J Mol Model (2008) 14:293–302
DOI 10.1007/s00894-008-0270-7

ORIGINAL PAPER

Hologram and 3D-quantitative structure toxicity relationship studies of azo dyes

F. A. Pasha · Muhammad Muddassar · Hwan Won Chung · Seung Joo Cho · Hoon Cho

Received: 25 August 2007 / Accepted: 11 January 2008 / Published online: 7 February 2008

Abstract Amino azobenzenes are important dyes in the food and textile industry but their application is limited due to their mutagenicity. Computational modeling techniques were used to help understand the factors responsible for mutagenicity, and several quantitative structure toxicity relationship (QSTR) models have been derived. HQSTR (hologram QSTR) analyses indicated that different substituents at sites on both rings contribute to mutagenicity. Fragment parameters such as bond (B) and connectivity(C), as well as donor-acceptor (DA)-based model provide significant results (q^2 = 0.59, r^2 = 0.92, $r^2_{predictive}$ = 0.63) explaining these harmful effect. HQSTR results indicated that a bulky group at ring "Y" and small group at ring "X" might help to decrease mutagenicity. 3D-QSTR based on comparative molecular field analyses (CoMFA) and comparative molecular similarity index analyses (CoMSIA) are also in agreement with HQSTR. The 3D QSTR studies reveal that steric and electrostatic field effects have a strong relationship with mutagenicity (for CoMFA: q^2 = 0.51, r^2 = 0.95, $r^2_{predictive}$ = 0.65 and for CoMSIA: q^2 = 0.51, r^2 = 0.93 and $r^2_{predictive}$ = 0.84). In summary, negative groups and steric bulk at ring "Y" and small groups at carbon-3 of ring "X" might be helpful in reducing the mutagenicity of azo dyes.

F. A. Pasha · M. Muddassar · H. W. Chung · S. J. Cho (✉)
Computational Science Center, Future Fusion Technology Division, Korea Institute of Science and Technology,
P.O. Box 131, Cheongryang,
Seoul 130–650, Republic of Korea
e-mail: chosj@kist.re.kr

H. Cho (✉)
College of Engineering, Chosun University,
Gwangju 501–579, South Korea
e-mail: hcho@chosun.ac.kr

Keywords 3D-QSTR · CoMFA · CoMSIA · HQSTR · Azo dyes · Mutagenicity

Introduction

The amino azo-benzenes belong to an important class of dye, and account for approximately 60–70% of all dyes used in the food and textile industry. However, despite their importance in industrial and daily life, azo dyes also shows mutagenicity. Several 4-aminoazobenzene, N-methyl-4-aminoazobenzene and N,N-dimethyl-4-aminoazobenzene derivatives are mutagenic and/or carcinogenic [1–4]. Interestingly, the toxicity 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. 3-Methoxy-4-amino-azobenzene (3-OMe-AAB) is a potent hepato-carcinogen 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 [5, 6]. The biochemical mechanisms responsible for such divergent behavior are not yet fully understood [7–10]. However, the azo linkage is the most labile portion of an azo dye molecule and may easily undergo enzymatic breakdown in mammals, including humans. This reaction is carried out by an enzyme named azo-reductase. In mammals, azo-reductases with different activities are found in various organs such as liver, kidney, lung, heart, brain, spleen and muscle tissues. Liver azo-reductase possesses the greatest enzymatic activity, followed by kidney azo-reductase. After cleavage of the azo-linkage, the component aromatic amines are absorbed in the intestine and excreted in the urine. However, the polarity of azo dyes influences their metabolism [11]. The aromatic amines require metabolic activation for carcinogenicity. The first step involves N-hydroxylation

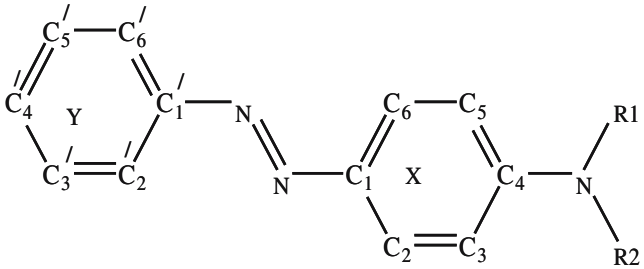

Fig. 1 The basic skeleton of amino azo benzene

and N-acetylation, and the second step involves O-acylation to yield acyloxy amines. These compounds can degrade to form highly reactive nitrenium and carbonium ions. These electrophilic reactants can readily bind covalently to genetic material, namely cellular DNA and RNA [12]. Sulfonation of azo dyes appears to decrease toxicity by enhancing urinary excretion of the dye and its metabolites. A generalized relationship between toxicity and molecular structure might be helpful to identify non-mutagenic dyes.

Table 1 The structure and observed mutagenicity of azo dyes in terms of logTA98 values [17]

