitution mutations) and that this effect is not accounted for by the increase of log P at increasing size of the molecules.

The above paper deserves two more comments. First, in a parallel work the authors[44] modeled the mutagenicity of nitroarenes. The main metabolic pathway of the nitroarenes is supposed to include the formation of the hydroxylamine by cytosolic reductase; then the fate of the activated compound should be identical to that of amines.[5,16] As to be expected, the equations reported for the nitroarenes are qualitatively very similar to the amine equations, with the major difference being that the HOMO term (related to the oxidative step of the amines) is missing.[44] This indicates that such equations not only provide a means for predicting mutagenicity, but can also reveal aspects of the activation mechanism. A second comment concerns inactive compounds. While the QSARs for the aromatic amines are quite good in modeling mutagenic potency, they are less satisfactory when one wants to predict the activity of the nonmutagenic amines: in many cases inactive compounds are incorrectly predicted to be highly mutagenic.[43]

For the same set of compounds considered by Debnath et al.,[43] the discrimination between mutagenic and nonmutagenic amines was studied more in detail by Benigni et al.[45] It appeared that lipophilicity alone had no discriminating power in TA98 and TA100, which is at odds with the major role played in the modulation of potency within the group of active compounds. Though statistically significant, discriminant functions separating mutagenic from nonmutagenic amines showed a reclassification rate of only about 70% accuracy. They were based mainly on electronic and steric hindrance factors. The same was true for the nitroarenes mutagenicity. In a second paper, the same group[46] tried to improve the discriminant models for the mutagenic activity of the amines in Salmonella. The best discrimination was obtained by splitting the amines into structural subclasses. The single-ring amines were best separated by electronic factors (first HOMO and second LUMO, in decreasing order of importance) (correct reclassification rate around 70%). This result confirmed the central role of metabolic transformation in the mutagenic activity of these chemicals. The diphenylmethanes were modeled by the contribution to molar refractivity of the substituents in ortho position to the functional group, thus indicating the negative effect of steric hindrance on the accessibility of the metabolizing system (correct reclassification rate: 87% for TA98; 93–100% for TA100). Steric factors, as measured by a similarity index, were also a key factor in the discrimination of biphenyls. The fused-rings amines were all mutagenic, so no discriminant model was necessary. The authors concluded that the minimum requirements for the mutagenicity of the aromatic amines (as modeled by the discriminant functions) were different from the factors ruling the modulation of potency.

Using their computer program CASE, Klopman et al.[47] analyzed a set of approximately 100 aromatic amines. The CASE methodology is a software package that selects its descriptors automatically from a learning set of molecules. It identifies single, continuous structural fragments that are embedded in the complete molecule and selects those that are statistically associated with activity or nonactivity or with increasing potency. Normally, the program screens the molecules for all the possible fragments ranging from 2 to 10 heavy (nonhydrogen) atoms. The program was used to examine mutagenicity in Salmonella strains TA98 and TA100 (with S9 activation) and yielded a number of structural features associated with mutagenicity and nonmutagenicity. This work was extended by Zhang et al.,[48] who studied 61 heterocyclic amines formed during food preparation. In both studies, the major feature leading to mutagenic activity was the aromatic amino group. Electronic parameters were also calculated, and the LUMO energy was found to correlate negatively with the mutagenic potency of the molecules. A model based on a number of fragments (the amino group in different combinations of atoms) together with the LUMO attained $r^2 = 0.857$.

Lewis et al.[49] studied a noncongeneric set of food mutagens, the majority being heterocyclic amines ($n = 17$). This study was in line with other studies of the same group aimed at highlighting the structural determinants that make the chemicals good substrates for cytochrome P4501 (CYP1). For the TA98 strain (frame shift mutations) of Salmonella, the best correlation of mutagenicity was with molecular diameter $r = 0.91$, hence with planarity. For the TA100 strain (base pair mutations), the best correlation was with the difference between the LUMO and HOMO energies: high mutagenicity was related to low values of the difference, hence to high chemical reactivity.

Basak and Grunwald[50] explored the suitability of "rough and fast" QSAR models based on easily calculable theoretical indices. For a set of 73 aromatic and heteroaromatic amines—previously studied by Debnath et al.[43]—the authors calculated a wide range ($n = 90$) of topological indices. Then they constructed five similarity spaces based on (a) counts of atom pairs, (b) principal components (PC) from the topological indices, (c) PCs from topological indices plus physicochemical parameters used by Debnath et al.,[43] (d) PCs from physicochemical parameters, and (e) physicochemical parameters. In each of the five similarity spaces, the mutagenic potency of every chemical was estimated by averaging the potency of its k-nearest neighbors ($k = 1-5$). It appeared that

the easily computable method based on atom pairs was almost as reliable ($r = 0.77$) as the similarity method based on physicochemical properties ($r = 0.83$). The disadvantage with this type of descriptors is, of course, that interpretability is very limited.

Hatch et al.[51] studied the mutagenic potency (frame shift mutations in TA98 or TA1538 Salmonella strains) of a series of heteroaromatic amines formed during the cooking of the food from two classes: aminoimidazoazaarene (AIA) ($n = 38$) and aminocarboline (AC) ($n = 23$). For the AIA compounds, the features relevant for the mutagenic activity were as follows: number of fused rings, number of heteroatoms in rings 2 and 3, methyl substitution on imidazo ring nitrogen atoms, and methyl substitution on ring carbon atoms ($r^2 = 0.78$). The relevant features for the AC compounds were as follows: position of the pyridine-type nitrogen atom in ring 1, position of the exocyclic amino group in ring 1, and methyl substitution at ring carbon atoms ($r^2 = 0.80$) The goodness of fit values referred to models including all the relevant features. In a further analysis, Hatch et al.[52] considered several molecular orbital properties calculated at different approximations, together with structural factors, for 16 AIA mutagens and their nitrenium ion metabolites. The major findings were as follows: (1) the potency increased with the size of the aromatic ring system, (2) potency was enhanced by the presence of an *N*-methyl group, (3) introduction of additional nitrogen atoms in pyridine, quinoline, and quinoxaline rings supported potency, (4) potency was inversely related to the LUMO energy, (5) potency was directly (although weakly) related to the LUMO energy of the derived nitrenium ions, (6) the calculated thermodynamic stability of the nitrenium ions was directly correlated with nitrenium LUMO energy and with the negative charge on the exocyclic nitrogen atom. The authors commented about the lack of a clear explanation for the role of LUMO energy, since the oxidation of the amine group was expected to be the main rate-limiting step in the metabolism of the amines. Hatch and Colvin[53] reconfirmed the above results in a wider set of 95 aromatic and heteroaromatic amines, together with the puzzling role of the LUMO energy.

Maran et al.[27] reevaluated the data set collected by Debnath et al.[43] with a very large set of descriptors ($n = 619$), including various constitutional, geometrical, topological, electrostatic, and quantum chemical descriptors. A final model with six descriptors was established ($r^2 = 0.8344$). The most important descriptor was the number of aromatic rings, followed by (in decreasing order of importance) γ-polarizability (second-order hyperpolarizability), hydrogen-acceptor surface area, hydrogen-donor surface area, maximum total interaction energy for the C−C bond, and maximum total interaction energy for a C−N bond. Maran et al.[27] concluded that the leading descriptor in their model (number of rings) was approximately proportional to the area of the hydrophobic aromatic hydrocarbon part of these molecules and was thus directly related to the hydrophobicity of polycyclic and condensed aromatic compounds (correlation coefficient between number of rings and log P $r^2 = 0.3715$). This correlation is only weak, and the authors stressed that they could not add log P to their model. There is probably a high multiple correlation between the entirety of their variables and log P which was not investigated. However, the number of rings was preferred by the authors to log P based on the argument that it is not an empirical parameter. The HOMO and LUMO energies did not appear in the model.

VI. QSARs for Carcinogenicity

Although the major concern posed by the aromatic amines derives from their carcinogenic potential, the number of QSAR studies is quite limited.

Yuta and Jurs[54] applied their ADAPT (automatic data analysis using pattern-recognition techniques) software system to a set of 157 aromatic amines; to be included into the data set, a compound was required to have biological activity data reported (either positive or negative) in at least three organ sites, it had to be aromatic amine, and it had to belong to one of five common structural classes (biphenol, stilbene, azo-compounds, fluorene, methylene). Topological and geometrical descriptors were used, and to avoid chance separations, multicollinearities were checked and the number of descriptors was reduced to 31. Particularly important were the molecular connectivity environment descriptors, based on structural features related to the theory on the mechanisms of action of the aromatic amines (e.g., primary or secondary amines, presence of bridging groups, etc.). The analyses were repeated with several pattern-recognition methods (Bayesian quadratic discrimination, Bayesian linear discriminant, *K*-nearest neighbor classification, iterative least-squares linear discrimination, simplex discriminating algorithm, linear learning machine). Each compound was considered to be either active (at least three active sites) or inactive (negative in all sites). The chemicals were divided in 11 possible subsets, according to organs and route of administration. Several QSAR analyses were performed, on the different subsets and on the entire set of chemicals, with the various pattern-recognition methods. The iterative least-squares program enjoyed the most success (classification rates around 90%). Overall, the analyses indicated that the number of rings (related to molecular volume or bulk) is an important descriptor relating aromatic amino structure to carcinogenic potential. Other important descriptors were those related to size and shape (e.g., smallest principal moment). Several subsets of descriptors supported linear discriminant functions that could separate carcinogens from noncarcinogens.

