ELSEVIER

Dyes and Pigments 55 (2002) 35–52

DYES and PIGMENTS

www.elsevier.com/locate/dyepig

Mutagenicity of aminoazobenzene dyes and related structures: a QSAR/QPAR investigation

Ashish Garg, Krishna L. Bhat, Charles W. Bock*

Department of Chemistry and Biochemistry, School of Science and Health and School of Textiles and Materials Technology, Philadelphia University, School House Lane and Henry Avenue, Philadelphia, PA 19144, USA

Received 17 May 2002; received in revised form 10 June 2002; accepted 27 June 2002

Abstract

Quantitative structure–activity/property–activity relationships (QSAR/QPARs) are developed that correlate the observed mutagenic activity of 43 aminoazobenzene derivatives with a variety of molecular descriptors calculated using quantum-chemical semiempirical methodology. Models based on multilinear regression techniques and artificial neural networks are presented that account for more than 80% of the variation in the reported relative mutagenicity of these compounds.

Keywords: Aminoazobenzene dyes; Structure–activity relationships; Mutagenicity; Artificial neural networks; CODESSA

1. Introduction

As early as 1937, Kinosita [1] reported that *N,N*-dimethyl-4-aminoazobenzene (DAB), an azo dye once used as a hair colorant, produces hepatomas and cholangiomas in the liver of rats when included as part of their diet over a period of several months [1]. Since then, it has been conclusively demonstrated that a variety of 4-aminoazobenzene (AAB), *N*-methyl-4-aminoazobenzene (MAB) and *N,N*-dimethyl-4-aminoazobenzene (DAB) derivatives are mutagenic and/or carcinogenic [2–5]. Interestingly, the potency of these compounds is strongly dependent on the nature and position of substituents with respect to both the aromatic rings and the amino nitrogen atom. For example, 3-methoxy-4-aminoazobenzene (3-OMe-AAB) is a potent hepatocarcinogen in rats and a strong mutagen in *Escherichia coli* and *Salmonella typhimurium*, whereas 2-OMe-AAB is apparently a non-carcinogen and an extremely weak mutagen under similar conditions [6]. The biochemical mechanisms responsible for such divergent behavior are not yet fully understood [7–11].

The purpose of this paper is to derive quantitative structure-activity/property-activity relationships (QSAR/QPARs) for the mutagenicity (rev/nmol) of a variety of AAB, MAB, and DAB derivatives in the *S. typhimurium* TA98 bacterial strain with S9 activation (TA98 + S9); this particular bacterial strain is well known to detect frameshift

* Corresponding author. Tel.: +1-215-951-2876; fax: +1-215-951-6812.

E-mail address: bockc@philau.edu (C.W. Bock).

0143-7208/02/$ - see front matter
PII: S0143-7208(02)00070-0

mutagens. Since these azo compounds require metabolic activation prior to reaction with cellular macromolecules [9], we have included some of their metabolites, e.g. N–OH–AAB, in our analyses, although none of their reductive-cleavage products have been utilized.

An array of QSAR/QPAR methods have been developed to analyze genotoxicity data and these methods have been applied with considerable success to various classes of molecules [12–22]. In particular, the computer-assisted substructural-based methods of Enslein and Borgstedt [19], Rosenkranz and Klopman (CASE) [20], and Rose and Juys (ADAPT) [21], have been used for some aminoazo dyes. For example, the CASE system predicted that the presence of a 1-amino-2-napthol residue in azo dyes is mutagenic, but that sulfonation at the C-3, C-4, or C-6 position reduces this mutagenicity [20]. Claxton et al. [22] trained the ADAPT system using a database of 25 active and 13 inactive azo dyes to correctly predict the activity/inactivity of 37 out of 44 dyes that were not included in the database; interestingly, only five simple descriptors (number of oxygen atoms, number of nitrogen atoms, and three geometrical parameters involving the principal axis) were required in their analysis. Our focus is on developing QSAR/QPARs for aminoazobenzene derivatives based on their relative mutagenic activity in specific bacterial strains, not just on whether a compound is active or inactive. We envision that this analysis will provide additional insight into those factors that contribute to mutagenic activity in this class of compounds.

We have restricted this initial study to derivatives that contain only a single azo linkage between two aromatic rings and for which there is an amino or substituted amino group at the 4-position; compounds that contain sodium or potassium salts of sulfonic acid groups as substituents were excluded from this investigation. For each of the compounds in this study a large set of molecular descriptors, i.e. numerical characteristics of molecules derived on the basis of their chemical constitution, topology, geometry, inherent wave function, and potential energy surface, were calculated using quantum-chemical semiempirical methodology. These descriptors were then used to construct multilinear regression and artificial neural network models for the observed mutagenic activity.

In general, QSAR/QPAR studies can be useful in establishing biochemical mechanisms/interactions, and in developing combinatorial strategies for the synthesis of environmentally safe chemicals. Such studies involving aminoazobenzene derivatives are particularly important because of their widespread use in the textile industry [23], and in the synthesis of polymer supported catalysts [24], dendrimers [25], and chemosensors for monosaccharides [26]. Recently, azo dyes have been utilized to selectively functionalize the explosive 2,4,6-trinitrotoluene to a surface-enhanced resonance Raman active species that allows for sensitive detection [27].

2. Methods

We employed the program CODESSA [28], in conjunction with the program AMPAC 5.0, [29] to develop correlation equations for the aminoazobenzene derivatives in this investigation; CODESSA has been used successfully to establish correlations in a variety of applications [30–33]. The structures of all the compounds involved in this study were fully optimized at the semiempirical AM1 computational level as implemented in AMPAC 5.0; many of these compounds have numerous local minima on their potential energy surfaces and extensive searches were performed to locate those conformers with the lowest heat of formation, ΔH_f. Structural and electronic properties for some of these azo derivatives calculated using density functional theory at the BP/DN** computational level [34] can be found in our previous papers [35–37]. The CODESSA/AMPAC integrated software package was used to calculate hundreds (>300) of molecular descriptors (constitutional, topological, geometrical, electrostatic, quantum-chemical, and thermodynamic) for each of the compounds in this study. The log of the octanol–water partition coefficient, logP [38], which has been implicated in a variety of mutagenicity QSAR/QPARs [15,16], and the log of the aqueous solubility, logS, were calculated

using the additive-constitutive approach implemented in the logD suite (version 4.5) from Advanced Chemical Development, Inc. [39], and added to the pool of descriptors. The entire collection of descriptors was then used in conjunction with the statistical facilities of CODESSA to develop multilinear regression models for the log of the measured mutagenicity (rev/nmol) in TA98 + S9, logTA98.

Using the pool of descriptors discussed above, the software package Neuroshell Predictor from Ward Systems Group, Inc. [40] was employed to develop artificial neural network (ANN) models [41] for logTA98. The basic building blocks of ANNs are simulated neurons which are connected via a network, see Fig. 1. The network in Neuroshell Predictor begins by finding *linear* relationships between the inputs (calculated values of molecular descriptors) and the outputs (measured values of mutagenic activity). Weight values are assigned to the links between the input and output neurons. After these linear relationships are established, neurons are added in a "hidden layer" so that *nonlinear* relationships can be found. Neural networks are known for their ability to model a wide range of functions without any a priori knowledge of the underlying concepts.