No.	R1	R2	Substituent	TA98[a]	logTA98
1	H	H	4′-NEt_2,3-OMe	0.007	−2.15
2	H	H	2-OMe	0.01	−2
3	H	H	4′-OH	0.053	−1.28
4	H	H	3′-Me-4′-OH	0.059	−1.22
5	H	H	4′-OH-2′,3-diMe	0.112	−0.95
6	H	H	-	0.204	−0.69
7	H	H	3′-Me	0.24	−0.62
8	H	H	3-OMe-4′-N(CH2CH2OH)2	0.39	−0.41
9	H	H	3′-CH2OH	0.596	−0.23
10	H	H	3-OH-AAB	0.687	−0.16
11	H	H	3-OCH2CH2OH-4′-N(CH2CH2OH)2	1.052	0.02
12	H	H	3-OCH2CH2OH	1.348	0.13
13	H	H	2′-CH2OH-3-Me	2.012	0.3
14	H	H	4′-OMe	2.3	0.36
15	H	H	2′,3-diMe	2.676	0.43
16	H	H	3-Obu	4.983	0.7
17	H	H	3-OEt	13.802	1.14
18	H	H	3-O-Pro	18.919	1.28
19	H	H	3-OMe	77.065	1.89
20	CH3	H	3′-Me-4′-OH	0.071	−1.15
21	CH3	H	3′-COOH	0.124	−0.91
22	CH3	H	4′-OH	0.14	−0.85
23	CH3	H	-	0.183	−0.74
24	CH3	H	4′-Me	0.283	−0.55
25	CH3	H	3′-Me	0.445	−0.35
26	CH3	H	3′-CH2OH	0.503	−0.3
27	CH3	CH3	3′-Me-40-OH	0.11	−0.96
28	CH3	CH3	-	0.14	−0.85
29	CH3	CH3	3′-COOH	0.201	−0.7
30	CH3	CH3	2-Me	0.22	−0.66
31	CH3	CH3	3′-Me	0.356	−0.45
32	CH3	CH3	3′-CHO	0.383	−0.42
33	CH3	CH3	3′-CH2OAC	0.518	−0.29
34	CH3	CH3	3′-CH2OH	0.601	−0.22
35	H	Ac	3′-Me	0.087	−1.06
36	H	OH	3′-Me-40-OH	0.089	−1.05
37	H	OH	2-OMe	0.11	−0.96
38	CH3	Ac	3′-Me	0.524	−0.28
39	CH3	OH	-	0.65	−0.19
40	CH3	OH	3′-Me	1	0
41	H	OH	N-OH	1.03	0.01
42	CH3	OH	4′-Me	1.132	0.05
43	H	OH	3-OMe	192	2.28

[a] Observed mutagenicity (rev/nmol) in the TA98 *Salmonella typhimurium* bacterial strain with S9 activation

For this purpose, computational chemistry based quantitative structure toxicity relationship (QSTR) and quantitative structure activity relationship (QSAR) techniques are in used [13–15]. The current study deals with the application of hologram QSTR (HQSTR) as well as 3-dimensional-QSTR (CoMFA and CoMSIA) [6, 16] models to establish better and more accurate toxicity–molecular structure relationships for the azo dyes.

Materials and methods

The basic skeleton of amino azo benzene is shown in Fig. 1. The 43 known amino azo benzenes are listed in Table 1 along with their mutagenicity (taken from the literature [17]). The mutagenicity (rev/nmol) of these derivatives was determined in the bacterial strain *S. typhimurium* TA98 with S9 activation (TA98 + S9); this particular bacterial strain is often used to detect mutagens. The dataset was randomly divided into a training set of 33 molecules and a test set of 10 molecules.

Computational details

Hologram QSTR

Hologram QSTR (HQSTR), recently introduced by Tripos (http://www.tripos.com) is a novel QSAR/QSTR method that eliminates the need for determination of 3D structure, putative conformations [18], and molecular alignment. In this method, each molecule in the dataset is divided into structural fragments that are then counted in the bins of a fixed length array to form a molecular hologram. The bin occupancies of the molecular hologram are structural descriptors (independent variables) encoding compositional and topological molecular information. A linear regression equation that correlates variation in structural information (as encoded in the hologram for each molecule) with variation in activity data is derived through PLS regression analysis to produce a QSTR model. Unlike other fragment-based methods, HQSTR encodes all possible molecular fragments (linear, branched, and overlapping). Optionally, additional 3D information, such as hybridization and chirality, may be encoded in the molecular holograms. Molecular holograms are generated in the same manner as hashed fingerprints, where different unique fragments may populate the same holographic bin, allowing the use of a fixed length hologram fingerprint. This hashing procedure emphasizes the importance of patterns of fragment distribution within the hologram bins, which more appropriately represents the nature of chemical structures. Since HQSTR models can be affected by a number of parameters concerning hologram generation, e.g., hologram length, fragment size, and fragment distinction, several combinations of these parameters were considered during the HQSTR modeling runs. Holograms were generated using 4–7 fragment sizes. HQSTR analysis was performed by screening the 12 default series of hologram length values from 51 to 401 bins. SYBYL 7.3 software running on Linux cluster was used, and fragment specification parameters such as atom (A), bonds (B), connectivity (C), hydrogen bond donor–acceptor (DA), etc., were employed to develop HQSTR models.

3D QSTR

Comparative molecular field analysis (CoMFA) and comparative molecular similarity index analysis (CoMSIA) were used for 3D-QSTR. All molecular modeling and 3D-QSTR studies were performed by SYBYL 7.3 with a standard TRIPOS force field on Linux cluster.

Table 2 Regression summery of different hologram quantitative structure toxicity relationship (HQSTR) models. *A* Atom, *B* bond, *C* connections, *H* hydrogen, *DA* donor–acceptor, *n* number of components, q^2 cross validated correlation coefficient, r^2 correlation coefficient, *SE* standard error

No.	Field	n	q^2	r^2	Ensemble	Length	SE	$r^2_{predictive}$
1	A	3	0.28	0.6	0.19	59	0.55	-
2	B	3	0.18	0.5	0.11	83	0.61	-
3	C	4	0.51	0.85	0.41	401	0.34	-
4	DA	4	0.47	0.82	0.34	257	0.37	-
5	A/B	4	0.24	0.66	0.18	61	0.51	-
6	A/C	4	0.42	0.83	0.31	97	0.36	-
7	A/DA	4	0.45	0.8	0.32	59	0.39	-
8	B/C	4	0.45	0.83	0.36	151	0.36	-
9	B/DA	6	0.43	0.93	0.33	307	0.25	-
10	C/DA	4	0.46	0.82	0.37	59	0.37	-
11	A/B/C	4	0.43	0.86	0.31	257	0.33	-
12	A/B/DA	6	0.34	0.9	0.22	83	0.3	-
13	B/C/DA	6	0.59	0.92	0.47	61	0.26	0.63