Loew et al.[55] challenged the capabilities of theoretical chemistry to characterize the chemicals as well as the physical and chemical interactions with the biological targets. Eight aromatic amines were selected for the study; the sample was small but consisted of four pairs of isomeric amines. One of each pair was an active carcinogen, while the other was inactive or of doubtful activity. Mutagenic potency data, even though not obtained with the same bacterial strain, paralleled the carcinogenic activity; the

weak mutagens were the inactive or more marginally active carcinogens. These pairs of isomers were selected as ideal tests of the ability of calculated electronic parameters alone to predict relative biological activity, since effects such as transport and elimination should be more nearly the same for both isomers of a given pair than for the group as a whole. Electronic reactivity parameters relevant to the relative ease of metabolic transformation of each parent compound to hydroxylamine by cytochrome P-450, as well as to other competing metabolic products involving ring epoxidation and hydroxylation, were calculated. Comparing the results for pairs of isomers, in each case the value of the N-atom superdelocalizability—chosen as an indicator of the extent of formation of hydroxylamine from parent compounds—was larger for the more potent mutagen/carcinogen. Moreover, the less potent isomer in each pair had the ring carbon which was most reactive (i.e., larger values of ring carbon superdelocalizability) to direct phenol formation, which appeared to be an effective detoxification pathway. Ring epoxidation (as measured by π-bond reactivity) appeared to be more activating than detoxifying. In addition, two measures of covalent adduct formation ability of the hypothesized intermediate reactive species (arylnitrenium ion) paralleled the biological activity within each pair (electron density on N and C_β atoms in the lowest energy empty molecular orbital of the arylnitrenium ion).

For completeness, the work of Vracko[56] and of Gini et al.[57] should be mentioned, which will, however, not be discussed in detail as it is not specifically concerned with aromatic amines and is thus out of the scope of this review. QSAR models based on theoretical descriptors were derived for noncongeneric sets of benzene derivatives, including different proportions of aromatic amines, using artificial neural networks. The models devised by Vracko[56] were able to describe the training set, but their prediction ability of carcinogenic potency (TD_{50}) was limited. Gini et al.[57] performed a retrospective study on 104 N-containing benzene derivatives that resulted in a quite good correlation after removal of several outliers.

VII. Original Model for the Carcinogenic Potency of Nonheterocyclic Aromatic Amines in Rodents

Whereas several QSAR models have been generated for the mutagenicity of the aromatic amines, we have found in the literature only two models specific for their rodent carcinogenicity;[54,55] moreover, only the yes/no activity was modeled. Vracko[56] considered the carcinogenic potency of a number of amines within a model for noncongeneric aromatic chemicals. For this reason and for the sake of completeness, we performed ad hoc for this paper a QSAR analysis of the carcinogenic potency of the nonheterocyclic aromatic amines.

1. Data and Methods

A. Carcinogenicity Data of Aromatic Amines

The carcinogenic potency data used for this study were the TD_{50} (mg/kg/day) values calculated by Gold et al.[58] The TD_{50} is the daily dose rate required to halve the probability of an experimental animal of remaining tumorless to the end of its standard life span. We used the TD_{50} values for rat and mouse as reported in the Carcinogenic Potency DataBase (CPDB), available at the Internet site http://potency.berkeley.edu/hybrid.html. These are harmonic means of the TD_{50} values for the different tumor types, averaged over the rodent species. For the scope of the QSAR analyses, carcinogenic potency was defined as follows: mice, $BRM = \log(MW/TD_{50})_{mouse}$; rats, $BRR = \log(MW/TD_{50})_{rat}$, where MW is the molecular weight.

B. Chemical Structures and Chemical Parameters

Table 1 summarizes the structures of the compounds (anilines, biphenylamines, naphthylamines, and aminofluorenes) for which carcinogenic potency data were available. Chemical structures are presented as substituted anilines according to the conventions outlined below.

To describe the chemical properties of the compounds, global and local parameters were used. Global electronic properties were characterized by the EHOMO (energy of the highest occupied molecular orbital) and ELUMO (energy of the lowest empty molecular orbital) calculated by the semiempirical molecular orbital method, AM1, after optimizing structures at the same level of theory (program system SYBYL, Tripos). Overall hydrophobicity is expressed in terms of log P computed from the program Tsar (Oxford Molecular).

In an attempt to gain some insight into possible local effects, ring substituents were characterized by hydrophobic, electronic, and steric substituent constants. Of the many parameters tried, the following appear in the resulting QSARs: $\mathcal{I}$ and $\mathcal{R}$ (inductive and resonance-polar electronic substituent constants according to Swain and Lupton), MR (molar refractivity; values scaled by 10^{-1}), and Charton's E_S values to characterize steric properties of substituents R at the functional amino group (all data from ref 59). To describe ring substituents, positions must be defined. The following conventions were used: (i) the functional amino group is always in position 1—additional amino groups are treated as substituents; (ii) if more than one amino group is present, we considered the functional group to be the one which has a substituent in an adjacent position (ortho substituent). Other conventions have also been tried but led to poorer results. (iii) If only one ortho substituent is present, this substituent is placed in position 2.

Biphenylamines, naphthylamines, and aminofluorenes were treated as substituted anilines. For the biphenylamines (see Figure 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 the ortho substituent and is fully parametrized with $\mathcal{I}$, $\mathcal{R}$, and MR. In the case of the

Table 1. Structures of Carcinogenic Compounds[a]

no.	ring	AnX	bridge X	R
1	N	3-C_4H_4-4		H
2	B	4-Ph-4-NH_2		H
3	F	3,4-Me_2		COMe
4	B	2-Cl,4-Ph-3-Cl,4-NH_2	CH_2	H
5	A	2-Me		H
6	B	4-C(=NH)-Ph-4-N$(Me)_2$	C=NH_2	Me_2
7	B	2-Ph		H
8	A	2,6-Cl_2,4-NH_2		H
9	A	2-NO_2,4-N$(C_2H_4OH)_2$		Me
10	B	4-CH_2-Ph-4-NH_2	CH_2	H
11	A	4-Cl		$CONMe_2$
12	B	4-O-Ph-4-NH_2	O	H
13	A	2-OEt,5-NHCOMe		H
14	F	3-Me,4-NEt		H
15	A	3-NO_2,4-OH		H
16	A	H		H
17	A	2-OMe		H
18	A	4-Cl		H
19	A	2-Cl,5-NH_2		H
20	A	2-NH_2,4-Cl		H
21	A	2-Me,4-OMe		H
22	A	2-OMe,5-Me		H
23	B	4-SO_2-Ph-4-NH_2	SO_2	H
24	A	2-OMe,5-NH_2		H
25	B	4-CH_2-Ph-4-N$(Me)_2$	CH_2	Me_2
26	B	4-CO-Ph-4-N$(Me)_2$	CO	Me_2
27	N	2-$C_3H_3C(NH_2)$-3		H
28	A	3-NO_2,4-OEt		COMe
29	A	2-OMe,5-NO_2		H
30	A	2-NO_2,4-NH_2		H
31	B	4-S-Ph-4-NH_2	S	H
32	A	2,6-$(NO_2)_2$,4-CF_3		$(nPr)_2$
33	A	2,4,5-Me_3		H
34	B	4-Ph		H
35	A	2-OH,4-NO_2		H
36	A	2-OH,5-NH_2		H
37	B	4-Ph		COMe
38	B	4-Ph-4-F		H
39	B	4-Ph-4-F		COMe
40	F	3,4-Me_2		$COCF_3$
41	B	2-Cl,4-Ph-3-Cl,4-NH_2		H
42	B	4-SO_2-Ph-4-NHCOMe	NH_2	COMe
43	A	4-OEt		COMe
44	A	4-F		Me,NO
45	A	H		Me,NO
46	A	2-NH_2		H
47	B	2-NH_2,4-Ph–3,4-$(NH_2)_2$		H
48	A	2,4,5,6-F_4,3-NH_2		H
49	A	2,4,6-Me_3		H
50	A	H		Me_2
51	A	4-Me		H
52	A	2-OH,5-NO_2		H
53	A	2,4,6-Cl_3		H
54	A	3-Me		H
55	B	2-OMe,4-Ph-3-OMe,4-NH_2		H
56	B	2-Me,4-Ph-3M_3,4-NH_2		H
57	A	2,5-Cl_2,3-COOH		H
58	B	2-Me,4-CH_2-Ph-3-Me,4-NH_2	CH_2	H

[a] A = anilines; B = biphenylamines; N = naphthylamines; F = aminofluorenes. Bridge: bridge between the phenyl rings in biphenylamines if present. AnX: ring substituent (all compounds described as substituted anilines; for definitions, see text). R = substituent at the functional amino group.

naphthylamines (see Figure 2), two situations are possible. They are treated as anilines substituted by $-C_4H_4-$ with an estimated MR of 0.8. This amount is equally distributed over the positions of substitution so that MR2 = MR3 = 0.8 in case 1 and MR3 = MR4 = 0.8 in case 2. To characterize electronic effects, the $\mathcal{I}$ and $\mathcal{R}$ values of CH=CH_2 are used for the two respective positions. If additional sub-

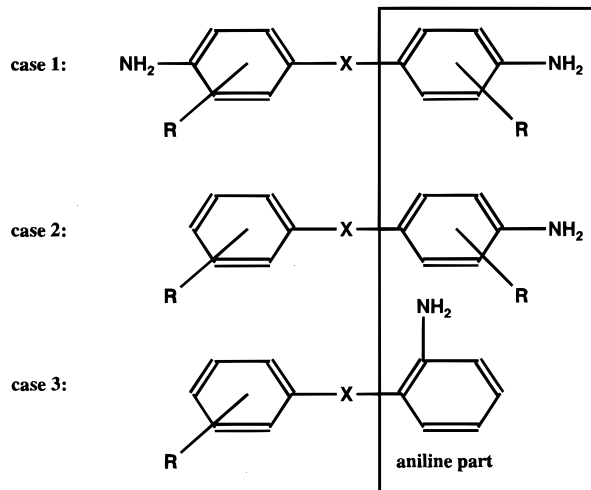


Figure 1. Treatment of biphenylamine.