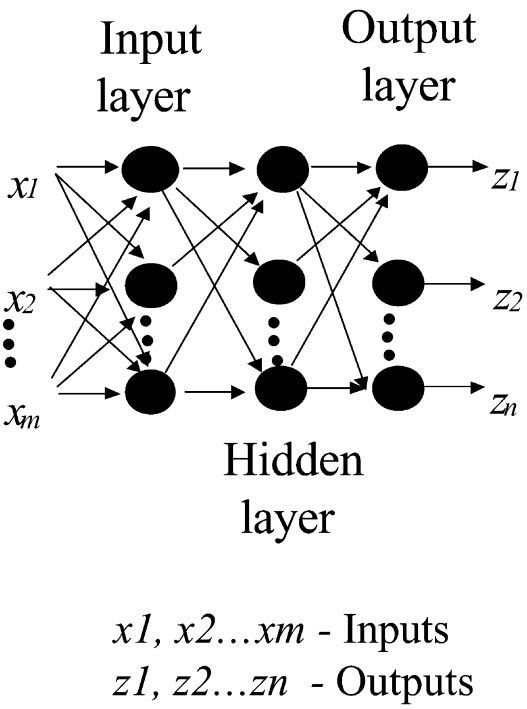


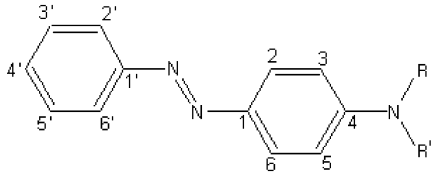
Fig. 1. Artificial neural network schematic.

3. Results and discussion

In Table 1 we list the observed mutagenicity data in the TA98 + S9 for the aminoazobenzene derivatives included in this study; calculated values of logP and logS and reported melting point values have also been included in this table [3,4,42–55]. Our literature survey found values for the mutagenic activity in TA98 + S9 of 43 aminoazobenzene derivatives. It should be noted that these measurements come from a variety of laboratories, over an extended period of time and, as a result, are likely to involve some "noise," e.g. the mutagenicity of 3′-Me-MAB has been reported as 445 rev/nmol [4] and also as 233 rev/nmol [51]. Fortunately, the mutagenic activity of these 43 compounds span over 4 orders of magnitude and provide a broad range of activity as input for QSAR/QPAR development.

The calculated values of logP (at pH = 7) for the molecules in Table 1 range from 1.6 to 5.2, showing that these azo derivatives have a relatively broad range of hydrophobic character; the average value of logP for these 43 compounds is 3.51 with a standard deviation of 0.83. Values of logP in this range are typical of those found for most drugs produced by the pharmaceutical industry, where logP is generally less than 5 [56]. The values of logS (at pH = 7) for the molecules in Table 1, calculated without using experimental melting point data, are generally negative, confirming their relatively low solubility in water. When values of the melting points of these azo derivatives could be found in the literature, their aqueous solubilities were recalculated with the ACD software [39] using this additional data and the resulting, generally more reliable [39], values of logS are also listed in Table 1. In cases where the melting point is unusually high, the predicted solubility is generally lower than that predicted in the absence of melting point data, e.g. the melting point of 3′-COOH-MAB has been reported as 207–208 °C

Table 1
Observed mutagenicity (rev/nmol) in the TA98 *Salmonella typhimurium* bacterial strain with S9 activation, calculated values [39] of logP and logS and melting points (°C), for derivatives of (A) AAB, (B) MAB, (C) DAB, and for (D) their metabolites



Compounds	Mutagenic activity in TA98 + S9 (rev/nmol) [Ref.]	LogP[a]	LogS[b]	Melting point (°C) [Ref.]
(A) AAB (R = R′ = H)				
4′-NEt_2-3-OMe-AAB	0.007 [42]	5.16	−3.41(−3.77)	147–149 [43]
2-OMe-AAB	0.010 [44]	3.87	−1.85(−2.43)	157–159 [3]
4′-OH-AAB	0.053 [45]	2.55	−0.66(−1.34)	180–181 [46]
3′-Me-4′-OH-AAB	0.059 [4]	3.01	−1.14	[c]
4′-OH-2′,3-diMe-AAB (4′-OH-OAT)	0.112 [47]	3.47	−1.62	[c]
AAB	0.204 [42]	3.13	−1.05(−1.39)	124–125 [48]
3′-Me-AAB	0.240 [4]	3.59	−1.52(−1.57)	89–91 [46]
3-OMe-4′-$N(CH_2CH_2OH)_2$-AAB	0.390 [42]	2.58	−1.48(−1.32)	129–131 [43]
3′-CH_2OH-AAB	0.596 [4]	1.94	−0.25(−0.23)	107–109 [49]
3-OH-AAB	0.687 [45]	2.97	−1.02	[c]
3-OCH_2CH_2OH-4′-$N(CH_2CH_2OH)_2$-AAB	1.052 [42]	1.60	−0.89(−0.54)	132–134 [43]
3-OCH_2CH_2OH-AAB	1.348 [42]	2.51	−0.92(−0.49)	65–67 [43]
2′-CH_2OH-3-Me-AAB	2.012 [47]	2.40	−0.72	[c]
4′-OMe-AAB	2.300 [44]	2.95	−1.09(−1.57)	155–159 [3]
2′,3-diMe-AAB (OAT)	2.676 [47]	4.05	−2.00(−2.14)	101–102 [50]
3-OBu-AAB	4.983 [42]	5.08	−3.15(−2.92)	63–65 [43]
3-OEt-AAB	13.802 [42]	4.02	−2.07(−2.20)	107–109 [43]
3-OPr-AAB	18.919 [42]	4.55	−2.60(−2.61)	97–98 [43]
3-OMe-AAB	77.065 [42]	3.48	−1.54(−1.68)	110–111 [43]
(B) MAB ($R = CH_3$, R′ = H)				
3′-Me-4′-OH-MAB	0.071 [4]	3.67	−1.77	[c]
3′-COOH-MAB	0.124 [4]	3.52	1.00(0.16)	207–208 [49]

(*continued on next page*)

Table 1 (*continued*)

Compounds	Mutagenic activity in TA98 + S9 (rev/nmol) [Ref.]	LogP[a]	LogS[b]	Melting point (°C) [Ref.]
4′-OH-MAB	0.140 [45]	3.21	−1.30	[c]
MAB	0.183 [51]	3.79	−1.68(−1.74)	87.5–88 [51]
4′-Me-MAB	0.283 [51]	4.25	−2.16(−2.35)	105–105.5 [51]
3′-Me-MAB	0.445 [4]	4.25	−2.16(−2.39)	109–109.5 [46]
3′-CH_2OH-MAB	0.503 [4]	2.60	−0.89(−0.10)	119–121 [4]
(C) DAB (R = R′ = CH_3)				
3′-Me-4′-OH-DAB	0.110 [52]	4.31	−2.41	[c]
DAB	0.140 [52]	4.43	−2.31(−2.62)	117–118 [46]
3′-COOH-DAB	0.201 [4]	4.17	0.37(−0.52)	212–213 [49]
2-Me-DAB	0.220 [52]	4.89	−2.80(−2.56)	64–66 [53]
3′-Me-DAB	0.356 [4]	4.89	−2.80(−3.52)	169–170 [49]
3′-CHO-DAB	0.383 [4]	3.91	−2.07(−2.06)	97–99 [49]
3′-CH_2OAC-DAB	0.518 [4]	4.14	−2.55(−2.13)	64–66 [49]
3′-CH_2OH-DAB	0.601 [4]	3.25	−1.52(−1.64)	122–123 [49]
(D) Metabolites (R = OH, Ac; R′ = H, CH_3)				
3′-Me-AAB-N-Ac	0.087 [4]	3.73	−1.92	[c]
3′-Me-4′-OH-AAB-N-Ac	0.089 [4]	3.15	−1.54(−2.13)	188 [49]
N-OH-2-OMe-AAB	0.110 [54]	4.08	−2.14(−2.59)	145–146[d] [44]
3′-Me-MAB-N-Ac	0.524 [4]	3.03	−1.43(−1.70)	149–150 [49]
N-OH-MAB	0.650 [55]	2.74	−0.92(−1.51)	171–174 [51]
N-OH-3′-Me-MAB	1.000 [51]	3.20	−1.40	[c]
N-OH-AAB	1.030 [54]	2.98	−1.03(−1.89)	195–197 [44]
N-OH-4′-Me-MAB	1.132 [51]	3.20	−1.40	[c]
N-OH-3-OMe-AAB	192.000 [54]	3.08	−1.30(−1.38)	113–114[d] [44]

[a] These values of logP are calculated at pH = 7 [39].