CoMFA

The steric and electrostatic potential fields for CoMFA were calculated at each lattice intersection of a regularly spaced grid of 2.0 Å. The lattice was defined automatically and is extended 4 units past Van der Waals volume of all molecules in *x*, *y*, and *z* directions. The Van der Waals potential and columbic terms, which represent steric and electrostatic fields, respectively, were calculated using Tripos force field. A distance-dependent dielectric constant was used. An sp^3 carbon atom with a Van der Waals radius of 1.52 Å and a charge of + 1.0 served as the probe atom to calculate steric and electrostatic fields. The steric and electrostatic contributions were truncated to ±30 kcal mol^{-1}, and electrostatic contributions were ignored at lattice intersections with maximum steric interactions. The CoMFA steric and electrostatic fields generated were scaled by the CoMFA standard option given in SYBYL.

Table 3 Observed and predicted activities of azo dye usnig the HQSTR model

No.	logTA98	Hologram	PA_{HQSTR}[a]	Residual
Training set				
1[b]	−2.15	13.255	1.448	−3.598
4	−1.23	7.295	−0.964	−0.266
5	−0.95	8.337	−0.702	−0.248
6	−0.69	5.26	−0.373	−0.317
7	−0.62	5.7	−0.408	−0.212
8	−0.41	21.399	−0.405	−0.005
9	−0.22	6.422	−0.364	0.144
11	0.02	19.21	0.099	−0.079
12	0.13	9.483	0.502	−0.372
13	0.3	7.854	0.046	0.254
14	0.36	6.581	0.005	0.355
15	0.43	6.846	−0.098	0.528
16	0.7	9.191	0.821	−0.121
17	1.14	8.707	1.134	0.006
19	1.89	8.065	1.428	0.462
21	−0.91	7.348	−0.796	−0.114
22	−0.85	7.114	−0.912	0.062
23	−0.74	5.94	−0.414	−0.326
24	−0.55	6.334	−0.352	−0.198
25	−0.35	6.407	−0.449	0.099
27	−0.96	9.341	−0.979	0.019
29	−0.7	8.65	−0.77	0.07
30	−0.66	8.473	−0.523	−0.137
31	−0.45	7.667	−0.422	−0.028
32	−0.42	7.874	−0.629	0.209
33	−0.29	9.129	−0.345	0.055
35	−1.06	7.917	−0.789	−0.271
36	−1.05	9.579	−1.346	0.296
37	−0.96	11.285	−1.266	0.306
39	−0.19	9.67	−0.069	−0.121
40	0	10.15	−0.104	0.104
41	0.01	7.433	0.011	−0.001
43	2.28	12.964	2.445	−0.165
Test set				
2	−2	8.071	−0.893	−1.107
3	−1.28	6.37	−0.871	−0.409
10	−0.16	7.781	0.416	−0.576
18	1.28	9.045	0.824	0.456
20	−1.15	8.053	−1.005	−0.145
26	−0.3	7.181	−0.405	0.105
28	−0.85	7.159	−0.387	−0.463
34	−0.22	8.485	−0.378	0.158
38	−0.28	9.823	−1.037	0.757
42	0.05	10.069	−0.007	0.057

[a] Predicted mutagenicity by model-13
[b] Outlier

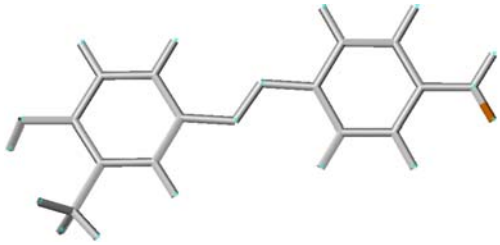

Fig. 2 Fragment contribution to mutagenicity of compound-4

CoMSIA

The reported CoMSIA method is based on molecular similarity indices [19] with the same lattice box as was used for the CoMFA calculations. Molecular similarity is expressed in terms of five different properties, viz., steric, electrostatic, hydrophobic, H-bond donors and acceptors, which were calculated [19] using a C^+ probe atom with a radius of 1 Å placed at a regular grid spacing of 2 Å. CoMSIA similarity indices (A_F) for a molecule j with atoms i at a grid point q are calculated using Eq. 1

$$A^q_{F,K}(j) = -\sum \omega_{prob,k}\omega_{ik}e^{-\alpha r^2}iq \tag{1}$$

Where k represents the following physicochemical properties: steric electrostatic, hydrophobic, H-bond donor and H-bond acceptor. A Gaussian-type distance-dependence was used between grid point q and each atom i of the molecule. A default value of 0.3 was used as the attenuation factor (α). The steric indices are related to the third power of the atomic radii, electrostatic descriptors are derived from atomic partial charges, hydrophobic fields are derived from atom-based parameters [20] and H-bond donor and acceptor indices are obtained by a rule-based method based on experimental results [21].