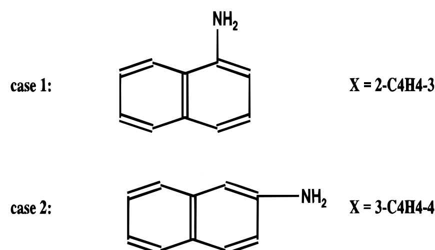


Figure 2. Treatment of naphthylamines.

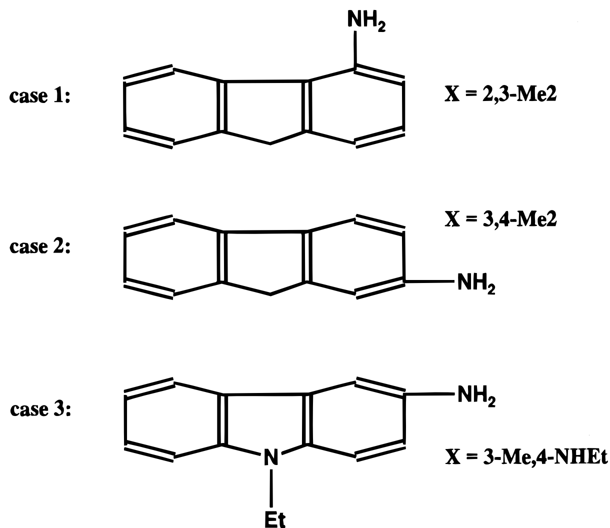


Figure 3. Treatment of aminofluorenes.

stituents occur, the MR values are correspondingly corrected; no values of electronic substituent constants are then available. For the aminofluorenes, finally, only steric effects (MR) could be parametrized following the scheme presented in Figure 3.

All chemical parameters appearing in the QSARs as well as carcinogenic potencies are summarized in Table 2. Table 3 summarizes the correlation matrix for the entirety of compounds in Table 2. There are no serious collinearities between the independent variables. However, multicollinearities occur which will be discussed in the context of the respective QSAR equations.

Some special features of the compounds are characterized by the following indicator variables: I(Bi)

Table 2. Chemical Descriptors and Carcinogenic Potencies (BRR, rats; BRM, mice) of the Carcinogenic Compounds in Table 1

									BRR		BRM	
no.	MR_3	$\Sigma MR_{2,6}$	$\Sigma \mathcal{F}_{2,6}$	$\Sigma \mathcal{R}_{2,6}$	E_S (R)	EHOMO	ELUMO	log P	observed	predicted[a]	observed	predicted[b]
1	0.8	0.2	0	0	0	−8.3724	−0.3671	2.27	0.37	0.31	0.59	0.97
2	0.1	0.2	0	0	0	−8.2341	0.0005	2.16	2.03	2.21	0.97	0.86
3	0.56	0.2	0	0	3	−8.4704	−0.3989	2.61	2.26	1.58	1.47	0.85
4	0.1	0.7	0.42	−0.19	0	−8.5595	0.0261	3.60	1.14	1.65		
5	0.1	0.66	0.42	−0.19	0	−8.5528	0.4059	1.73	0.39	0.13	−0.89	−0.45
6	0.1	0.2	0	0	2	−8.3483	−0.0286	3.02	1.39	1.45	0.63	0.59
7	0.1	2.64	0.12	−0.13	0	−8.5626	0.1005	2.95			−0.82	−1.11
8	0.1	1.2	0.84	−0.38	0	−8.2613	−0.089	1.52			−0.66	−0.32
9	0.1	0.84	0.65	0.13	1	−8.7212	−1.0742	0.34	−0.44	−0.36	0.47	0.14
10	0.1	0.2	0	0	0	−8.4373	0.3141	2.56	1.00	1.29	0.79	0.29
11	0.1	0.2	0	0	5	−8.7767	−0.1195	1.64	0.18	0.09		
12	0.1	0.2	0	0	0	−8.3253	0.1629	1.91	1.32	1.06	0.78	0.45
13	0.1	1.35	0.26	−0.5	0	−8.717	−0.1664	0.2			−1.03	−1.39
14	0.56	0.2	0	0	0	−8.0133	−0.2496	2.39	0.57	1.51	0.74	1.04
15	0.74	0.2	0	0	0	−9.039	−0.7543	0.93	−0.30	−0.15		
16	0.1	0.2	0	0	0	−8.6088	0.4153	1.26	−0.46	−0.04		
17	0.1	0.89	0.29	−0.55	0	−8.5707	0.2834	1.01	0.62	−0.13	−0.89	−1.19
18	0.1	0.2	0	0	0	−8.5906	0.1051	1.78			0.15	0.42
19	0.1	0.7	0.42	−0.19	0	−8.3803	0.1761	1.00	−0.34	−0.13	−0.94	−0.44
20	0.1	0.64	0.08	−0.74	0	−8.3819	0.111	1.00	−0.18	−0.13	−0.97	−0.3
21	0.1	0.66	0.01	−0.18	0	−8.5193	0.2607	1.48	−0.53	0.04		
22	0.1	0.89	0.29	−0.55	0	−8.5439	0.2527	1.48	0.15	0.04		
23	0.1	0.2	0	0	0	−9.0261	−0.3937	1.31	1.04	0.85		
24	0.1	0.89	0.29	−0.55	0	−8.1651	0.3817	0.23	−0.12	−0.4	−0.82	−0.83
25	0.1	0.2	0	0	2	−8.2965	0.3429	3.71	1.19	1.69	0.09	0.39
26	0.1	0.2	0	0	2	−8.6024	−0.2779	2.85	1.68	1.39	0.50	0.51
27	0.8	0.9			0	−7.9895	−0.3544	1.48	0.36	0.04	−0.01	−0.07
28	0.74	0.2	0	0	3	−9.5438	−1.1895	0.94			−1.01	−1.29
29	0.1	0.89	0.29	−0.55	0	−9.1996	−1.1264	0.96			−0.34	−0.3
30	0.1	0.84	0.65	0.13	0	−8.5709	−0.9914	0.43	−0.60	−0.33		
31	0.1	0.2	0	0	0	−8.5894	−0.1632	2.25	1.77	1.18	0.81	0.6
32	0.1	1.48	1.3	0.26	6	−10.2645	−1.997	4.25			0.01	0.4
33	0.1	0.66	0.01	−0.18	0	−8.361	0.3996	2.67	0.60	0.45	1.34	0.64
34	0.1	0.2	0	0	0	−8.4687	−0.0638	2.95			1.91	1.82
35	0.1	0.38	0.33	−0.7	0	−9.1425	−1.0677	0.93	0.14	−0.15		
36	0.1	0.38	0.33	−0.7	0	−8.0935	0.3428	0.20			−0.06	−0.15
37	0.1	0.2	0	0	3	−8.6681	−0.2232	2.58	2.25	2.35		
38	0.1	0.2	0	0	0	−8.5644	−0.3045	3.09			2.22	2.12
39	0.1	0.2	0	0	3	−8.9676	−0.6448	2.72	2.36	2.40		
40	0.56	0.2	0	0	5.4	−9.1706	−0.8891	3.73	2.23	1.97		
41	0.1	0.7	0.42	−0.19	0	−8.341	−0.3259	3.2				
42	0.1	0.2	0	0	3	−9.1575	−0.5674	0.57	0.78	0.59		
43	0.1	0.2	0	0	3	−8.6972	0.1085	0.99	−0.84	−0.14	−1.08	−1.03
44	0.1	0.2	0	0		−9.7645	−0.4175	1.83	2.78	2.91		
45	0.1	0.2	0	0		−9.2655	−0.183	1.69	2.98	2.86		
46	0.1	0.64	0.08	−0.74	0	−8.3321	0.4	0.48	−0.36	−0.31	−0.83	−0.75
47	0.1	0.64	0.08	−0.74	0	−8.1015	−0.009	0.60			−0.26	0.13
48	0.54	0.18	0.9	−0.78	0	−9.0632	−0.7425	1.04			0.32	−0.13
49	0.1	1.12	0.02	−0.36	0	−8.3657	0.4378	2.67	1.42	0.45	0.74	0.1
50	0.1	0.2	0	0	2	−8.4447	0.4541	1.84	−0.01	0.16		
51	0.1	0.2	0	0	0	−8.4698	0.393	1.73			0.11	0.17
52	0.1	0.38	0.33	−0.7	0	−9.0747	−1.026	0.93	−0.74	−0.15		
53	0.1	1.2	0.84	−0.38	0	−8.7178	−0.3013	2.82			−0.12	0.61
54	0.56	0.2	0	0	0	−8.5625	0.4141	1.73			−1.13	−0.49
55	0.1	0.89	0.29	−0.56	0	−8.2224	−0.1597	1.66	2.37	2.03		
56	0.1	0.66	0.01	−0.18	0	−8.1338	0.00173.1	2.53	2.54	0.87		
57	0.69	0.7	0.42	−0.19	0	−8.9882	−0.6344	2.00				
58	0.1	0.66	0.01	−0.18	0	−8.3558	0.3482	3.50	1.49	1.62		

[a] Predicted from eq 20. [b] Predicted from eq 13.

Table 3. Correlation Matrix (all compounds in Table 2 included)

	BRM	BRR	log P	EHOMO	ELUMO	$\Sigma MR_{2,6}$	MR_3	E_s (R)	$\Sigma \mathcal{F}_{2,6}$	$\Sigma \mathcal{R}_{2,6}$
BRM	1	0.717	0.496	0.208	0.012	−0.267	−0.095	−0.095	−0.274	0.369
BRR		1	0.603	0.234	0.126	−0.374	−0.026	0.256	−0.392	0.302
log P			1	0.008	0.018	0.014	0.0900	0.253	0.112	0.311
EHOMO				1	0.386	−0.128	0.133	−0.362	−0.245	−0.099
ELUMO					1	−0.028	−0.030	−0.105	−0.072	−0.048
$\Sigma MR_{2,6}$						1	−0.166	−0.196	−0.475	−0.329
MR_3							1	0.212	−0.114	0.154
E_s (R)								1	−0.063	0.398
$\Sigma \mathcal{F}_{2,6}$									1	−0.200
$\Sigma \mathcal{R}_{2,6}$										1

= 1 for biphenylamines; I(BiBr) = 1 for biphenylamines with a bridge between the phenyl rings; I(F) = 1 for aminofluorenes; $I(NO_2)$ = 1, if a NO_2 group is present; I(RNNO) = 1, if the amino group is substituted with (Me)NO; $I(monoNH_2)$ = 1, if only one amino group is present; $I(diNH_2)$ = 1, if more than one amino group is present

Compounds 22 and 57 behaved as outliers in the case of BRM and compounds 18, 41, and 47 in the case of BRR; these compounds were not included in the analyses.