[b] The values of *S* (g/l) were calculated at pH = 7. The values in parentheses were recalculated using the melting point data; this is expected to give better estimates of the aqueous solubility [39].

[c] Melting points for these compounds could not be found.

[d] Melting points of hydrochloride salts.

and, using this data, the calculated value of logS is reduced from 1.00 to 0.16.

4. Multilinear regression models

The statistical models we developed primarily used the best multilinear regression (BMLR) method implemented in CODESSA [28]; this method selects the best two-descriptor regression model, the best three-descriptor regression model, etc., based on the highest value of the square of the regression coefficient, R^2. The models are constituted from a reduced set of non-collinear descriptors as determined by the pair correlation matrix. The heuristic method implemented in CODESSA [28] was employed in a few instances to provide alternative correlation equations (albeit with lower R^2 values) to increase the number and diversity of descriptors that were considered in developing ANN models.

In Table 2 we list BMLR equations using N descriptors ($N = 3$–5) for the relative mutagenic activity of the 43 aminoazobenzene compounds listed in Table 1; Fisher criterion F-values, F, variances, s^2, and squares of the regression correlation coefficients, R^2, are also given in Table 2A. The descriptors in the correlation equations in Table 2 are identified in Table 3 [28,29,38,39,57–64]. Correlation plots of the 4- and 5-descriptor BMLR models are shown in Fig. 2A and B, respectively; the observed values of logTA98 and those predicted from the various models are given in Table 4. As can be seen from Table 2, a 4-descriptor equation accounts for about 79% of the variation in mutagenic activity of the azo derivatives in Table 1, and increasing the number of descriptors to 5, leads to an equation that accounts for 85% of the variation.

It is interesting to note that the hydrophobicity descriptor logP does not appear in any of these BMLR correlation equations. At first sight this may seem somewhat surprising since the QSAR studies of Debnath et al. [15] suggest that hydrophobicity plays an important role in regulating mutagenic activity in TA98 + S9 for a variety of aromatic and heteroaromatic amines. It should be noted, however, that their investigation involved a much more diverse collection of parent structures, e.g. anthracenes, phenanthrenes, fluorenes, etc. but no aminoazo derivatives. Furthermore, it is evident from Table 1 that these aminoazobenzene compounds can have rather similar values of logP but quite different mutagenic activities. Nevertheless, we used the heuristic method implemented in CODESSA to force logP into a 5-descriptor

Table 2
Quantitative structure–activity/structure–property relationships for the mutagenicity of the aminoazobenzene derivatives in Table 1[a]

QSAR/QPAR	N	R^2	s^2	F
(A) BMLR equations				
$\log TA98 = 261.03(\pm 55.35) + 842.33(\pm 151.21)Q_1 - 1.06(\pm 0.21)Q_2 + 0.01(\pm 0.00)E_1$	3	0.66	0.28	25.22
$\log TA98 = -2.03(\pm 1.41) - 4.23(\pm 0.60)E_2 + 22.54(\pm 3.05)Q_3 + 701.94(\pm 104.01)Q_1 - 0.40(\pm 0.07)H_1$	4	0.79	0.18	34.56
$\log TA98 = -60.51(\pm 41.61) - 4.22(\pm 0.52)E_2 - 0.46(\pm 0.07)Q_4 + 573.29(\pm 95.55)Q_1 - 0.41(\pm 0.06)H_1 + 75.01(\pm 19.34)Q_5$	5	0.85	0.13	40.41
(B) Heuristic equation				
$\log TA98 = 6.02(\pm 1.76) + 20.94(\pm 3.07)Q_3 - 7.55(\pm 1.52)G_1 - 0.29(\pm 0.09)O_1 - 3.81(\pm 0.66)E_2 + 436.31(\pm 90.05)Q_1$	5	0.80	0.18	28.96

[a] The molecular descriptors are identified in Table 3.

Table 3
Molecular descriptors for the QSAR/QPAR models in Table 2

Classification/label	Molecular descriptors	Reference
Geometrical		
G_1	XY Shadow/XY Rectangle[a]	[28,57]
Electrostatic		
E_1	WPSA-2 Surface weighted partial positive surface area[b]	[28,58,59]
E_2	Polarity parameter/(distance)^2 $=(Q_{max}-Q_{min})$/(distance)^2	[28,60,61]
Quantum-chemical		
Q_1	Average electrophilic reactivity index for a N atom	[28,62]
Q_2	Maximum electron-nuclear attraction for a C–C bond	[28,29]
Q_3	Minimum net atomic charge for a C atom	[28]
Q_4	Maximum electron-nuclear attraction for a C atom	[28,63]
Q_5	Maximum bond order of an N atom	[28,64]
Thermodynamic		
H_1	Final heat of formation/number of atoms	[28,29]
Others		
O_1	LogP	[38,39]

[a] This is one of six indices defined by Rohrbaugh and Jurs [57] that encode size and shape information about a molecule in three mutually perpendicular perspectives defined by the axes of inertia.

[b] This parameter is one of the charged partial surface area descriptor invented by Stanton and [58] and is calculated as (PNSA-2)*TMSA/1000, where PNSA-2 is the total charge weighted partial positive surface area and TMSA is the total molecular surface area.

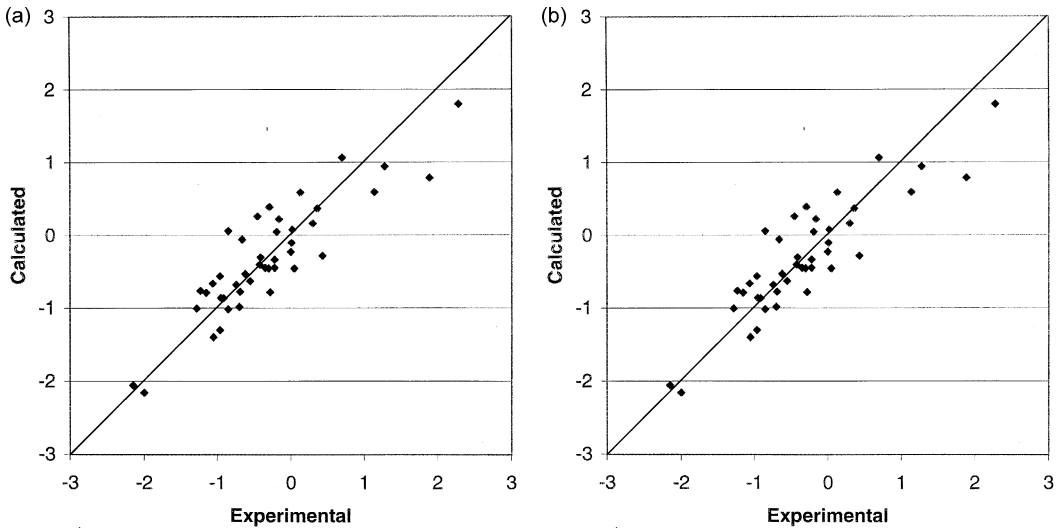


Fig. 2. Correlation Plots for the (A) 4-descriptor and (B) 5-descriptor BMLR equations in Table 2.