Partial least square analysis and validation of QSAR models

To derive 3D-QSTR models, the CoMFA and CoMSIA descriptors were used as independent variables and the logTA98 as the dependent variable. The partial least square (PLS) analysis method [22, 23] was used to linearly correlate these CoMFA and CoMSIA descriptors to mutagenicity. The

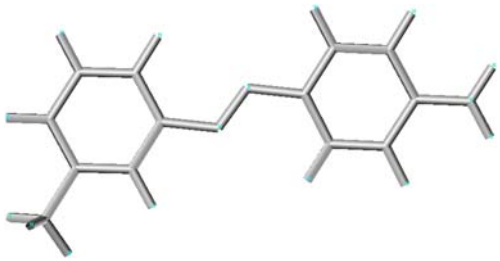

Fig. 3 Fragment contribution to mutagenicity of compound-7

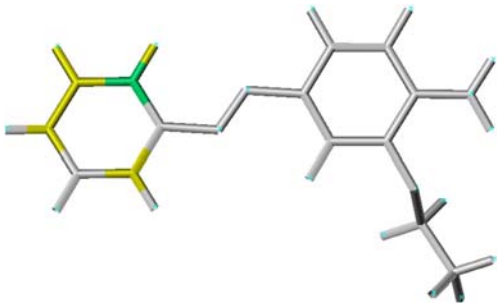

Fig. 4 Fragment contribution to mutagenicity of compound-17

CoMFA cutoff values were set to 30 kcal mol^{-1} for both steric and electrostatic fields, and all fields were scaled by the default options in SYBYL. The cross validation analysis was performed using the leave one out (LOO) method, in which one compound is removed from the dataset and its activity is predicted using the model derived from the rest of the dataset. The cross-validated correlation coefficient (q^2) that resulted in the optimum number of components and lowest standard error of prediction was considered for further analysis and calculated using Eq. 2:

$$q^2 = 1 - \frac{\sum_y \left(y_{pred} - y_{observed}\right)^2}{\sum_y \left(y_{observed} - y_{mean}\right)^2} \tag{2}$$

where, γ_{pred}, γ_{actual} and γ_{mean} are predicted, actual, and mean values of the target property (pIC_{50}), respectively, and PRESS is the sum of predictive sum of squares. The non-cross-validated PLS analyses were performed with a column filter value of 2.0, to reduce analysis time with small effect on the q^2 values. To further assess the robustness and statistical confidence of the derived models, bootstrapping analysis for ten runs was performed.

To assess the predictive power of the 3D-QSTR models derived using the training set, mutagenicity of an external test set of 12 molecules was predicted. The predictive ability of the models is expressed by the predictive r^2 value, which is analogous to cross-validated r^2 (q^2) and is calculated using Eq. 3:

$$r^2_{pred} = \frac{SD - PRESS}{SD} \tag{3}$$

where SD is the sum of the squared deviations between the biological activities of the test set and mean activities of the

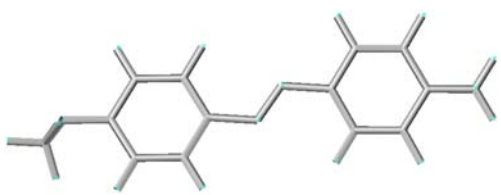

Fig. 5 Fragment contribution to mutagenicity of compound-14

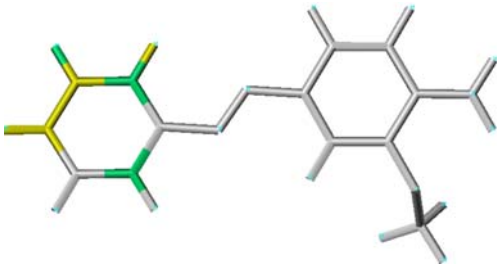

Fig. 6 Fragment contribution to mutagenicity of compound-19

training molecules, and PRESS is the sum of squared deviation between the predicted and observed mutagenicity of the test set molecules calculated using Eq. 4:

$$PRESS = \sum_{y} \left(y_{predicted} - y_{observed}\right)^2 \quad (4)$$

Results

Hologram QSTR

Hologram QSTR models of 43 azo dyes were constructed using fragment parameters atom (A), bonds (B), connectivity (C), hydrogen bond donor and acceptor (DA) in different combinations, e.g., A/B, A/B/C, A/B/C/DA, A/C/DA and B/C/DA, with hologram lengths 53, 59, 61, 71, 83, 97, 151, 199, 257, 307, 353, and 401 bins and fragment size 4–7. The optimum number of components was selected in each case. The regression summary of the different models (1–13) is reported in Table 2. Hydrogen bonding is an important feature of dyes and the DA site varies among different molecules, which might be significant for mutagenicity. On the other hand, the connection parameter "C" is also an important variable. Both the parameter "C"-based model-3 ($q^2 = 0.51$, $r^2 = 0.85$), and parameter "DA"-based model-4 ($q^2 = 0.47$, $r^2 = 0.82$) are significant. In an effort to obtain the best model, combinations of different fragments were considered. Model-13, based on fragment B/C/DA gave better results ($q^2 = 0.59$, $r^2 = 0.92$, ensemble = 0.47, SE = 0.26, component = 6, length = 61and $r^2_{predictive} = 0.63$). The regression summary and mutagenic values of all molecules predicted by model-13 are reported in Tables 2 and 3, respectively. Molecule 1 was identified as an outlier

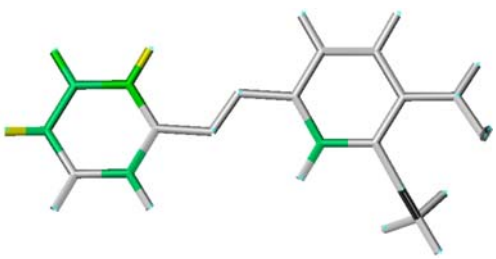

Fig. 7 Fragment contribution to mutagenicity of compound-43

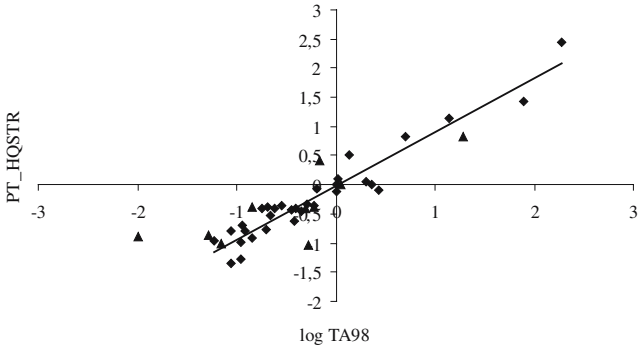

Fig. 8 Trend of observed and predicted mutagenicity of training (n) and test (p) set by hologram quantitative structure toxicity relationship (HQSTR) model-16