2. Results

A. Modeling the Carcinogenic Potency in Mice

The following equation is obtained for BRM

$$\text{BRM} = 0.56 (\pm 0.18) \log P + 1.03 (\pm 0.74)\ \text{EHOMO} - 1.19 (\pm 0.58)\ \text{ELUMO} - 0.79 (\pm 0.37) \sum \text{MR}_{2,6} - 0.93 (\pm 0.90)\ \text{MR}_3 - 0.22 (\pm 0.19)\ E_S(\text{R}) + 8.51 (\pm 6.31) \quad (8)$$

$n = 37 \quad r = 0.845 \quad r^2 = 0.714 \quad s = 0.485 \quad F = 12.5 \quad p < 0.001$

The two most important variables in eq 8 are log P and $\sum MR_{2,6}$ accounting for 32.8% and 15.6% of the data variance, respectively. Carcinogenic potency increases with increasing hydrophobicity and increasing energy of the highest occupied molecular orbital and decreases with the energy of the lowest empty molecular orbital and with bulk in the ortho positions. Some steric effect is also evident for substitutions at position 3 (MR_3 term). The E_S(R) term, finally, indicates that substitution at the amino nitrogen is unfavorable for potency and that this effect becomes stronger as the bulkiness of the substituent(s) increases.

Though significant, eq 8 is of only moderate statistical quality, explaining only 71.4% of the data variance. An analysis in subgroups of compounds was considered a strategy to gain deeper insight. There are two possibilities to select subgroups: (i) according to the ring system (anilines and aminobiphenyles (there are not enough compounds to treat aminofluorenes or naphthylamines separately)), (ii) according to functionality (compounds with only one and compounds with more than one (substituted) amino group). The first possibility did not result in an improvement, but separating compounds with one and more than one amino group led to interesting relationships.

For the monoamines, eq 9 is obtained.

$$\text{BRM} = 0.74 (\pm 0.31) \log P + 2.60 (\pm 1.27)\ \text{EHOMO} - 1.65 (\pm 0.97)\ \text{ELUMO} - 0.85 (\pm 0.46) \sum \text{MR}_{2,6} - 1.46 (\pm 1.20)\ \text{MR}_3 + 21.77 (\pm 11.19) \quad (9)$$

$n = 17 \quad r = 0.936 \quad r^2 = 0.877 \quad s = 0.394 \quad F = 15.7 \quad p < 0.001$

The most important quantities again are log P and $\sum MR_{2,6}$, which alone already explain 61.6% of the data variance; the least important is MR_3, which contributes only 7%. In contrast to eq 8, no E_S(R) term appears in eq 9. The reason might be that in this subgroup there are only four N-substituted compounds with either R = H,COMe or R = $(n\text{-Pr})_2$ so that the variation of properties in R is very limited.

Equation 9 can be improved by adding an indicator variable accounting for the occurrence of NO_2 as a ring substituent.

$$\text{BRM} = 1.03 (\pm 0.31) \log P + 3.37 (\pm 1.11)\ \text{EHOMO} - 0.97 (\pm 0.89)\ \text{ELUMO} - 0.96 (\pm 0.36) \sum \text{MR}_{2,6} - 1.41 (\pm 0.92)\ \text{MR}_3 + 2.21 (\pm 0.89)\ I(\text{NO}_2) + 27.73 (\pm 9.48) \quad (10)$$

$n = 17 \quad r = 0.968 \quad r^2 = 0.937 \quad s = 0.281 \quad F = 25.0 \quad p < 0.001$

The meaning of the $I(NO_2)$ term is not clear. It could represent an electronic correction of the EHOMO/ELUMO terms but could also indicate a special (potency increasing) role of the NO_2 group.

EHOMO and ELUMO in eq 9 can be replaced by electronic substituent constants for substituents in the ortho positions. If Swain–Lupton constants are used, eq 9 transforms into eq 11.

$$\text{BRM} = 1.45 (\pm 0.36) \log P - 1.30 (\pm 0.78) \sum \mathcal{F}_{2,6} - 2.45 (\pm 1.37) \sum \mathcal{R}_{2,6} - 1.13 (\pm 0.47) \sum \text{MR}_{2,6} - 2.32 (\pm 0.76) \quad (11)$$

$n = 17 \quad r = 0.940 \quad r^2 = 0.883 \quad s = 0.384 \quad F = 22.6 \quad p < 0.001$

Even though the MR_3 term (which is of only marginal importance in eq 9) is no longer significant at the 95% level, eq 11 shows a better fit than eq 9. Obviously electron-releasing substituents in the ortho position enhance carcinogenic potency. It is unusual—but not without examples in the QSAR field—that electronic substituent effects occur for substituents in only one position. The reason probably is that the variation of electronic properties in the ortho position is greater than in the other positions (considerably higher variances of electronic substituent constants); in addition, some position dependence might also be operative.

For compounds with more than one free or substituted amino group, the following relationship is obtained.

$$\text{BRM} = 0.32 (\pm 0.22) \log P + 0.83 (\pm 0.82)\ \text{EHOMO} - 1.39 (\pm 0.47)\ \text{ELUMO} - 1.21 (\pm 0.58) \sum \text{MR}_{2,6} - 1.07 (\pm 1.06)\ \text{MR}_3 - 0.31 (\pm 0.29)\ E_S(\text{R}) + 7.41 (\pm 7.21) \quad (12)$$

$n = 20 \quad r = 0.923 \quad r^2 = 0.852 \quad s = 0.283 \quad F = 12.5 \quad p < 0.001$

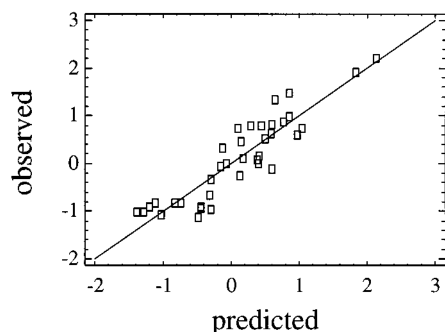


Figure 4. Plot of observed values of BRM against predicted values from eq 13.

The most important contribution in eq 12 comes from $\sum MR_{2,6}$ accounting for 50% of the data variance followed by ELUMO accounting for 13%; log P accounts for only 10% of the data variance. Replacement of EHOMO and ELUMO by electronic substituent constants is not possible in this case as the additional amino group(s) also is a potential reaction center.

The equations obtained for the subgroup of compounds with only one or more than one (substituted) amino group show a considerably better fit than the relationship obtained for the entirety of compounds (eq 8). Obviously the separation of the series into compounds with only one and compounds with more than one amino groups does make sense. Equation 12 tells the same story as the relationships obtained for the monoamines: bulk in positions adjacent to an amino group is unfavorable for carcinogenic potency, potency decreases with the energy of the lowest empty molecular orbital and increases with the energy of the highest occupied molecular orbital as well as with hydrophobicity, and bulky substituents at the nitrogen and in position 3 are unfavorable.

There are, however, also differences between eqs 9−11 on the one hand and eq 12 on the other hand: the coefficients of the log P- and of the EHOMO terms differ (the latter affords the difference of the intercepts). EHOMO shows a much smaller variance (0.25 in the monoamines and 0.07 in the compounds with more than one amino group) than log P (0.97 in the monoamines and 1.23 in the compounds with more than one amino group), so that the difference in the log P terms is of primary importance. Obviously the dependence of carcinogenic potency on hydrophobicity is much stronger in the case of the monoamines. This reflects itself not only in the higher regression coefficient of log P in eqs 9−11 as compared with eq 12, but also in the much higher part of the data variance explained by the log P term in these equations. Thus, the poor fit of eq 8 is mainly due to a different relationship of potency with log P for compounds with only one and with more than one (substituted) amino group, respectively. If this is accounted for by allowing different slopes for the log P term, compounds with one and more than one amino group can be treated in one equation (plot of regression results in Figure 4).

$$\begin{aligned} \text{BRM} = {} & 0.88\,(\pm 0.27)\log P\,I(\text{monoNH}_2) + 0.29\,(\pm 0.20)\log P\,I(\text{diNH}_2) + \\ & 1.38\,(\pm 0.76)\,\text{EHOMO} - 1.28\,(\pm 0.54)\,\text{ELUMO} - 1.06\,(\pm 0.34)\sum \text{MR}_{2,6} - 1.10\,(\pm 0.80)\,\text{MR}_3 - \\ & 0.20\,(\pm 0.16)\,E_S(\text{R}) + 0.75\,(\pm 0.75)\,I(\text{diNH}_2) + 11.16\,(\pm 6.68) \end{aligned} \quad (13)$$

$n = 37 \quad r = 0.907 \quad r^2 = 0.823 \quad s = 0.381 \quad F = 16.3 \quad p < 0.001$

Instead of EHOMO and ELUMO separately, the difference ELUMO − EHOMO representing the "hardness" can also be used.

$$\begin{aligned} \text{BRM} = {} & 0.88\,(\pm 0.26)\log P\,I(\text{monoNH}_2) + 0.30\,(\pm 0.19)\log P\,I(\text{diNH}_2) - \\ & 1.27\,(\pm 0.54)\,(\text{ELUMO} - \text{EHOMO}) - 1.08\,(\pm 0.31)\sum \text{MR}_{2,6} - 1.09\,(\pm 0.79)\,\text{MR}_3 - \\ & 0.22\,(\pm 0.12)\,E_S(\text{R}) + 0.77(\pm 0.75)\,I(\text{diNH}_2) + 10.25(\pm 4.85) \end{aligned} \quad (14)$$

$n = 37 \quad r = 0.907 \quad r^2 = 0.822 \quad s = 0.382 \quad F = 19.2 \quad p < 0.001$

The hardness is related to the barrier for the reaction of an electrophile with an aromatic compound. Debnath et al.[43] found a similar relationship for the mutagenic potency of aromatic amines.