Table 4
Observed values of logTA98 and predicted values of logTA98 from the BMLR and heuristic QSAR/QPAR equations in Table 2 for the 43 compounds in Table 1

Compounds	Observed values of logTA98	Predicted values of logTA98			
		BMLR			Heuristic
		3-Descriptor	4-Descriptor	5-Descriptor	5-Descriptor
(A) AAB					
4′-NEt_2-3-OMe-AAB	−2.15	−1.93	−2.06	−2.11	−2.32
2-OMe-AAB	−2.00	−1.41	−2.16	−2.17	−1.59
4′-OH-AAB	−1.28	−1.01	−1.00	−1.24	−1.03
3′-Me-4′-OH-AAB	−1.23	−0.93	−0.76	−0.96	−0.80
4′-OH-2′,3-diMe-AAB (4′-OH-OAT)	−0.95	−0.26	−0.86	−0.87	−1.19
AAB	−0.69	−0.75	−0.78	−0.85	−0.41
3′-Me-AAB	−0.62	−0.61	−0.53	−0.55	−0.24
3-OMe-4′-$N(CH_2CH_2OH)_2$-AAB	−0.41	−0.96	−0.31	−0.37	−0.36
3′-CH_2OH-AAB	−0.22	−0.38	−0.34	−0.44	−0.08
3-OH-AAB	−0.16	+0.36	+0.22	+0.11	+0.14
3-OCH_2CH_2OH-4′-$N(CH_2CH_2OH)_2$-AAB	+0.02	+0.20	+0.07	+0.02	+0.06
3-OCH_2CH_2OH-AAB	+0.13	+0.83	+0.59	+0.54	+0.69
2′-CH_2OH-3-Me-AAB	+0.30	+0.53	+0.16	+0.30	−0.21
4′-OMe-AAB	+0.36	−0.63	+0.37	+0.10	+0.34
2′,3-diMe-AAB (OAT)	+0.43	−0.22	−0.29	+0.30	−0.48
3-OBu-AAB	+0.70	+0.94	+1.06	+1.00	+1.10
3-OEt-AAB	+1.14	+0.45	+0.59	+0.54	+0.62
3-OPr-AAB	+1.28	+0.72	+0.94	+1.04	+0.84
3-OMe-AAB	+1.89	+0.26	+0.79	+0.92	+0.87
(B) MAB					
3′-Me-4′-OH-MAB	−1.15	−0.83	−0.79	−0.98	−0.92
3′-COOH-MAB	−0.91	−1.21	−0.87	−0.84	−0.74
4′-OH-MAB	−0.85	−0.91	−1.01	−1.26	−1.02

(*continued on next page*)

Table 4 (*continued*)

Compounds	Observed values of logTA98	Predicted values of logTA98			
		BMLR			Heuristic
		3-Descriptor	4-Descriptor	5-Descriptor	5-Descriptor
MAB	−0.74	−0.50	−0.68	−0.77	−0.35
4′-Me-MAB	−0.55	−0.55	−0.63	−0.76	−0.62
3′-Me-MAB	−0.35	−0.36	−0.45	−0.48	−0.44
3′-CH_2OH-MAB	−0.30	−0.16	−0.46	−0.49	−0.13
(C) DAB					
3′-Me-4′-OH-DAB	−0.96	−0.46	−0.56	−0.82	−1.06
DAB	−0.85	−0.58	+0.06	−0.01	−0.30
3′-COOH-DAB	−0.70	−1.32	−0.98	−0.97	−1.13
2-Me-DAB	−0.66	−0.59	−0.06	−0.69	−0.33
3′-Me-DAB	−0.45	−0.43	+0.26	+0.21	−0.19
3′-CHO-DAB	−0.42	−1.23	−0.41	−0.50	−0.31
3′-CH_2OAC-DAB	−0.29	+0.16	+0.39	+0.34	+0.10
3′-CH_2OH-DAB	−0.22	−0.23	−0.45	−0.32	−0.45
(D) Metabolites					
3′-Me-AAB-N-Ac	−1.06	−1.06	−0.67	−0.27	−0.75
3′-Me-4′-OH-AAB-N-Ac	−1.05	−1.04	−1.40	−1.27	−1.60
N-OH-2-OMe-AAB	−0.96	−0.29	−1.30	−1.05	−0.81
3′-Me-MAB-N-Ac	−0.28	−0.32	−0.78	−0.32	−0.85
N-OH-MAB	−0.19	+0.36	+0.04	+0.16	+0.27
N-OH-3′-Me-MAB	0.00	+0.49	−0.23	+0.04	−0.12
N-OH-AAB	+0.01	+0.08	−0.10	+0.04	+0.28
N-OH-4′-Me-MAB	+0.05	+0.35	−0.46	−0.31	−0.48
N-OH-3-OMe-AAB	+2.28	+1.37	+1.80	+1.94	+1.95

multilinear regression model; the remaining 4 descriptors were chosen as usual to optimize R^2. The resulting correlation equation is listed in Table 2, where it can be seen to account for about 80% of the variation in mutagenic activity; a reoptimized 4-parameter multilinear regression equation using the same descriptors, but without logP, accounts for only 57% of the variation.

In general, it is not possible to identify unambiguously the mechanistic significance of each descriptor in a QSAR/QPAR model, particularly when the available data is rather limited. However, several of the descriptors that appear in these BMLR models at least seem reasonable. For example, all three of the BMLR models in Table 2 suggest that the relative mutagenicity of these aminoazobenzene derivatives increases as the *average* electrophilic reactivity index, $\overline{\mathrm{ERI}}$, for all the nitrogen atoms in the compound increases. The ERI for a nitrogen atom is defined as

$$\mathrm{ERI}_N = \sum_{j\in \mathrm{N}} \frac{C_{j^2_{\mathrm{LUMO}}}}{\varepsilon_{\mathrm{LUMO}} + 10}, \qquad (1)$$

where $C_{j\mathrm{LUMO}}$ denotes the jth atomic orbital coefficient of the lowest unoccupied molecular orbital (LUMO) and $\varepsilon_{\mathrm{LUMO}}$ is the orbital energy of the LUMO [28]. Since metabolic reductive-cleavage of the azo linkage and/or hydroxylation at the amino nitrogen atom (and subsequent conversion to an arylnitrenium ion) are believed to be important steps in the onset of mutagenesis [7–11], the presence of this descriptor in these correlation equations seems appropriate. The ERI is a measure of the acidity of a nitrogen atom: larger values of this index occur at azo nitrogen atoms and smaller values occur at amino nitrogen atoms. We can gain some insight into factors that affect $\overline{\mathrm{ERI}}_N$ by comparing ERI_N values for the different types of nitrogen atoms in the two positional isomers 2-OMe-AAB and 3-OMe-AAB, whose relative mutagenic activities are so different. It should be noted that AM1 finds the structures of these isomers to be somewhat different: the methoxy group in 3-OMe-AAB is significantly out-of-the-plane of the benzene ring to which it is attached, whereas in 2-OMe-AAB it is in the plane. The value of ERI_N for the amino nitrogen atom in 2-OMe-AAB is ~8% larger than it is in 3-OMe-AAB, whereas the value for each of the two azo nitrogen atoms in 3-OMe-AAB are ~14% larger. This may suggest that some of the enhanced mutagenic activity of 3-OMe-AAB compared to 2-OMe-AAB is linked to its capacity to undergo reductive cleavage. It should also be mentioned that an additional $-NR_2$ ($R = CH_2CH_3$ or CH_2CH_2OH) group on an aminoazobenzene derivative lowers $\overline{\mathrm{ERI}}_N$ and the correlation equations predict lower mutagenic activity. This is a reflection, for example, of the low activity of 4′-NEt_2-3-OMe-AAB compared to that of the parent compound 3-OMe-AAB, see Table 1.

Since the electrostatic descriptor $(Q_{\mathrm{MAX}}-Q_{\mathrm{MIN}})/D^2$ is common to two of the BMLR models in Table 2 and also appears in two of our ANN models (*vide infra*), it deserves some further discussion. The values of Q_{MAX} and Q_{MIN} in this composite parameter are the maximum and minimum of the *empirical* partial charges calculated using the Sanderson electronegativity scale [65] and D is the distance between the positions of Q_{MAX} and Q_{MIN}. As can be seen from the 4- and 5-descriptor correlation equations, an increase in the value of this descriptor for a compound in Table 1 decreases its mutagenic activity. The values of the so-called polarity parameter (Q_{MAX}–Q_{MIN}) for the aminoazobenzene derivatives in Table 1 vary over a relatively narrow range, 0.15–0.27e, but the positions of Q_{MAX} and Q_{MIN} can differ substantially and the values of $(Q_{\mathrm{MAX}}$–$Q_{\mathrm{MIN}})/D^2$ vary over some two orders of magnitude. Thus, when the centers of maximum and minimum atomic charge (based on electronegativity assignments) are spatially localized, the relative mutagenic activity of the compound is decreased.