as the predicted value of mutagenicity is greater than 1.5 times the interquartile range. The contributions of fragment parameters to mutagenicity are shoen in Figs. 2, 3, 4, 5, 6 and 7, and the trend of observed and predicted mutagenicity is shown in Fig. 8. Compound-4 (Fig. 2) has a similar structure to compound-7 (Fig. 3), except for one additional hydroxyl group at the 4′ position, which makes compound-4 the least mutagenic; the 4-NH_2 group shows a negative contribution to mutagenicity as indicated by the brown color in Fig. 2. Compound-7 and compound-17 are similar in structure but compound-17 has an ethoxy group at position 3 instead of a methyl group at position 3′. This slight variation in structure dramatically changes the mutagenicity; the contribution of the "Y" ring to mutagenicity is shown by the yellow and green color in Fig. 4.

Similarly, compound-14 (Fig. 5) and compound-19 (Fig. 6) are similar in structure but compound-14 carries a methoxy group at position-4′ while compound-19 holds a methoxy group at position-3. The positional variation of methoxy group from 4′ to 3 significantly affects the mutagenicity, and the ring "Y" contributes positively to mutagenicity, as clear from Fig. 6. Compound-19 (Fig. 6) and compound-43 (Fig. 7) have similar structures except for the -OH group at the amino nitrogen. In compound-19, the ring "Y" (Fig. 6) contributes to mutagenicity but in compound-43 both the "X" and "Y" rings contribute to mutagenicity as indicated by the green and yellow color in

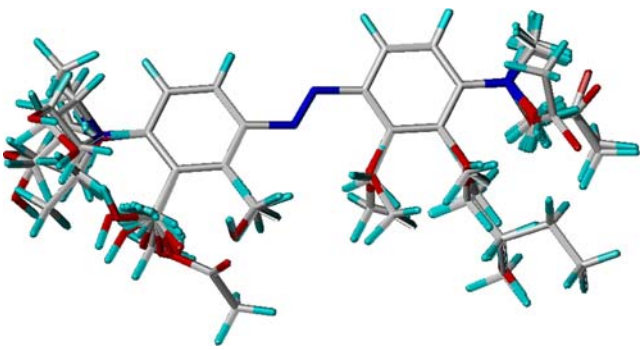

Fig. 9 The aligned structure of all 43 molecules

Table 4 The regression summery of different comparative molecular field analysis (CoMFA) models. *S* steric, *E* electrostatic, *n* number of components, q^2 leave one out cross validated correlation coefficient, r^2 correlation coefficient, *SE* standard error

Model No.	Field	q^2	n	r^2	F	SE	r^2_{bs}	SD	$r^2_{predictive}$
14	S	0.59	6	0.94	61.9	0.23	0.97	0.01	0.48
15	E	0.016	4	0.75	20	0.44	0.81	0.09	0.05
16	0.70S/0.30E	0.51	6	0.95	84.0	0.2	0.99	0.006	0.65

Table 5 Observed and predicted activities of training and test set of compounds by 3D QSTR models

No.	logTA98	PT_{CoMFA} [a]	Residual	PT_{CoMSIA} [b]	Residual
Training set					
1[c]	−2.15	1.189	−3.339	0.412	−2.562
4	−1.23	−1.07	−0.16	−0.83	−0.4
5	−0.95	−0.665	−0.285	−0.674	−0.276
6	−0.69	−0.37	−0.32	−0.342	−0.348
7	−0.62	−0.561	−0.059	−0.572	−0.048
8	−0.41	−0.338	−0.072	−0.399	−0.011
9	−0.22	−0.276	0.056	−0.35	0.13
11	0.02	−0.041	0.061	0.046	−0.026
12	0.13	0.061	0.069	0.012	0.118
13	0.3	0.295	0.005	0.318	−0.018
14	0.36	0.18	0.18	0.02	0.34
15	0.43	0.101	0.329	0.014	0.416
16	0.7	0.629	0.071	0.603	0.097
17	1.14	1.147	−0.007	1.499	−0.359
19	1.89	1.728	0.162	1.682	0.208
21	−0.91	−0.775	−0.135	−0.754	−0.156
22	−0.85	−1.186	0.336	−1.207	0.357
23	−0.74	−0.289	−0.451	−0.488	−0.252
24	−0.55	−0.405	−0.145	−0.504	−0.046
25	−0.35	−0.481	0.131	−0.718	0.368
27	−0.96	−1	0.04	−0.83	−0.13
29	−0.7	−0.631	−0.069	−0.53	−0.17
30	−0.66	−0.809	0.149	−0.594	−0.066
31	−0.45	−0.481	0.031	−0.557	0.107
32	−0.42	−0.5	0.08	−0.618	0.198
33	−0.29	−0.369	0.079	−0.288	−0.002
35	−1.06	−1.103	0.043	−1.23	0.17
36	−1.05	−1.074	0.024	−0.945	−0.105
37	−0.96	−0.864	−0.096	−0.948	−0.012
39	−0.19	0.054	−0.244	0.102	−0.292
40	0	−0.107	0.107	−0.132	0.132
41	0.01	−0.245	0.255	−0.047	0.057
43	2.28	2.464	−0.184	2.271	0.009
Test set					
2	−2	−0.687	−1.313	−1.267	−0.733
3	−1.28	−1.129	−0.151	−1.022	−0.258
10	−0.16	−0.184	0.024	−0.237	0.077
18	1.28	0.574	0.706	1.312	−0.032
20	−1.15	−1.166	0.016	−1.039	−0.111
26	−0.3	−0.395	0.095	−0.632	0.332
28	−0.85	−0.295	−0.555	−0.33	−0.52
34	−0.22	−0.307	0.087	−0.424	0.204
38	−0.28	−0.186	−0.094	−0.671	0.391
42	0.05	−0.019	0.069	0.122	−0.072