The statistical fit of eqs 13 and 14 is much better than that for eq 8. They "explain" more than 80% of data variance, which is an acceptable result for the type of biological data considered. The log P terms in eqs 13 and 14 account for about 33% of the data variance, $\sum MR_{2,6}$ for about 15%, and the electronic terms (EHOMO and ELUMO) for about 20%. These terms reflect the most important effects. Thus, in agreement with eqs 9−12, eqs 13 and 14 show that the key factors for the carcinogenic potency in mice are as follows: (i) potency increases with increasing hydrophobicity (this effect seems to be more pronounced in compounds with only one amino group), (ii) increasing values of the energy of the highest occupied molecular orbital and decreasing values of the lowest empty molecular orbital enhance potency, (iii) potency decreases with increasing bulk in the positions adjacent to the amino group. In addition, bulk in position 3 and at the amino nitrogen inhibit carcinogenic activity. The $I(\text{diNH}_2)$ term in eqs 13 and 14 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 monoamines.

B. Modeling the Carcinogenic Potency in Rats

With the results for BRM in mind, analyses for BRR were also performed for the compound with only one and with more than one amino group separately as well as for the entirety of compounds. The following relationship is obtained for compounds with one amino group (plot of regression results Figure 11).

$$\text{BRR} = 0.46\ (\pm 0.35)\ \log P + 2.00\ (\pm 0.76)\ I(\text{Bi}) + 1.70\ (\pm 0.67)\ I(\text{F}) + 0.93\ (\pm 0.71) \sum \text{MR}_{2,6} + 2.99\ (\pm 0.63)\ I(\text{RNNO}) - 1.10\ (\pm 0.57) \quad (15)$$

$n = 20 \quad r = 0.969 \quad r^2 = 0.939 \quad s = 0.317 \quad F = 43.3 \quad p < 0.001$

The most important variable in eq 15 is log P, explaining 44% of the data variance, followed by I(RNNO)(36%), whereas only 12% of the data variance is explained by I(Bi) and I(F), indicating that these parameters are of less importance. The $\sum \text{MR}_{2,6}$ term, finally, is of only marginal importance (3.5%).

In agreement with the results obtained for BRM, potency increases with hydrophobicity as the key factor. There are, however, neither electronic terms nor effects of bulk in positions 3 or for the substitution at the nitrogen (significant factors for BRM). The $\sum \text{MR}_{2,6}$ term has a positive sign in contrast to the relationships obtained for BRM, indicating that substitution in the ortho position might support carcinogenic potency in rats. Since this term contributes very little to the explained data variance, this conclusion is only tentative. These differences in the QSARs for BRM and BRR highlight specific features that may account for the poor correspondence between the BRM and BRR quantities (see discussion).

The I(RNNO) term in eq 15 accounts for the unusually high potency of compounds 74 and 75, which possess the N(Me)N=O moiety instead of a simple amino group. Finally, the positive I(Bi) and I(F) terms lead to the conclusion that larger ring systems tend to be more carcinogenic.

For compounds with more than one (substituted) amino group, eq 16 is obtained.

$$\text{BRR} = 0.70\ (\pm 0.49)\ \text{EHOMO} + 2.43\ (\pm 0.38)\ I(\text{Bi}) - 0.77\ (\pm 0.40)\ I(\text{BiBr}) + 0.56\ (\pm 0.57)\ I(\text{F}) + 5.58\ (\pm 4.13) \quad (16)$$

$n = 21 \quad r = 0.972 \quad r^2 = 0.945 \quad s = 0.230 \quad F = 68.4 \quad p < 0.001$

In contrast to the result obtained for the monoamines, eq 16 seems to indicate that for compounds with more than one amino group hydrophobicity is not important for carcinogenic potency while electronic properties (expressed by EHOMO) come into play. However, even though eq 16 shows very good statistics, it is not the only possible result due to the internal data structure. For the compounds considered in eq 16, a multiple relationship between log P, EHOMO, I(Bi), and I(BiBr) exists with $r = 0.863$. It is, therefore, possible to replace EHOMO in eq 16 by log P:

$$\text{BRR} = 0.22\ (\pm 0.18)\ \log P + 2.14\ (\pm 0.47)\ I(\text{Bi}) - 1.08\ (\pm 0.40)\ I(\text{BiBr}) - 0.34\ (\pm 0.26) \quad (17)$$

$n = 21 \quad r = 0.960 \quad r^2 = 0.921 \quad s = 0.275 \quad F = 66.2 \quad p < 0.001$

According to eq 17, carcinogenic potency does not depend on the electronic properties but on hydrophobicity. Even though eq 16 shows a somewhat better fit than eq 17, there is no way to decide between these two possibilities. This illustrates how difficult it is to arrive at unambiguous results if multicollinearities occur. As will become clear, the relationship with log P is more likely to reflect the real situation (see below).

The I(Bi) and I(F) (the latter significant at only $P = 94\%$) terms in eq 16 indicate that biphenylamines and aminofluorenes are intrinsically more active than predicted by their log P or EHOMO values, respectively, which agrees with eq 15 for the monoamino compounds. In eq 17, the I(F) term is no longer significant. If a bridge between the phenyl rings in biphenylamines occurs, activity is decreased as follows from the I(BiBr) term in eqs 16 and 17.

If all compounds are to be considered, eq 15 is to be combined with either eq 16 or eq 17. Combination of eqs 15 and 16 yields

$$\text{BRR} = 0.37\ (\pm 0.15)\ \log P\ I(\text{monoNH}_2) + 0.61\ (\pm 0.43)\ \text{EHOMO} + 2.27\ (\pm 0.33)\ I(\text{Bi}) + 1.32\ (\pm 0.51)\ I(\text{F}) - 0.56\ (\pm 0.47)\ I(\text{BiBr}) + 3.23\ (\pm 0.70)\ I(\text{RNNO}) + 4.79\ (\pm 3.72) \quad (18)$$

$n = 41 \quad r = 0.947 \quad r^2 = 0.896 \quad s = 0.358 \quad F = 48.7 \quad p < 0.001$

Equation 19 is obtained if eqs 15 and 17 are combined.

$$\text{BRR} = 0.41(\pm 0.18)\ \log P\ I(\text{monoNH}_2) + 0.22(\pm 0.21)\ \log PT\ I(\text{diNH}_2) + 2.07\ (\pm 0.47)\ I(\text{Bi}) + 1.15\ (\pm 0.58)\ I(\text{F}) - 0.85\ (\pm 0.48)\ I(\text{BiBr}) + 2.67\ (\pm 0.61)\ I(\text{RNNO}) - 0.51\ (\pm 0.38) \quad (19)$$

$n = 41 \quad r = 0.942 \quad r^2 = 0.887 \quad s = 0.373 \quad F = 44.6 \quad p < 0.001$

According to eq 18, carcinogenic potency in rats increases with hydrophobicity (in this case, only of the monoamino compounds) and the energy of the highest occupied molecular orbital, and biphenylamines as well as aminofluorenes are intrinsically more active than the other compounds. The electronic term, though statistically significant, contributes only very little to the explained data variance. The presence of a bridge between the phenyl rings of biphenylamines is unfavorable (negative I(BiBr) term).The I(RNNO) term indicates the high values of BRR of the two compounds with R = NO. Equation 19, on the other hand, does not show any electronic effect: potency increases with hydrophobicity in such a way that this effect seems to be more pronounced in the monoamino compounds, analogous to eq 13 for the carcinogenic potency in mice. Again, as in the case of eqs 16 and 17, there is no way to decide whether eq 18 or 19 is to be preferred.

Both, eqs 18 and 19, seem to indicate that the hydrophobic effect is more pronounced in the compounds with only one amino group, which is in keeping with the different coefficients of the log P term in eqs 15 and 17. As, however, the confidence intervals of the regression coefficients associated with

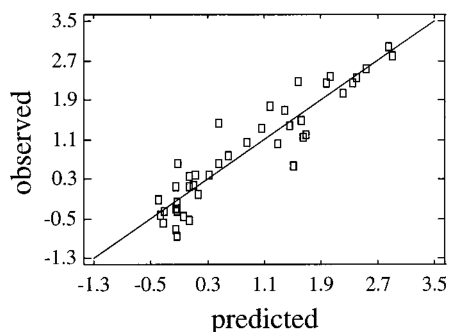


Figure 5. Plot of observed values of BRR against predicted values from eq 20.

log $P \times I(\text{monoNH}_2)$ and log $P \times I(\text{diNH}_2)$ in eq 19 overlap, it is possible to reunite these two variables into a common log P term (plot of regression results Figure 5).

$$\text{BRR} = 0.35\ (\pm 0.18) \log P + 1.93\ (\pm 0.48)\ I(\text{Bi}) + 1.15\ (\pm 0.60)\ I(\text{F}) - 1.06\ (\pm 0.53)\ I(\text{BiBr}) + 2.75\ (\pm 0.64)\ I(\text{RNNO}) - 0.48\ (\pm 0.30) \quad (20)$$

$$n = 41 \quad r = 0.933 \quad r^2 = 0.871 \quad s = 0.398 \quad F = 47.4 \quad p < 0.001$$

An electronic term cannot replace log P nor can such a term be added to eq 20. The conclusion is that hydrophobicity is the most important factor. Because of the data structure, however, it can neither be proven nor be ruled out that some effect of EHOMO and ELUMO does exist in addition to the hydrophobic effect. The key variable in eq 20 is log P which explains 42% of the data variance followed by I(RN-NO) with 20%. As already mentioned, the I(RNNO) term is afforded by compounds 74 and 75, which show very high potency. Elimination of these compounds does not affect eq 20 (the IRNNO term is, of course, no longer needed).

$$\text{BRR} = 0.35\ (\pm 0.18) \log P + 1.93\ (\pm 0.48)\ I(\text{Bi}) + 1.15\ (\pm 0.61)\ I(\text{F}) - 1.06\ (\pm 0.47)\ I(\text{BiBr}) - 0.48\ (\pm 0.30) \quad (21)$$

$$n = 39 \quad r = 0.918 \quad r^2 = 0.843 \quad s = 0.398 \quad F = 45.7 \quad p < 0.001$$

In this equation log P accounts for 53% of the data variance, clearly demonstrating that hydrophobicity is indeed the key factor for BRR. Again, no electronic term can be added to eq 21.