All three BMLR equations, as well as the heuristic equation, in Table 2 involve at least one descriptor associated with the carbon atoms in the molecule. Detoxification mechanisms, e.g. C-hydroxylation, are believed to play a role in the relative mutagenic activity of aminoazobenzene derivatives [47] and descriptors involving the carbon atoms may help incorporate such features into the models.

In order to provide some measure of how well these BMLR models perform in a predictive

capacity, we chose 13 azo derivatives from the literature that have been classified either as "active" or "inactive" in TA98 + S9, and predicted their relative mutagenicity from each of the models, see Table 5 [43,51,55,66][1]; calculated values of logP and LogS for these compounds and their experimental melting points are also listed in this table. It should be noted that the classification of a compound as "inactive" may be somewhat misleading, since the lower limits of detection in the reported experiments are difficult to assess. For example, Hashimoto et al. [44] classify 2-OMe-AAB as

Table 5
Predicted values of logTA98 from the 3-, 4-, and 5-descriptor BMLR and 5-descriptor heuristic models listed in Table 2 [28] for the 43 compounds in Table 1

Compounds	LogP	LogS	Melting points (°C) [Ref.]	Predicted values of log TA98			
				BMLR			Heuristic
				3-Descriptor model	4-Descriptor model	5-Descriptor model	5-Descriptor model
(A) Active compounds							
2′-COOH-DAB	4.91	0.26(0.46)	179–182 [66]	−0.77	−0.56	+0.47	−0.94
2′-Me-DAB	4.89	−2.80(−2.72)	72–73 [67]	−0.46	+0.28	+0.75	−0.01
2′-CH_2OH-DAB	3.25	−1.52	[a]	+0.29	−0.02	+0.02	−0.25
4′-Me-DAB	4.89	−2.80(−3.49)	165–167 [68]	−0.65	+0.09	−0.07	−0.54
4′-CH_2OH-DAB	3.25	−1.52	[a]	−0.39	−0.59	−0.66	−1.04
4′-MC-N-OH-MAB	3.09	−1.60(−1.96)	164–167 [55]	−0.11	+1.22	+2.10	+0.54
(B) Inactive compounds							
3-OPr-4′-NEt_2-AAB	6.22	−4.49(−4.46)	98–100 [43]	−1.26	−1.91	−1.95	−1.94
3-OEt-4′-$N(CH_2CH_2OH)_2$-AAB	3.11	−2.03(−1.62)	100–102 [43]	−1.00	−0.69	−0.71	−1.05
3-OPr-4′-$N(CH_2CH_2OH)_2$-AAB	3.64	−2.57(−2.10)	92–93 [43]	−0.56	−0.65	−0.67	−0.80
3-OBu-4′-$N(CH_2CH_2OH)_2$-AAB	4.17	−3.11(−2.72)	100–102 [43]	−0.01	−0.60	−0.63	−0.44
N-Et_2-AAB	5.49	−3.37	[a]	−0.48	−0.53	−0.62	−1.24
3-OCH_2CH_2OH-4′-NEt_2-AAB	4.18	−2.80	143–144 [43]	−1.59	−1.98	−2.04	−1.74
4′-MC-MAB	4.13	−2.36(−2.92)	171–173 [55]	−1.54	−0.85	−0.28	−1.26
(C)							
N-OAc-MAB	2.94	−1.37(−1.01)	74–76 [55]	+0.65	−0.39	+0.08	−0.10
N-OAc-3′-Me-MAB	3.40	−1.85(−1.57)	85.5–86.5 [51]	+0.87	−0.28	+0.23	−0.32
N-OAc-4′-Me-MAB	3.40	−1.85(−1.68)	98–99 [51]	+0.73	−0.39	−0.02	−0.31

[a] Melting points for these compounds could not be found.

[1] Compound synthesized in our laboratories. Private communication from U. Osterwinter, ACROS Organics BVBA, Belgium.

inactive in TA98 + S9, but the more recent work by Freeman et al. [42] shows that this compound has a low, but detectable, mutagenic activity in this bacterial strain. In general, we chose the particular aminoazobenzene derivatives in Table 5 to have some structural similarity to the compounds used to develop the QSAR/QPARs in Table 2, e.g. the mutagenic activity of 3-OCH_2CH_2OH-4′-$N(CH_2CH_2OH)_2$-AAB was used to develop the correlation models which were then used to predict the activity of the related inactive compound 3-OCH_2CH_2OH-4′-NEt_2-AAB.[2] On the other hand, no values for the relative mutagenicity of halogen substituted AAB derivatives is currently available and we have not included them in Table 5. We have, however, included 4′-methoxycarbonyl-MAB or 4′-methylcarboxylate-MAB (4′-MC-MAB) and its metabolite 4′-MC-N-OH-MAB although no methoxycarbonyl compounds were used in the development of the regression equations in Table 2.

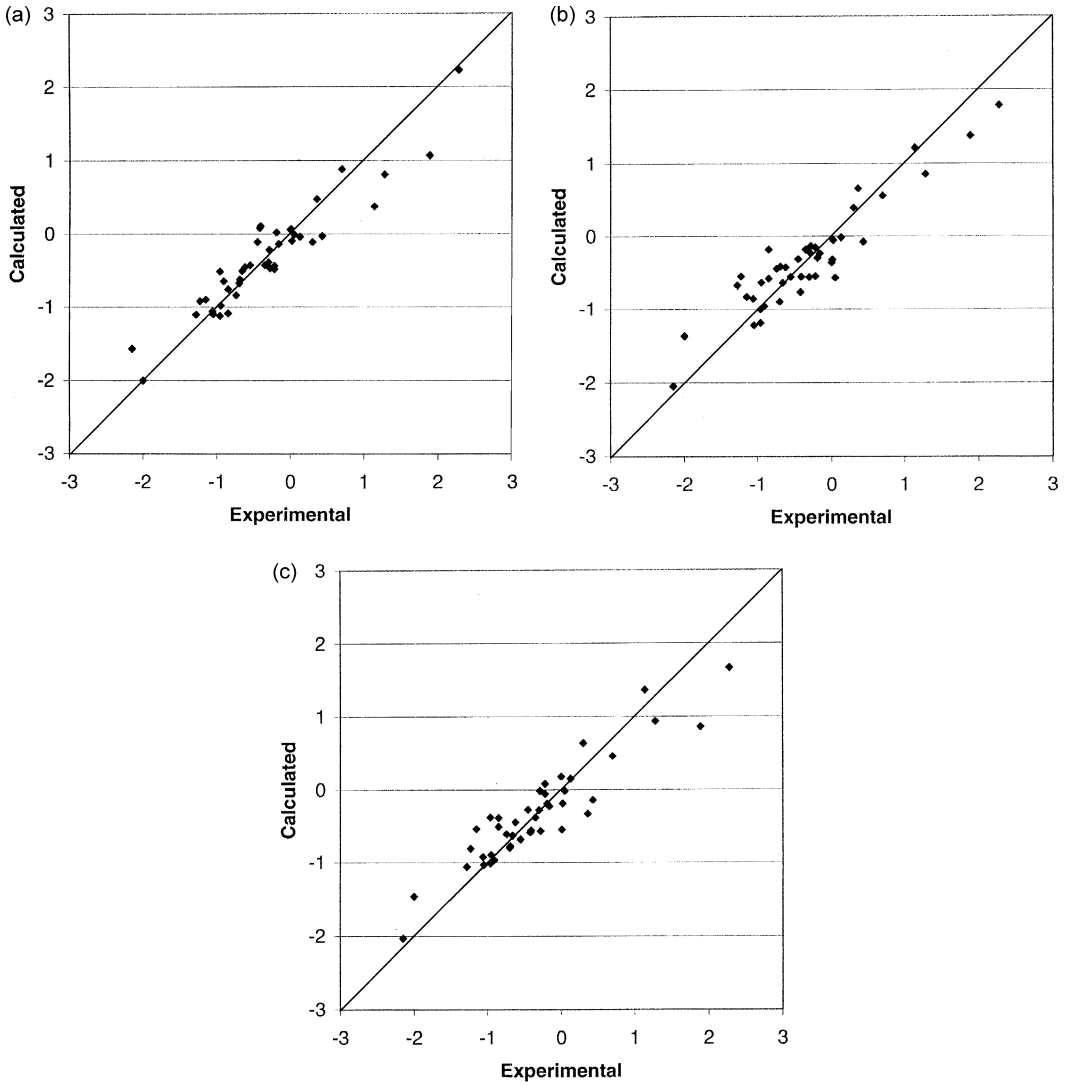


Fig. 3. Correlation plots for three 3-descriptor ANN models.