[a] Predicted mutagenicity by model-16
[b] Predicted mutagenicity by model-22
[c] Outlier

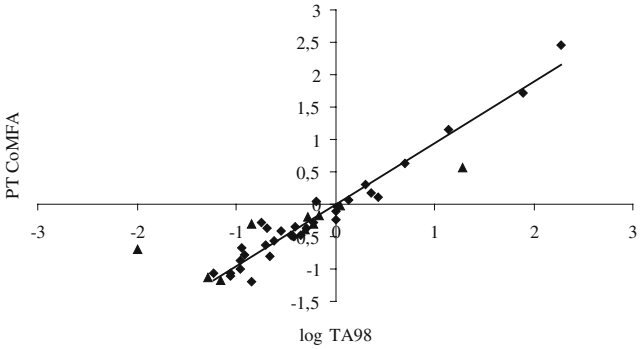

Fig. 10 Trend of observed and predicted mutagenicity of training (n) and test (p) set by comparative molecular field analysis (CoMFA) based model-16

Fig. 7. A general trend appears, i.e., mutagenicity increases on increasing substituents near to carbon 4, and the ring "Y" significantly contributes to mutagenicity. This might be due to high electron density at ring "Y". Further investigation reveals that carbon 4′ is exactly opposite carbon-4. Compound-3 and compound-5 have similar structures but the mutagenicity of compound-5 is high, which might be due to the two additional methyl groups at carbons 2′ and 3. Compound-15 has a similar structure but high mutagenicity to compound-5. This might be due to the lack of substitution at carbon 4′ of compound-15. The fragment maps indicate that substitutions near to carbon 4 are not desirable, but are favorable at ring "Y" to reduce mutagenicity, which is also in agreement with experimental data.

3D QSTR

Molecular alignment

CoMFA and CoMSIA require that 3D structures are aligned according to a suitable conformational template [24]. Due to the lack of structural data supporting a specific 'active' conformation, we have assumed that a compound is active in an energetically minimized conformation. In the present study, the most mutagenic molecule (compound-43) was used as template, initially minimized at Tripos force field [25] with MMFF94 charge by using conjugate gradient method, with a convergence criterion of 0.005 kcal mol^{-1}. A systematic search routine was used in the conformational analysis and all rotatable bonds were searched in 10° increments from 0° to 360°. Conformational energies were computed with electrostatic terms, and low energy conformers were selected for superimposition. The template molecule was further optimized by a semi-empirical PM3 [26, 27] approach and modified for other azo compounds. All molecules were aligned on the template by using the common sub-structure method as shown in Fig. 9 and subsequently used for CoMFA/CoMSIA probe interaction energy calculations.

CoMFA Analysis

CoMFA models were developed with individual fields (steric, electrostatic) and both fields together. The steric field based model-14 showed good correlation, but conjunction with the electrostatic field (0.70S/0.30E) gave even better results (q^2 = 0.51, r^2 = 0.95 and r^2_{bs} = 0.99, and $r^2_{predictive}$ = 0.65) as clear from model-16. The regression summary and activities predicted by model-16 are presented in Tables 4 and 5, respectively. The trend of observed and predicted activities of training and test sets by CoMFA based model-16 is shown in Fig. 10. The study reveals that steric and electrostatic interactions are both important but that the steric contribution is dominant.

CoMFA MAPS

Model-16-based CoMFA contour maps of the most mutagenic compound-43 are shown in Fig. 11. Sterically favored areas (contribution level of 80%) are represented by green polyhedra, and disfavored areas (contribution level of 20%) are represented by yellow polyhedra. Similarly positive charged favored areas (contribution level of 80%) are represented by blue polyhedra and negatively charged favored areas (contribution level of 20%) by red polyhedra. The CoMFA steric map encompasses green contours around carbon-3, and indicates that steric bulk might increase mutagenicity. The yellow polyhedron bordering carbon-2 and 3′ suggests that steric bulk around these positions might reduce mutagenicity. In Fig. 11, red contour between position-R2 and carbon-3 indicates that negatively charged groups favor mutagenicity. Blue contours between position R1 and carbon-6, as well as near to carbon-4′, indicate areas within the lattice where the positive group might increase mutagenicity. Since high mutagenicity is not desirable, the reverse trend should be followed to design the least mutagenic dyes. It is clear from the maps that small substituents at carbon-3 and positive substituents at carbon-4′ are favorable to reduce mutagenicity.