3. Discussion of the Carcinogenic Potency QSARs

Significant Hansch equations are obtained for both BRM and BRR. The key factor for the gradation of carcinogenic potency is hydrophobicity: both BRM and BRR increase with increasing log P. The influence of hydrophobicity is stronger for compounds with one amino group in comparison with compounds with more than one amino group. This effect is more pronounced for BRM than for BRR. In the case of BRM, the different dependence of potency on log P for compounds with one and with more than one amino group explicitly has to be taken into account when formulating QSARs for all compounds by allowing for different regression coefficients of log P for mono- and diamines. This is not necessary in the case of BRR. For BRM, in addition to hydrophobicity, electronic factors also play a role: potency increases with increasing energy of the highest occupied and with decreasing energy of the lowest empty molecular orbital. As these orbital energies are related to electronic substituent constants, they can be replaced by electronic substituent constants in the group of monoamines where the reaction center is clearly defined: potency is decreased by electron-attracting substituents in the positions adjacent to the functional amino group (ortho positions). For BRR, electronic effects are much less important; whether such effects are operating cannot unambiguously be decided because of multicollinearities in the data structure. Carcinogenic potency also depends on the type of ring system: aminobiphenyls (and, in the case of BRR, also fluorenamines) are intrinsically more active than anilines or naphthylamines. A bridge between the rings of the biphenyls decreases potency. Steric factors are involved in the case of BRM but are not important in the case of BRR. BRM strongly decreases with bulk in the positions adjacent to the functional amino group, and bulky substituents at the nitrogen and in position 3 also decrease potency. The latter effects are, however, not as important. Deeper insight into the role of substituents at the nitrogen would require that not only the size, but also the nature of such substituents is taken into consideration, which, however, is not possible with the limited variation in this position present in the series. In the case of BRR, R = (Me)NO strongly enhances potency (compounds with this substituent have no measured value for BRM).

The QSAR models for BRM and BRR show common features as well as specific differences. The most important common feature is the principal role of hydrophobicity, whereas the differences mainly reside in steric and electronic factors that are important for BRM but not for BRR. This situation reflects itself also in the correlation between BRM and BRR, which is statistically significant but “explains” only 58% of the data variance ($r = 0.764$).

Table 4 summarizes the situation for noncarcinogenic aromatic amines. If their carcinogenic potencies are predicted from the QSARs, they all appear as weak carcinogens (Table 5). In other words, the Hansch equations permit the recognition of strong carcinogens and the estimation of the gradation of potency within active compounds but cannot separate weak carcinogens from inactive compounds. This is not an uncommon situation with Hansch equations as the properties connected with the gradation of potency need not be identical with those discriminating between active and inactive compounds: there can be many reasons outside the parameter space related to potency which can render a compound inactive. What is significant, however, is that the noncarcinogens are placed in the region of very low to low potency with the only exception being compound 70 which shows high values for both BRM and

Table 4. Structures of Noncarcinogenic Compounds[a]

no.	ring	AnX	bridge X	R
59	A	3-Cl		COOiPr
60	A	2-Me,3-NH_2		H
61	A	2-COOH		H
62	A	4-$COCH_2Cl$		COMe
63	A	2-Cl,4-NH_2		H
64	A	2,4-OMe_2		H
65	A	2,6-Me_2,4-OCONMe		Me_2
66	N	2-C_4H_4-3		$C_2H_4NH_2$
67	A	2-COOH,5-NO_2		H
68	A	2-NH_2,4-NO_2		H
69	A	4-NH_2		H
70	B	4-NH−Ph-4-NH_2	NH	H
71	A	H		$CSNH_2$
72	A	2-Me,4-NH_2		H
73	A	2-Cl,4-Me		H

[a] A = anilines; B = biphenylamines; N = naphthylamines; F = aminofluorenes. Bridge: bridge between the phenyl rings in biphenylamines if present. AnX: ring substituent (all compounds described as substituted anilines; for definitions, see text). R = substituent at the functional amino group.

BRR. The situation is much better than in the case of the mutagenic potency of aromatic amines where the predicted potency for most of the nonmutagenic amines spans the entire range of values up to very high potency.[45] To separate active from inactive compounds, classification methods such as discriminant analysis or SIMCA have to be applied. This will be the subject of further investigations.

VIII. The QSARs of Aromatic Amines in Perspective

Despite the very complex nature of the processes involved, the QSARs obtained by the various authors are generally in good agreement with already published observations pertaining to mechanisms of carcinogenicity and mutagenicity, respectively, of aromatic amines. In general, aromatic amines require metabolic activation to yield the ultimate carcinogen or mutagen, and the principal pathway(s) of this bioactivation involves formation of a hydroxylamine which decomposes to a reactive nitrenium ion intermediate[5,16] (see Scheme 1). This bioactivation mechanism for aromatic amines is believed to be the same in carcinogenesis and mutagenesis, and for this reason, the present results can be compared with QSARs for the mutagenic potency of aromatic amines.

In the most mechanistically oriented QSAR analyses, the toxic activity of the amines was demonstrated to correlate with the ease of formation of the *N*-hydroxylamine,[55] with the stability of the nitrenium ion,[41,52] and with the ease of formation of epoxides on the aromatic ring.[55] Loew[55] also found that the ease of formation of phenols (a detoxifying pathway) is actually negatively correlated with the carcinogenic activity.

All studies[40,43] considering lipophilicity, including the present investigation, confirmed its central role. In particular, Debnath et al.[43] found that lipophilicity is the main determinant of mutagenic potency, with a linear increase in potency observed with increasing log *P*. Electronic effects were of secondary importance, with potency increasing with EHOMO and decreasing with ELUMO. 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 analogues. This finding is in keeping with the result obtained for the carcinogenic potency in a mouse in the present investigation, i.e., that potency of the monoamines shows a much stronger dependence on log *P* than potency of the diamines (our analysis of carcinogenic potency). The majority of compounds in the paper by Debnath et al.[43] for which the QSARs hold consists of monoamines, and practically all of the "too hydrophilic" compounds are diamines. For the latter, an inverse relationship with log *P* was suggested which, however, is not well supported by the data. It would be worth trying to find out what the result for mutagenic potency is if mono- and diamines are treated separately.

The HOMO and LUMO energies were found to have a role both for the mutagenic activity in Salmonella[43,46,48,49,52] and for the carcinogenic potency in a mouse (our analysis). The role of the HOMO energy can be easily rationalized in terms of the propensity of the toxic amines to form the intermediate metabolite *N*-hydroxylamine. The role of the LUMO energy is quite puzzling. Debnath et al.[43]

Table 5. Chemical Descriptors and Predicted Carcinogenic Potencies (BRR, rats; BRM, mice) for the Noncarcinogens in Table 4

no.	MR_3	$\Sigma MR_{2,6}$	$\Sigma /_{2,6}$	$\Sigma \mathcal{R}_{2,6}$	E_S (R)	EHOMO	ELUMO	log *P*	predicted carcinogenic potency	
									BRR[a]	BRM[b]
59	0.6	0.2	0	0	4	−9.162	−0.1543	2.79	0.50	−0.50
60	0.54	0.66	0.01	−0.18	0	−8.3607	0.4333	0.95	−0.15	−1.20
61	0.1	0.79	0.34	0.11	0	−8.8284	−0.455	0.96	−0.14	−0.54
62	0.1	0.2	0	0	3	−9.3254	−0.8105	0.80	−0.20	−0.89
63	0.1	0.7	0.42	−0.19	0	−8.1632	0.1491	1.00	−0.13	−0.11
64	0.1	0.89	0.29	−0.56	0	−8.3083	0.2602	0.76	−0.22	−1.02
65	0.1	1.12	0.02	−0.36	2	−8.9385	0.0892	2.25	0.31	−1.01
66	0.8	0.9	0.13	−0.17	3	−8.5284	−0.4132	1.69	0.11	−1.03
67	0.1	0.79	0.34	0.11	0	−9.4286	−1.5938	0.92	−0.16	0.04
68	0.1	0.64	0.08	−0.74	0	−9.0498	−1.0257	0.43	−0.33	0.07
69	0.1	0.2	0	0	0	−8.0719	0.411	0.48	−0.31	0.07
70	0.1	0.2	0	0	0	−8.046	0.119	2.38	1.23	1.67
71	0.1	0.2	0	0	3.2	−8.6991	−0.735	1.86	0.17	0.77
72	0.1	0.66	0.01	−0.18	0	−8.0712	0.4025	0.95	−0.15	−0.27
73	0.1	0.7	0.42	−0.19	0	−8.5409	0.1441	2.25	0.31	0.31

[a] From eq 20. [b] From eq 13.

discussed several possibilities. One is that the two terms LUMO and HOMO could be linked together through the concept of "hardness" (η = (LUMO−HOMO)/2) as a measure of chemical reactivity. Another hypothesis is that LUMO energy accounts for the reduction of the nitro group present, together with the amino group, in a number of their set of amines. However, Zhang et al.[48] and Hatch et al.[52] found a LUMO term in data sets without nitroarenes. Another explanation could rely on a very recent finding by King at al.[60] They found a new enzymatic mechanism of carcinogen detoxification: a microsomal NADH-dependent reductase that rapidly converts the *N*-hydroxyarylamine back to the parent compound. In this case, a low LUMO energy cold favor the detoxification. However, the LUMO energy of the metabolite is not necessarily coincident with that of the parent amine; thus, the entire matter needs further clarification.