[2] Private communication from Prof. Freeman HS, North Carolina State University, Raleigh, NC 27695, USA.

As can be seen from Table 5, on average, the BMLR models do find the compounds classified as active to have higher relative mutagenicity than those classified as inactive, e.g. using the 5-descriptor BMLR model, the average value of logTA98 for the six active compounds is +0.44 and for the seven inactive compounds is −1.21. Interestingly, each of the models predicts the mutagenic activity of 4′-MC-N-OH-MAB to be greater than that of the parent compound 4′-MC-MAB, as would be expected; however, the predicted increase, 2.4 log units with the 5-descriptor BMLR model, is greater than that observed for similar N-hydroxylations, <1 log unit in going from MAB to N-OH-MAB, see Table 1. It is important to note that all the regression models in Table 2 consistently predict that four of the seven inactive compounds are actually active to some extent; only 3-OPr-4′-NEt_2-AAB and 3-OCH_2-CH_2OH-4′-NEt_2-AAB could be labeled reasonably as inactive in the context of these models. Part of the problem in using our correlation equations with inactive compounds may be that only 2 of the 43 compounds used to develop these equations might be classified as inactive, e.g. 3-OMe-4′-NEt_2-AAB and 2-OMe-AAB. Additional data on weakly mutagenic azobenzene derivatives are needed to enhance the predictive power of these regression models. On the other hand, some questions have been raised about the inactivity of at least one of the compounds in Table 5, solvent yellow 56, N-Et_2-AAB [67].

We also used the BMLR models to predict mutagenicities for three of the postulated metabolites of MAB: N-OAc-MAB, N-OAc-3′-Me-MAB, and N-OAc-4′-Me-MAB, see Table 5C. Mutagenic activities for these compounds *in the absence of S9* have been reported and they are extremely high, 26.0, 28.0, and 55.0 rev/nmol, respectively [51]. As can be seen from the Table 5, their predicted values *in the presence of S9* suggest that these compounds are active, but their relative mutagenicities seem rather low.

5. Artificial neural network (ANN) models

We next set about training ANNs to learn the mutation data in Table 1. We started with a pool of about 50 descriptors; these were chosen based on our BMLR and heuristic studies described above, as well as, on the findings of other authors using different classes of molecules [12–22]. Only three descriptors were needed to adequately train these ANN models with the limited data available, and we found a variety of 3-descriptor models for

Table 6
Molecular descriptors and their relative importance (%) for the artificial neural network models [40]

Model	R^2 [a]	Molecular descriptors	Relative importance (%)	Ref.
I	0.88(0.97)	Polarity parameter/(distance)^2	55.9	[60,61]
		Minimum net atomic charge for a carbon atom	37.4	[28]
		Final heat of formation/number of atoms	0.7	[28,29]
II	0.86(0.95)	Polarity parameter/(distance)^2	49.6	[60, 61]
		Average electrophilic reactivity index for nitrogen atom	42.3	[62]
		Log$P$	8.1	[38]
III	0.84(0.91)	Average electrophilic reactivity index for nitrogen atom	54.4	[62]
		Average information content (order 0)[b]	34.9	[68–70]
		Final heat of formation/number of atoms	10.7	[28,29]

[a] The values of R^2 in parentheses were obtained with the Enhanced Generalization option turned *off*.

[b] The average information content (IC) is calculated using the equation $\overline{IC} = \Sigma\ N_i \log_2 N_i$ where $N_i = n_i/n$ and n_i is the number of atoms in the ith class and n is the total number of atoms in the molecule. The division of atoms into different classes depends on the coordination sphere taken into account which leads to indices of different order κ.

Table 7
Observed values of logTA98 and predicted values of logTA98 from the ANN models [40] in Table 2 for the 43 compounds in Table 1

Compounds	Observed values of logTA98	Predicted values of logTA98		
		Model I	Model II	Model III[a]
(A) AAB				
4′-NEt_2-3-OMe-AAB	−2.15	−1.57	−2.05	−2.03(−2.15)
2-OMe-AAB	−2.00	−2.00	−1.37	−1.46(−2.01)
4′-OH-AAB	−1.28	−1.11	−0.68	−1.06(−1.27)
3′-Me-4′-OH-AAB	−1.23	−0.92	−0.55	−0.81(−1.11)
4′-OH-2′,3-diMe-AAB (4′-OH-OAT)	−0.95	−0.98	−0.64	−0.90(−1.23)
AAB	−0.69	−0.62	−0.41	−0.78(−0.69)
3′-Me-AAB	−0.62	−0.45	−0.43	−0.45(−0.60)
3-OMe-4′-$N(CH_2CH_2OH)_2$-AAB	−0.41	−0.11	−0.56	−0.56(−0.41)
3′-CH_2OH-AAB	−0.22	−0.48	−0.15	+0.08(+0.13)
3-OH-AAB	−0.16	−0.14	−0.23	−0.23(−0.15)
3-OCH_2CH_2OH-4′-$N(CH_2CH_2OH)_2$-AAB	+0.02	−0.09	−0.05	−0.19(+0.02)
3-OCH_2CH_2OH-AAB	+0.13	−0.04	−0.02	+0.15(+0.20)
2′-CH_2OH-3-Me-AAB	+0.30	−0.11	+0.39	+0.64(+0.27)
4′-OMe-AAB	+0.36	+0.48	+0.65	−0.33(−0.29)
2′,3-diMe-AAB (OAT)	+0.43	−0.03	−0.08	−0.14(+0.16)
3-OBu-AAB	+0.70	+0.88	+0.55	+0.46(+0.46)
3-OEt-AAB	+1.14	+0.38	+1.21	+1.36(+1.84)
3-OPr-AAB	+1.28	+0.81	+0.85	+0.94(+1.11)
3-OMe-AAB	+1.89	+1.07	+1.37	+0.86(+1.33)
(B) MAB				
3′-Me-4′-OH-MAB	−1.15	−0.90	−0.83	−0.54(−0.76)
3′-COOH-MAB	−0.91	−0.65	−0.96	−0.97(−0.89)
4′-OH-MAB	−0.85	−1.09	−0.58	−0.39(−0.39)
MAB	−0.74	−0.76	−0.45	−0.61(−0.55)

(*continued on next page*)

Table 7 (*continued*)