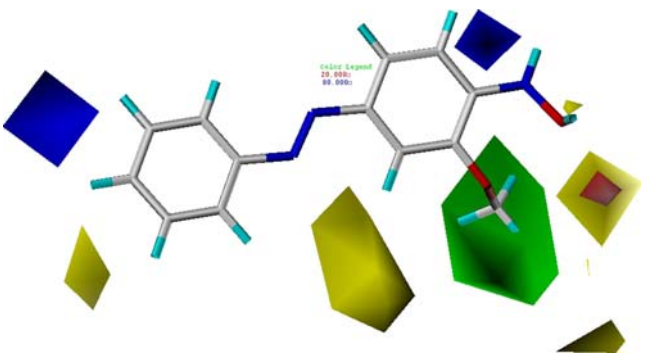

Fig. 11 CoMFA contour maps according to model-16

Table 6 Regression summery of different CoMSIA models

Model No.	Field	q^2	n	r^2	F	SE	r^2_{bs}	SD	$r^2_{predictive}$
17	S	0.45	8	-	-	-	-	-	-
18	E	0.46	4	-	-	-	-	-	-
19	H	0.29	2	-	-	-	-	-	-
20	D	0.12	2	-	-	-	-	-	-
21	A	0.082	10	-	-	-	-	-	-
22	0.48S/0.52E	0.51	6	0.93	55.9	0.24	0.97	0.01	0.84
23	S/H	0.35	6	-	-	-	-	-	-
24	S/D	0.47	9	-	-	-	-	-	-
25	S/A	0.48	8	-	-	-	-	-	-
26	E/H	0.38	4	-	-	-	-	-	-
27	0.54E/0.46D	0.55	9	0.92	27.9	0.28	0.93	0.027	0.75
28	0.52E/0.48A	0.58	6	0.86	24.7	0.35	0.92	0.04	0.50
29	H/D	0.37	9	-	-	-	-	-	-
30	H/A	0.46	9	-	-	-	-	-	-
31	D/A	0.07	2	-	-	-	-	-	-

CoMSIA

CoMSIA models of the same series with five field descriptors, namely steric, electrostatic, hydrophobic hydrogen bond donor and hydrogen bond acceptor, were developed. Descriptors were assessed individually and in different combinations to correlate with mutagenicity. Model-22, based on steric and electrostatic field (0.48S/0.52E), showed a better relationship ($q^2 = 0.51$, $r^2 = 0.93$, $r^2_{bs} = 0.97$) than other models. The test set predictivity ($r^2_{predictive} = 0.84$) of model-22 is acceptable. The mutagenicities predicted by model-22 and the regression summary are reported in Tables 5 and 6, respectively. The trend of observed and predicted toxicities of training and test sets by the CoMSIA-based model is shown in Fig. 12. The results are in accordance with CoMFA and HQSTR results.

CoMSIA MAPS

CoMSIA maps of steric and electrostatic field effects with the most mutagenic molecule of the series (molecule-43) are shown in Fig. 13. As with CoMFA, the sterically favored areas are represented by green polyhedra and disfavored areas are represented by yellow polyhedra. Similarly, positive charged favored areas are represented by blue polyhedra and negatively charged favored areas by red polyhedra.

Position-3 holds a green contour, which indicates that bulky substituents might improve mutagenicity, hence a small group is desirable. A red followed by blue contour also appears around same region, indicating that a negative group with a positive chain is desirable to reduce mutagenicity. The data reveals that, in general, molecules (molecules-11, 12, 16 to 19 and 43) with alkoxy groups at position-3′ posses high mutagenicity. A blue contour also appears around position-4′, indicating that a positive group favors mutagenicity, hence a negative group is desirable.

Discussion

Amino-azobenzenes are important industrial colorants. The ability of azo dye to induce cancer was first described in 1906. Different azo compounds have been shown to be

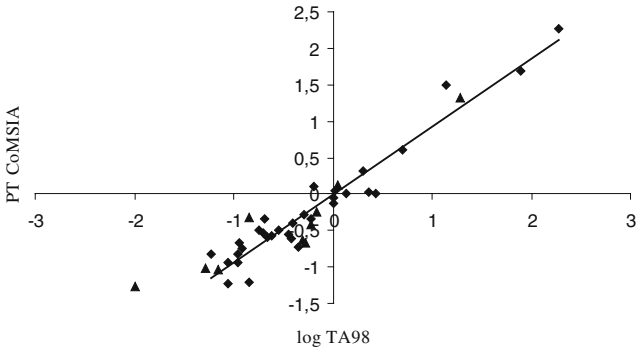

Fig. 12 Trend of observed and predicted mutagenicity of training (n) and test (p) set of dyes by CoMSIA based model-22