A number of qualitative rules have been proposed regarding the properties of aromatic amines that affect carcinogenic potency.[5] Bulky substituents at the nitrogen of the amino group generally inhibit bioactivation. This is in keeping with the E_S(R) contribution found by us for N-substituents in the functional amino group in a mouse and with the findings of Trieff et al. (inhibiting effect of the acetylation of the amino group).[40] A general rule states that carcinogenic potency decreases with steric bulk in the ortho position[5] (see also Trieff et al.[40] and Benigni et al.[46]). This rule is consistent with the negative $\sum MR_{2,6}$ term observed by us for the carcinogenic potency in a mouse. A mechanistic rationale for these observations is that steric bulk prevents enzymatic access to the nitrogen and formation of the reactive intermediate.

As was also found in our analysis of the carcinogenic potency, ring substituents have been proposed to exert electronic and steric effects. In fact, according to Vracko,[56] substitution of a chloro group or methyl or methoxy group ortho to the amino group is considered to often enhance potency. This statement requires clarification. Our present QSARs for the carcinogenic potency show that ortho substituents can operate through at least three effects: direct steric, electronic, and hydrophobic. Thus, potency can increase or decrease depending on the nature of an ortho substituent. Electron-donating ortho substituents would stabilize a positive nitrenium ion, whereas electron-withdrawing substituents would destabilize such an intermediate.

It is more difficult to put into context work based on topological and substructural parameters, as the results are difficult to interpret and do not lend themselves easily to comparison and generalization. However, a finding common to various authors was the correlation between activity and the number of aromatic rings,[27,49,52,54] which has been interpreted in different ways: (a) indicator for the planar systems apt to induce frameshift mutations in TA98 Salmonella strain;[49] (b) indicator for the hydrophobicity of polycyclic and condensed aromatic rings;[27] (c) indicator for the presence of extended conjugated systems that favor the formation of reactive intermediates.[5] Obviously the simple empirical correlation between the number of rings and toxic activity cannot suggest which (or what combination) of the above hypotheses is correct. A thoughtful insight into this issue was provided by Debnath et al.;[43] they showed that besides log *P*, an additional contribution to the mutagenic potency in TA98 was given by the presence of three or more fused rings. This effect was absent in TA100 strain and was related to the specificity of TA98 for frameshift mutations.

The latter result is a brilliant demonstration of the importance of using a common language for QSAR modeling. A common language is the only approach that can tell us if and to what extent QSAR models (and the underlying chemical biological interactions) are similar. In the present review, the models based on the Hansch approach and on the use of the operational definition of hydrophobicity in terms of log *P* have permitted us to conclude that the toxic activity of the aromatic amines follows similar mechanisms in the Ames test and carcinogenicity in mouse and rat. The evidence for this similarity reassures us about the use of bacterial and animal data for risk assessment.

A more general perspective can be gained by comparing the QSARs for the aromatic amines with those for other classes of mutagens and carcinogens. Several reviews have appeared on this subject,[61−69] and we will not duplicate them here. Briefly, the QSARs for the aromatic amines confirm the central role of hydrophobicity in the activity of mutagens and carcinogens, especially for those requiring metabolic activation (hence interaction with the metabolism enzymes).[63,64] Debnath et al.[64] also showed how consistently the coefficient of log *P* was close to the same value (around 1.0) in different bacterial systems, thus pointing to a specific characteristic of these systems. They also reviewed examples of direct-acting mutagens (not requiring metabolic activation): most of the models did not contain a term for hydrophobicity.

The successful modeling of in vivo data (carcinogenic potency) provided in this paper deserves a further comment. Whereas experimental results from in vitro systems (bacteria or cultured cells) are normally considered as reliable enough for building models, the quality of in vivo data is often considered as one of the major obstacles to sound QSAR analyses. In particular, even what end point should be used for carcinogenicity is questioned. The long list of claimed difficulties also includes the several steps and competing reactions that "compose" an in vivo effect. However, this review showed that not only the bacterial data lend themselves to successful QSAR modeling, but also the in vivo data were translated into models both statistically reliable and scientifically informative. To understand this result, some comments on the end point selected by us (TD_{50}) are necessary. A rodent carcinogenicity experiment provides a large amount of information, namely, (a) yes/no activity of the compound, (b) potency, (c) target organs. Yes/no activity is the most important information, and it is highly predictive of the effect in humans.[3,70−72] On the other hand, the target organs

vary very much from species to species and also depend on age, sex, route of administration, etc.;[72,73] thus, this information is not suitable for extrapolating the risk to humans. A commonly used measure for carcinogenic potency is TD_{50} (daily dose that halves the probability of remaining tumorless with respect to the control animals).[58] Since several tumor types can be induced in the same experiment, at very different rates, a TD_{50} for each target organ can be calculated. These TD_{50}s have a wide range of variability, due to the myriad of organ and tissue differences in terms of partitioning, bioavailability, toxifying/detoxifying processes. On the contrary, the (harmonic) average TD_{50} values have been demonstrated to be correlated between males and females in both rat and mouse; moreover, they are correlated between rat and mouse.[72] In addition, it has been demonstrated that the recognized human carcinogens have the same ranking of potency in humans and rodents.[74,75] This is an indirect demonstration that the average TD_{50} values point to the "intrinsic" carcinogenic potential of the chemicals, independently from local organ and tissue specific effects. From an operational point of view, our results demonstrated that the average TD_{50}s are a reliable basis for modeling the carcinogenic effects, although the experimental data originated from different laboratories. On the contrary, we were not successful in modeling the carcinogenic potency at the level of the individual organs (unpublished results). Overall, the results presented in this review demonstrate that not only the in vitro data can be modeled, but also in vivo effects lend themselves to QSAR analysis.

To date, most efforts to develop carcinogenicity prediction models have considered large, structurally diverse data sets and focused on qualitative predictions of activity class, i.e., positive or negative predictions independent of potency. Although some rule-based expert approaches, such as OncoLogic system,[76] have attempted semiquantitative estimations of carcinogenic potency (e.g., low, moderate, high) based on mechanistic considerations, these efforts have not attempted to incorporate any type of quantitative modeling or QSAR analysis. We hope that the QSAR results for the aromatic amines from our laboratory and from other authors have demonstrated the feasibility and value of such analysis and will stimulate other investigators to consider applying QSAR methods to predicting relative potency within other well-defined classes of chemical carcinogens. As more QSARs are discovered, it increases our ability to apply comparative QSAR analysis to uncover meaningful biofunctional associations across chemical classes and, thus, to broaden and deepen our understanding of the structural basis for chemical carcinogenicity. Ultimately, the goal would provide the scientific community with both classification models as well as an array of refined QSAR models for the prediction of the carcinogenicity of untested compounds.

VIII. Acknowledgments

The continued interest and unvaluable comments of Ann M. Richard are gratefully aknowledged.