Compounds	Observed values of logTA98	Predicted values of logTA98		
		Model I	Model II	Model III[a]
4′-Me-MAB	−0.55	−0.43	−0.56	−0.68(−0.57)
3′-Me-MAB	−0.35	−0.43	−0.18	−0.38(−0.24)
3′-CH_2OH-MAB	−0.30	−0.39	−0.56	−0.28(−0.46)
(C) DAB				
3′-Me-4′-OH-DAB	−0.96	−0.51	−1.00	−0.38(−0.98)
DAB	−0.85	−0.75	−0.18	−0.51(−0.39)
3′-COOH-DAB	−0.70	−0.67	−0.90	−0.80(−0.70)
2-Me-DAB	−0.66	−0.51	−0.64	−0.63(−0.47)
3′-Me-DAB	−0.45	−0.11	−0.31	−0.28(−0.08)
3′-CHO-DAB	−0.42	+0.08	−0.76	−0.59(−0.35)
3′-CH_2OAC-DAB	−0.29	−0.22	−0.23	−0.02(−0.29)
3′-CH_2OH-DAB	−0.22	−0.44	−0.55	−0.06(−0.30)
(D) Metabolites				
3′-Me-AAB-N-Ac	−1.06	−1.06	−0.86	−0.92(−1.13)
3′-Me-4′-OH-AAB-N-Ac	−1.05	−1.10	−1.22	−1.03(−1.05)
N-OH-2-OMe-AAB	−0.96	−1.12	−1.19	−1.01(−0.54)
3′-Me-MAB-N-Ac	−0.28	−0.47	−0.13	−0.57(−0.61)
N-OH-MAB	−0.19	+0.02	−0.29	−0.19(−0.24)
N-OH-3′-Me-MAB	0.00	+0.06	−0.36	+0.18(+0.07)
N-OH-AAB	+0.01	+0.07	−0.32	−0.55(−0.46)
N-OH-4′-Me-MAB	+0.05	−0.01	−0.57	−0.02(+0.05)
N-OH-3-OMe-AAB	+2.28	+2.22	+1.78	+1.66(+2.13)

[a] The values in parentheses are the predicted values with the Enhanced Generalization Option turned *off*.

which R^2 was above 0.90. Since we believe the mutagenicity data in Table 1 is rather "noisy," however, we used the Enhanced Generalization option (at 50%) available in Neuroshell Predictor; this feature typically reduces the value of R^2 for the data in the training set, but tends to enhance the predictive power of the resulting models.

Correlation plots for three ANN models are shown in Fig. 3; these particular models were chosen to illustrate some of the features of this

approach. The molecular descriptors involved in these models are listed in Table 6; values of R^2 and the relative importance of the various descriptors in these models are also given in this Table. The correlation plots in Fig. 3 clearly demonstrate how well 3-descriptor ANN models (even with the Enhanced Generalization option turned *on*) can learn the mutagenicity data in Table 1; the predicted values of logTA98 for these three ANN models are given in Table 7. (For comparison, we have also listed the predicted values of logTA98 using Model III with the Enhanced Generalization option turned off.) Model II includes logP as a descriptor, but its relative importance to the model is only about 8%. The average ERI and $(Q_{MAX}–Q_{MIN})/D^2$ descriptors, which played such important roles in the BMLR equations, each appear in two of the ANN models and their relative importance in the context of these models is quite high, see Table 6.

Table 8
Predicted values of logTA98 from the 3-descriptor ANN models in Table 6

Compounds	Predicted values of TA98		
	Model I	Model II	Model III[a]
(A) Active compounds			
2′-COOH-DAB	−0.79	−0.65	−0.56 (+0.11)
2′-Me-DAB	−0.17	−0.19	−0.01 (+0.38)
2′-CH_2OH-DAB	−0.49	+0.22	+2.36 (+5.84)
4′-Me-DAB	−0.07	−0.64	−0.67 (−0.52)
4′-CH_2OH-DAB	−0.45	−0.84	−0.36 (−0.51)
4′-MC-N-OH-MAB	+5.57	−1.49	+0.50 (+3.81)
(B) Inactive compounds			
3-OPr-4′-NEt_2-AAB	−2.52	−1.76	−8.96(−49.7)
3-OEt-4′-$(CH_2CH_2OH)_2$-AAB	−0.52	−2.67	−2.50(−12.1)
3-OPr-4′-$(CH_2CH_2OH)_2$-AAB	−0.51	−3.13	−5.36(−29.5)
3-OBu-4′-$(CH_2CH_2OH)_2$-AAB	−0.44	−2.64	−9.26(−53.4)
N-Et_2-AAB	−1.27	+0.14	−0.19 (+0.13)
3-OCH_2CH_2OH-4′-NEt_2-AAB	−1.87	−3.59	−4.97(−22.0)
4′-MC-MAB	+0.18	−1.55	−1.36 (+0.82)
(C)			
N-OAc-MAB	+0.64	+0.09	+1.35 (−1.52)
N-OAc-3′-Me-MAB	+0.77	−0.27	+1.11 (−1.78)
N-OAc-4′-Me-MAB	+0.77	−0.46	+1.26 (+0.28)

[a] The values in parentheses are the predicted values with the Enhanced Generalization Option turned *off*.

The predicted relative mutagenicities from these three ANN models for the 13 compounds classified as active/inactive and the three MAB metabolites discussed above are listed in Table 8. All three models generally predict that the six compounds classified as active are active, although specific values for some compounds, e.g. 2′-CH_2OH-DAB and 4′-MC-N-OH-MAB, vary substantially among the models. On the other hand, these three models behave quite differently with respect to the seven compounds classified as inactive. For example, with Model I 4 of the 7 compounds are predicted to show some mutagenic activity, whereas for Models II and III only one compound, N-Et_2-AAB, is predicted to be active and, as mentioned previously, there are some questions about the classification of this molecule [67]. Although all three models suggest mutagenic activity for the metabolites of MAB, models I and III predict significantly greater activity, see Table 8 C.

6. Conclusions

Aminoazobenzene derivatives and their related structures are extremely important industrial colorants that have found a variety of new uses [23–27]. Unfortunately, some of these compounds have been shown definitely to be mutagenic/carcinogenic [2–5], although the detailed biochemical mechanisms responsible for this behavior have not been clearly delineated. The pioneering work by Freeman et al. [42,43] has clearly demonstrated

that some small structural modifications of these aminoazobenzene derivatives can reduce or eliminate their mutagenic activity, while maintaining the physical and/or chemical properties that make them useful industrial chemicals. Developing QSAR/QPARs that correlate the relative mutagenic activity of aminoazobenzene derivatives with various molecular descriptors can help identify factors that alter their relative mutagenicity.

Although the available quantitative mutagenicity data for aminoazobenzene derivatives is rather limited and likely to involve considerable noise, we have shown that a 5-descriptor multilinear regression model is capable of accounting for some 85% of the reported variation in mutagenic activity of 43 compounds in the TA98 bacterial strain with S9 activation. This is encouraging since these compounds involve AAB, MAB and DAB derivatives, for which there are at least two potential activation pathways: reductive-cleavage and N-hydroxylation at the amino nitrogen with subsequent conversion to an arylnitrenium ion. Furthermore, DAB derivatives may require an additional N-demethylation step in their activation pathway, although the rate of N-demethylation does not appear to be correlated with carcinogenic potency [2]. While it is difficult to assess these BMLR models as predictive tools, the 5-descriptor model gives plausible results for a small set of related compounds that have been classified in the literature as active or inactive. There appear to be several false positives from this model, which may be the result of including only active compounds in the development of our correlation equations. Furthermore, there is quantitative data available for only two aminoazobenzene derivatives with mutagenic activity below 0.05 rev/nmol.

Several 3-descriptor ANN models were trained that can account for over 90% of the variation in mutagenic activity of the compounds in Table 1. However, in view of the limited quality of the mutagenicity data set available and our desire to expand the predictive power of these ANN models, we opted to accept somewhat lower values of R^2 (0.84–0.88) for the training set by employing the Enhanced Generalization feature of the software. These ANN models also do well in predicting active compounds and have the added advantage of predicting less in the way of false positives than the BMLR models, assuming that the observed classifications are reliable.

There is little doubt that the predictive power of our BMLR and ANN models could be significantly improved if additional values for the relative mutagenicities of aminoazobenzene dyes and their related structures were available. We are currently incorporating mutagenicity data on the reductive-cleavage products of the compounds into the model, as well as, developing models for different bacterial strains. The results of these investigations will be reported in future publications.