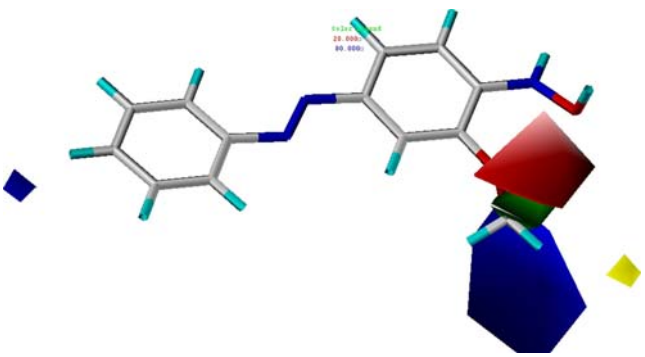

Fig. 13 CoMSIA contour maps according to model-22

definitely mutagenic/carcinogenic [1–3]. However, the detailed biochemical mechanisms responsible for this behavior have not been clearly understood. The pioneering work of Freeman et al. [28, 29] clearly demonstrated that some small structural modifications of these amino-azo-benzene derivatives can reduce or eliminate their mutagenic activity. Many studies have been performed to correlate the relative mutagenic activity of amino-azo-benzene derivatives with various molecular descriptors. Shahin et al. [30] demonstrated that bulky groups at position 4 and 4′ can reduce mutagenicity. Rosenkranz et al. [31] reported the substitution sites responsible for mutagenicity. Chung et al. reviewed mutagenicity and found that azo-reductase cleavage products for different azo dyes were also mutagenic [32, 33]. Enslein et al. [34] suggested the electron donating and withdrawing groups as a possible factor in mutagenicity of azo compounds. A recent study by Maran et al. [35] is based on both the conventional and theoretical molecular descriptors obtained from quantum chemical calculations. The descriptors in the models suggest possible structural factors influencing mutagenicity. The distinct structural factors are the bulkiness of the compound, the charge distribution, hydrogen bonding, and reactivity. The overall mechanism responsible for mutagenicity is not clear but there is strong evidence that mutagenicity of azo dyes is related to substitution sites. Our study indicates the positions of substitution sites as important factors for mutagenicity. Raghuvir et al. [36] recently presented in silico modeling using 3D QSAR techniques, which suggests an immediate next direction for our work. We hope that, by using advanced in silico modeling, our studies will enable us to design new effective and non-mutagenic dyes in the near future. The least mutagenic compound-1 (identified outlier) cannot be included in any of the proposed three models as its predicted mutagenicity falls more than 1.5 times outside the interquartile range. This may result from other possible factors such as metabolic activation [37].

Conclusions

HQSTR results indicate that the non-substituted ring “Y” contributes significantly to mutagenicity. Inspection of the data reveals that the contribution of ring “Y” to mutagenicity increases as substitution increases at ring “X” near to carbon-4. Consequently, small substituents at positions R1 and at carbon-3, and bulky substituents at ring “Y”—particularly at carbon-4′—are desirable for low mutagenicity. The 3D QSTR results also indicate that small groups at carbon-3 of ring “X”, and a bulky and negative group at carbon-4′ of ring “Y” might be helpful in reducing mutagenicity. Substitution at ring “Y” might also be favorable in reducing mutagenicity.

Acknowledgments This work was supported by KIST Linux Supercomputers. We also thank Jung Soo Oh for help.

References

1. Kimura T, Kodama M, Nagata C (1982) Carcinogenesis 3:1393–1396
2. Miller JA, Miller EC (1961) Cancer Res 21:1068–1072
3. Mori Y, Niwa T, Toyoshi K, Nagai H, Koda A, Kawada K, Ojima A, Takahashi Y (1980) Carcinogenesis 1:533–535
4. Sugiura K, Halter CR, Kensler CJ, Rhoads CP (1945) Cancer Res 5:235–238
5. Kojima M, Degawa M, Hashimoto Y, Tada M (1991) Biochem Biophys Res Commun 179:817–823
6. Cramer RD, Patterson DE, Bunce JD (1988) J Am Chem Soc 110:5959–5967
7. Ashby J, Lefevre PA, Callander RD (1983) Mutat Res 116:271–279
8. Chung KT (1983) Mutat Res 114:269–281
9. Hashimoto Y, Degawa M, Watanabe HK, Tada M (1981) Gann 72:937–943
10. Lin JK, Miller JA, Miller EC (1975) Cancer Res 35:844–850
11. Pinheiro HM, Touraud E, Thomas O (2004) Dyes Pigm 61:121–139
12. Brown MA, Devito SC (1993) Crit Rev Environ Sci Technol 23:249–324
13. Roy DR, Sarkar U, Chattaraj PK, Mitra A, Padmanabhan J, Parthasarathi R, Subramanian V, Van Damme S, Bultinck P (2006) Mol Divers 10:119–131
14. Pasha FA, Srivastava HK, Srivastava A, Singh PP (2007) Qsar & Combinatorial Science 26:69–84
15. Pasha FA, Srivastava HK, Singh PP (2005) Bioorganic Med Chem 13:6823–6829
16. Klebe G, Abraham U (1999) J Comput-Aided Mol Des 13:1–10
17. Garg A, Bhat KL, Bock CW (2002) Dyes Pigm 55:35–52
18. Tong W, Lowis DR, Perkins R, Chen Y, Welsh WJ, Goddette DW, Heritage TW, Sheehan DM (1998) J Chem Inf Comput Sci 38:669–77
19. Klebe G, Abraham U, Mietzner T (1994) J Med Chem 37:4130–4146
20. Viswanadhan VN, Ghose AK, Revankar GR, Robins RK (1989) J Chem Inf Comput Sci 29:163–172
21. Klebe G (1994) J Mol Biol 237:212–35
22. Geladi P, Xie YL, Polissar A, Hopke P (1998) J Chemom 2:231
23. Wold S, Ruhe A, Wold H, Dunn WJ (1984) Siam J Sci Statist Comput 5:735–743
24. Kim KH, Greco G, Novellino E (1998) Perspect Drug Discov Des 12:257–315
25. Clark M, Cramer RD, Vanopdenbosch N (1989) J Comput Chem 10:982–1012
26. Stewart JJP (1989) J Comput Chem 10:221–264
27. Stewart JJP (1989) J Comput Chem 10:209–220
28. Freeman HS, Esancy MK, Esancy JF, Claxton LD (1991) Chemtech 21:438–445
29. Freeman HS, Kim SD, Gilbert RD, Mcgregor R (1991) Dyes Pigm 17:83–100
30. Shahin MM (1989) Mutagenesis 4:115–25
31. Rosenkranz HS, Klopman G (1989) Mutat Res 221:217–34
32. Chung KT, Fulk GE, Andrews AW (1981) Appl Environ Microbiol 42:641–648
33. Chung KT, Fulk GE, Egan M (1978) Appl Environ Microbiol 35:558–562
34. Enslein K, Borgstedt HH (1989) Toxicol Lett 49:107–21
35. Maran U, Sild S (2003) Artif Intell Rev 20:13–38
36. Pissurlenkar RR, Shaikh MS, Coutinho EC (2007) J Mol Model 13:1047–71
37. Chung KT, Cerniglia CE (1992) Mutat Res 277:201–20