IX. References

(1) Benigni, R. *Mutat. Res.* **1997**, *387*, 35−45.
(2) Benigni, R. In *QSARs for endocrine disruption, pollution prevention, toxicity screening and persistence prediction*; Walker, J. D., Ed.; SETAC: Washington, D.C., 2000.
(3) Fung, V. A.; Barrett, J. C.; Huff, J. *Environ. Health Perspect.* **1995**, *103*, 680−683.
(4) Cronin, M. T. D.; Dearden, J. C. *Quant. Struct.−Act. Relat.* **1995**, *14*, 1−7.
(5) Lai, D. Y.; Woo, Y. T.; Argus, M. F.; Arcos, J. C. In *Designing safer chemicals. Green chemistry for pollution prevention*; DeVito, S. C., Garrett, R. L., Eds.; American Chemical Society: Washington, D.C., 1996; pp 1−13.
(6) Sugimura, T. *Mutat. Res.* **1997**, *376*, 211−219.
(7) Skog, K. I.; Johansson, M. A.; Jagerstad, M. I. *Food Chem. Toxicol.* **1998**, *36*, 879−896.
(8) Vineis, P.; Pirastu, R. *Cancer, Causes Control* **1997**, *8*, 346−355.
(9) Dolin, P. J. *Br. J. Cancer* **1992**, *65*, 476−478.
(10) Rubino, G. F.; Scansetti, G.; Piolatto, G.; Pira, E. *Environ. Res.* **1982**, *27*, 241−254.
(11) Stasik, M. J. *Int. Arch. Occup. Environ. Health* **1988**, *60*, 21−24.
(12) Bartsch, H.; Malaveille, C.; Friesen, M.; Kadlubar, F. F.; Vineis, P. *Eur. J. Cancer* **1993**, *29A*, 1199−1207.
(13) Talaska, G.; Schamer, M.; Skipper, P. *Cancer Epidemiol. Biomarkers Prev.* **1991**, *1*, 61−66.
(14) Huff, J. *Environ. Health Perspect.* **1993**, *100*, 201−210.
(15) Kadlubar, F. F.; Hammons, G. J. In *Mammalian cytochromes P-450*; Guengerich, F. P., Ed.; CRC Press: Boca Raton, FL, 1987; Vol. II, pp 81−130.
(16) Beland, F. A.; Kadlubar, F. F. In *Chemical carcinogenesis and mutagenesis I*; Cooper, C. S., Grover, P. L., Eds.; Springer-Verlag: Berlin, 1990; pp 265−325.
(17) Kadlubar, F. F.; Beland, F. A. In *Polyciclic hydrocarbons and carcinogenesis*; Harvey, R. G., Ed.; American Chemical Society: Washington, D.C., 1985; pp 341−370.
(18) Radomski, J. L. *Annu. Rev. Pharmacol. Toxicol.* **1979**, *19*, 129−157.
(19) Purchase, I. F. H.; Kalinowski, A. E.; Ishmael, J.; Wilson, J.; Gore, C. W.; Chart, I. S. *Br. J. Cancer* **1981**, *44*, 892−901.
(20) Hammons, G. J.; Guengerich, F. P.; Weis, C. C.; Beland, F. A.; Kadlubar, F. F. *Cancer Res.* **1985**, *45*, 3578−3585.
(21) Belman, S.; Troll, W.; Teebor, G.; Mukai, F. *Cancer Res.* **1968**, *28*, 535−542.
(22) Dooley, K. L.; Beland, F. A.; Bucci, T. J.; Kadlubar, F. F. *Cancer Res.* **1984**, *44*, 1172−1177.
(23) Kadlubar, F. F.; Miller, J. A.; Miller, E. C. *Cancer Res.* **1978**, *38*, 3628−3638.
(24) Murofushi, Y.; Hashimoto, Y.; Shudo, K.; Okamoto, T. *Chem. Pharm. Bull.* **1981**, *29*, 2730−2732.
(25) Hein, D. W. *Biochim. Biophys. Acta* **1988**, *948*, 37−66.
(26) King, C. M.; Land, S. J.; Jones, R. F.; Debiec-Rychter, M.; Lee, M. S.; Wang, C. Y. *Mutat. Res.* **1997**, *376*, 123−128.
(27) Maran, U.; Karelson, M.; Katritzky, A. R. *Quant. Struct.−Act. Relat.* **1999**, *18*, 3−10.
(28) Arcos, J. C.; Argus, M. F.*Aromatic amines and azo compounds*; Academic Press: New York, 1974.
(29) Ashby, J.; Styles, J. A.; Patton, D. *Br. J. Cancer* **1978**, *38*, 34−50.
(30) Milman, H. A.; Peterson, C. *Environ. Health Perspect.* **1984**, *56*, 261−273.
(31) Esancy, J. F.; Freeman, H. S.; Claxton, L. D. *Mutat. Res.* **1990**, *238*, 23−38.
(32) McCann, J.; Choi, E.; Yamasaki, E.; Ames, B. N. *Proc. Natl. Acad. Sci. U.S.A.* **1975**, *75*, 5135−5139.
(33) McCann, J.; Ames, B. N. *Proc. Natl. Acad. Sci.* **1976**, *73*, 950−954.
(34) Tennant, R. W.; Margolin, B. H.; Shelby, M. D.; Zeiger, E.; Spalding, J. W.; Caspary, W.; Resnick, M. A.; Anderson, B.; Minor, R.; Haseman, J. K.; Stasiewicz, S. *Science* **1987**, *236*, 933−941.
(35) Zeiger, E.; Haseman, J. K.; Shelby, M. D.; Margolin, B. H.; Tennant, R. W. *Environ. Mol. Mutagen.* **1990**, *16,* 1−14.
(36) Zeiger, E. *Cancer Res.* **1987**, *47*, 1287−1296.
(37) Fetterman, B. A.; Soo Kim, B.; Margolin, B. H.; Schidcrout, J. S.; Smith, M. G.; Wagner, S. M.; Zeiger, E. *Environ. Mol. Mutagen.* **1997**, *29*, 312−322.
(38) Chung, K.; Kirkovsky, L.; Kirkovsky, A.; Purcell, W. P. *Mutat. Res.* **1997**, *387*, 1−16.
(39) Colvin, M. E.; Hatch, F. T.; Felton, J. S. *Mutat. Res.* **1998**, *400*, 479−492.
(40) Trieff, N. M.; Biagi, G. L.; Sadagopa Ramanujam, V. M.; Connor, T. H.; Cantelli-Forti, G.; Guerra, M. C.; Bunce, H., III; Legator, M. S. *Mol. Toxicol.* **1989**, *2*, 53−65.
(41) Ford, G. P.; Griffin, G. R. *Chem. Biol. Interact.* **1992**, *81*, 19−33.

(42) Ford, G. P.; Herman, P. S. *Chem. Biol. Interact.* **1992**, *81*, 1−18.
(43) Debnath, A. K.; Debnath, G.; Shusterman, A. J.; Hansch, C. *Environ. Mol. Mutagen.* **1992**, *19*, 37−52.
(44) Debnath, A. K.; Lopez de Compadre, R. L.; Shusterman, A. J.; Hansch, C. *Environ. Mol. Mutagen.* **1992**, *19*, 53−70.
(45) Benigni, R.; Andreoli, C.; Giuliani, A. *Environ. Mol. Mutagen.* **1994**, *24*, 208−219.
(46) Benigni, R.; Passerini, L.; Gallo, G.; Giorgi, F.; Cotta-Ramusino, M. *Environ. Mol. Mutagen.* **1998**, *32*, 75−83.
(47) Klopman, G.; Frierson, M. R.; Rosenkranz, H. S. *Environ. Mutagen.* **1985**, *7*, 625−653.
(48) Zhang, Y. P.; Klopman, G.; Rosenkranz, H. S. *Environ. Mol. Mutagen.* **1993**, *21*, 100−115.
(49) Lewis, D. F. V.; Ioannides, C.; Walker, R.; Parke, D. V. *Food Addit. Contam.* **1995**, *12*, 715−723.
(50) Basak, S. C.; Grunwald, G. D. *Chemosphere* **1995**, *31*, 2529−2546.
(51) Hatch, F. T.; Knize, M. G.; Felton, J. S. *Environ. Mol. Mutagen.* **1991**, *17*, 4−19.
(52) Hatch, F. T.; Colvin, M. E.; Seidl, E. T. *Environ. Mol. Mutagen.* **1996**, *27*, 314−330.
(53) Hatch, F. T.; Colvin, M. E. *Mutat. Res.* **1997**, *376*, 87−96.
(54) Yuta, K.; Jurs, P. C. *J. Med. Chem.* **1981**, *24*, 241−251.
(55) Loew, G. H.; Poulsen, M.; Kirkjian, E.; Ferrel, J.; Sudhindra, B. S.; Rebagliati, M. *Environ. Health Perspect.* **1985**, *61*, 69−96.
(56) Vracko, M. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 1037−1043.
(57) Gini, G.; Lorenzini, M.; Benfenati, E.; Grasso, P.; Bruschi, M. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 1076−1080.
(58) Gold, L. S.; Slone, T. H.; Manley, N. B.; Garfinkel, G. B.; Hudes, E. S.; Rohrbach, L.; Ames, B. N. *Environ. Health Perspect.* **1991**, *96*, 11−15.
(59) Hansch, C.; Leo, A.; Hoekman, D. *Exploring QSAR. Hydrophobic, electronic and steric constants*; American Chemical Society: Washington, D.C., 1995.
(60) King, R. S.; Teitel, C. H.; Shaddock, J. G.; Casciano, D. A.; Kadlubar, F. F. *Cancer Lett.* **1999**, *143* (2), 167−171.
(61) Frierson, M. R.; Klopman, G.; Rosenkranz, H. S. *Environ. Mutagen.* **1986**, *8*, 283−327.
(62) Benigni, R.; Andreoli, C.; Giuliani, A. *Mutat. Res.* **1989**, *221*, 197−216.
(63) Hansch, C. *Sci. Tot. Environ.* **1991**, *109/110*, 17−29.
(64) Debnath, A. K.; Shusterman, A.; Lopez de Compadre, R. L.; Hansch, C. *Mutat. Res.* **1994**, *305*, 63−72.
(65) Richard, A. M. *Mutat. Res.* **1994**, *305*, 73−97.
(66) Cronin, M. T. D.; Dearden, J. C. *Quant. Struct.−Act. Relat.* **1995**, *14*, 518−523.
(67) Benigni, R.; Giuliani, A. *Med. Res. Rev.* **1996**, *16*, 267−284.
(68) Benigni, R.; Giuliani, A. In *Computer-Assisted lead finding and optimization*; van der Waterbeemd, H., Testa, B., Folkers, G., Eds.; Wiley-VCH: Basel, 1997.
(69) Richard, A. M. *Mutat. Res.* **1998**, *400*, 493−507.
(70) Tomatis, L.; Huff, J.; Hertz-Picciotto, I.; Sandler, D. P.; Bucher, J.; Boffetta, P.; Axelson, O.; Blair, A.; Taylor, J.; Stayner, L.; Barrett, J. C. *Carcinogenesis* **1997**, *18*, 97−105.
(71) Huff, J. In *Carcinogenicity. Testing, predicting, and interpreting chemical effects*; Kitchin, K. T., Ed.; Marcel Dekker: New York, 1999; pp 21−123.
(72) Benigni, R.; Giuliani, A. *Environ. Carcinog. Ecotoxicol. Rev.* **1999**, *C17*, 45−67.
(73) Bucci, T. J. *Toxicol. Pathol.* **1985**, *13*, 105−109.
(74) Goodman, G.; Wilson, R. *Regulat. Pharmacol. Toxicol.* **1991**, *14*, 118−146.
(75) Allen, B. C.; Crump, K. S.; Shipp, A. M. *Risk Anal.* **1988**, *8*, 531−544.
(76) Woo, Y. T.; Lai, D. Y.; Arcos, J. C.; Argus, M. F.; Cimino, M. C.; DeVito, S.; Keifer, L. *Environ. Carcinog. Ecotoxicol. Rev.* **1997**, *C15*, 139−160.

CR9901079