Acknowledgements

The authors would like to thank Revathy Iyer from the University of Pennsylvania for technical assistance. We would also like to acknowledge the National Textile Center (Grant No. COO-PO1) for financial support of this work.

References

[1] Kinosita R. Tr Soc Path Jap 1937;27:665.
[2] Kimura T, Kodama M, Nagata C. Carcinogenesis 1982; 3:1393.
[3] Miller JA, Miller EC. Can Res 1961;21:1068.
[4] Mori Y, Niwa T, Hori T, Toyoshi K. Carcinogenesis 1980; 1:121.
[5] Sugiuva K, Halter CR, Kensler CJ, Rhoads CP. Can Res 1944:235.
[6] Kojima M, Degawa M, Hasimoto Y, Tada M. Biochem Biophy Res Commun 1991;179:817.
[7] Ashby J, Lefevre PA, Callander RD. Mut Res 1983; 116:271.
[8] Chung KT. Mut Res 1983;114:269.
[9] Hashimoto Y, Degawa M, Watanabe HK, Tada M. Gann 1981;72:937.
[10] Lin JK, Schmall B, Sharpe ID, Miuva I, Miller JA, Miller EC. Can Res 1975;35:832.
[11] Kabuldar FF, Miller JA, Miller EC. Can Res 1976; 36:2350.
[12] Shahin MM. Mut Res 1989;221:165.
[13] Chung KT, Kirkovsky L, Kirkovsky A, Purcell WP. Mut Res 1997;387:1.
[14] Chung KT, Cernighia CE. Mut Res 1992;277:201.
[15] Debnath AK, Debnath G, Shusterman AJ, Hansch C. Env Mol Mutagenis 1992;19:37.
[16] Zhang L, Sannes K, Shusterman AJ, Hansch C. Chem Biol Interactions 1992;81:149.

[17] Cunningham AR, Klopman G, Rosenkranz HS. Mut Res 1998;405:9.
[18] Ford GP, Griffen GR. Chem Biol Interactions 1992;81:19.
[19] Enslein K, Borgstedt HH. Toxicol Lett 1989;49:107.
[20] Rosenkranz HS, Klopman G. Mut Res 1989;221:217.
[21] Rose SL, Juys PC. J Med Chem 1982;25:769.
[22] Claxton LD, Walsh DB, Esancy JF, Freeman HS. Prog Clin Biol Res 1990;340:11.
[23] Stead CV. In: Shore J, editor. Chemistry of azo colorants in colorants and auxiliaries vol. 1. Society of Dyers and Colorists, 1990. p. 147.
[24] Bergbreiter DE, Osburn PL, Li C. Org Letts 2002;4:737.
[25] Wang S, Advincula RC. Org Letts 2001;3:3831.
[26] Ward CJ, Patel P, James TD. Org Letts 2002;4:477.
[27] McHugh CJ, Keir R, Graham D, Smith WE. J Chem Soc Chem Commun 2002;6:580.
[28] CODESSA^{TM}, v2.0, Semichem, 7204 Mullen, Shawnee, KS 66216, USA.
[29] AMPAC 5.0 ©1994 Semichem, 7128 Summit, Shawnee, KS 66216.
[30] Huibers PDT, Lobanov VS, Katritzky AR, Shah DO, Karelson M. Langmuir 1996;12:1462.
[31] Katritzky AR, Sild S, Karelson M. J Chem Inf Comput Sci 1998;38:1171.
[32] Katritzky AR, Karelson M, Lobanov VS. Pure Appl Chem 1997;69:245.
[33] Mu L, Drago RS, Richardson DE. J Chem Soc Perkin II 1998:159.
[34] SPARTAN v5.0, Wavefunction Inc, 18401 Von Karmen Avenue, Suite 370, Irvine, CA 92612, USA.
[35] Bhat KL, Freeman HS, Velga J, Sztandera L, Trachtman M, Bock CW. Dyes and Pigments 2000;46:109.
[36] Bhat KL, Trachtman M, Bock CW. Dyes and Pigments 2001;48:197.
[37] Bhat KL, Garg A, Trachtman M, Bock CW. Dyes and Pigments 2001;50:133.
[38] Bhat KL, Garg A, Bock CW. Dyes and Pigments 2002; 52:145.
[39] ACD module: v4.5, Advanced Chemistry Development, Inc, 90 Adelaide St. W., Suite 702, Toronto, Ontario, Canada M5H 3V9.
[40] Neuroshell Predictor, Ward Systems Group, Inc, Executive Park West, 5 Hillcrest Drive, Frederick, MD 21703, USA.
[41] Sigman ME, Rives SS. J Chem Inf Comput Sci 1994; 34:617.
[42] Freeman HS, Esancy M, Esancy JF, Claxton LD. Chem Tech 1991:439.
[43] Freeman HS, Esancy JF, Claxton JD. In: Peters AT, Freeman HS, editors. Colour chemistry—the design and synthesis of organic dyes and pigments. Elsevier Scientific Publications; 1991. p. 85.
[44] Hashimoto Y, Watanabe HK, Degawa M. Gann 1981; 72:921.
[45] Degawa M, Shoji Y, Masuko K, Hashimoto Y. Can Lett 1979;8:71.
[46] Sato K, Poirier LA, Miller JA, Miller EC. Can Res 1966; 26:1678.
[47] Degawa M, Kanazawa C, Hashimoto Y. Carcinogenesis 1982;3:1113.
[48] Miller EC, Kadlubar FF, Miller JA, Pitot HC, Drinkwater NR. Can Res 1979;39:3411.
[49] Mori Y, Hori T, Toyoshi K, Horie M. J Pharm Dyn 1979; 1:192.
[50] Miller JA, Sapp RW, Miller EC. Can Res 1949;9:652.
[51] Degawa M. Miyairis, Hashimoto Y Gann 1978;69:367.
[52] Yahagi T, Degawa M, Seino Y, Matsushima T, Nagao M, Sugimura T, Hashimoto Y. Can Letts 1975;1:91.
[53] Kitagawa T, Pitot HC, Miller EC, Miller JA. Can Res 1979;39:112.
[54] Hashimoto Y, Watanabe H, Degawa M. Gann 1977; 68:373.
[55] Degawa M, Hashimoto Y. Chem Pharm Bull 1976; 24:1485.
[56] Lipinski CA, Lombardo F, Domiay BW, Freeney PJ. Advanced Drug Delivery Reviews 1997;23:3.
[57] Rohrbaugh RH, Jurs PC. Analytica Chimica Acta 1987; 199:99.
[58] Stanton DT, Jurs PC. Anal Chem 1990;62:2323.
[59] Stanton DT, Egolf LM, Jurs PC, Hicks MG. J Chem Inf Comput Sci 1992;32:306.
[60] Kirpichenok MA, Zefirov NS. Zh Org Khim 1987;23:4.
[61] Zefirov NS, Kirpichenok MA, Izmailov FF, Trofimov MI. Dokl Akad Nauk SSSR 1987;296:886.
[62] Fukui K. Theory of orientation and stereoselection. Berlin: Springer-Verlag; 1975.
[63] Strouf O. Chemical pattern recognition. New York: Wiley; 1986.
[64] Sannigrahi AB. Adv Quant Chem 1992;23:301.
[65] Sanderson RT. Chem bonds and bond energy. New York: Acad. Press; 1976, p. 224.
[66] Aldrich handbook of fine chemicals and laboratory equipment, 2000–2001, p. 1159.
[67] Simmon VF. J Natl Cancer Inst 1979;62:893.
[68] Kier LB. J Pharm Sci 1980;69:807.
[69] Boncher D. Information theoretic indices for characterization of chemical structure. New York: Wiley-Interscience; 1983.
[70] Basak SC, Harriss DK, Magnuson VR. J Pharm Sci 1984; 73:429.