Hologram Quantitative Structure Activity Relationship (HQSAR) Study of Mutagen X

Seung Joo Cho

Life Science Division, Korea Institute of Science and Technology, Seoul 130-650, Korea. E-mail: chosj@kist.re.kr
Received September 15, 2004

MX and its analogs are synthesized and modeled by quantitative structure activity relationship (QSAR) study including comparative molecular field analysis (CoMFA). As a result, factors affecting this class of compounds have been found to be steric and electrostatic effects. Because hologram quantitative structure activity relationship (HQSAR) technique is based on the 2-dimensional descriptors, this is free of ambiguity of conformational selection and molecular alignment. In this study we tried to include all the data available from the literature, and modeled with the HQSAR technique. Among the parameters affecting fragmentation, connectivity was the most important one for the whole compounds, giving good statistics. Considering additional parameters such as bond specification only slightly improved the model. Therefore connectivity has been found to be the most appropriate to explain the mutagenicity for this class of compounds.

Key Words : Mutagen X, HQSAR, QSAR, Molecular Modeling

Introduction

Chlorine bleaching disinfects our drinking water by reducing the water-mediated diseases. However, some of the bi-products caused by this disinfection process are highly mutagenic.[1] Although how MX is produced in water is not clearly understood,[2-4] MX is a potent mutagen ever tested in Ames test with test strain TA100.[5] The mutagenicity of MX has been reported 3430-13800 induced reversants per nanomole in the Ames assay without S9 mix. This unusual high mutagenicity attracted considerable attention from many researchers.[6-10] Until recently, MX was assumed to pose little carcinogenic risk due to its low exposure, high reactivity and short residence time.[11] But recent identification of DNA adducts[12,13] and evidence of carcinogenicity along the gastro-intestinal lining in rodents following MX exposure has heightened concern for this class of chemicals. MX can alter the metabolic pathway when it is administered in rats in high dosage.[14] It is also found to induce apoptosis of HL-60 cells.[15] A relatively large number of MX analogs have been synthesized,[16,17] tested for mutagenicity,[18-20] subject to many experimental studies. As a result, the resultant MX analogs show wide range of mutagenicity.[21-34] They are modeled by structure-activity relationship methods.[35-37] In spite of this multitude of studies, basic questions concerning the nature of the reactive species and the mechanism of interaction of these compounds with DNA to produce their remarkable mutagenic potency in SAL TA100 remain unresolved. MX exists as an equilibrium mixture of both ring and open form in water as shown in Figure 1. The relative concentration of ring and open form depends heavily on the pH of the solution.[35] If the aqueous solution is highly acidic, the ring form is dominant species. At pH 5.5 the ratio of ring form and open form is 1 : 1. The relative concentration of open form becomes high as the solution gets more basic. This is a fast equilibrium process.[38] To study factors affecting the mutagenicity, there have been a few quantitative structure activity relationship (QSAR) studies. The structural and electronic properties were calculated using the semi-empirical AM1 (Austin Model 1) method. The lowest unoccupied frontier orbital (LUMO) was found to be important by using this quantum mechanical method.[22,23] This may imply that MX acts as an electron acceptor. In particular, LUMO electron density and partial charge of the C3 were correlated with mutagenicity. Electron density near C3 also showed negative linear dependency by NMR study. Comparative molecular field analysis (CoMFA) results indicated that the steric properties of MX analogs with their electron-accepting ability, explain their mutagenic activity almost completely.[39] However, these studies are based on a few reports and some of the structurally relevant compounds were never considered for QSAR studies. In this study, we tried to include all the data available from the literature and summarized in Table 1. At a glance, as the degree of halogen substitution increases, the mutagenicity also increases.

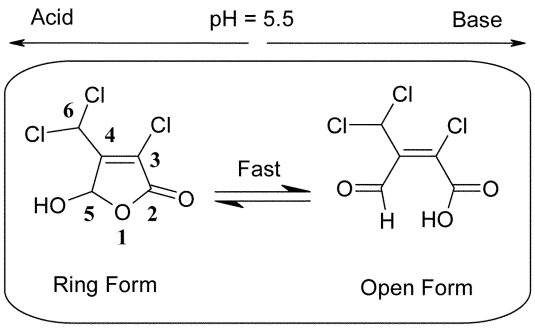

Figure 1. Two forms of Mutagen X in equilibrium.

The mutagenicity of MX is the average value of 9 different studies. All the activity values are within the order of magnitude (3430-13800). Thus the average value is considered as highly reliable. The whole set comprises of 37 compounds. The range of activity is fairly well spread for any particular family as well as for the whole set. All the

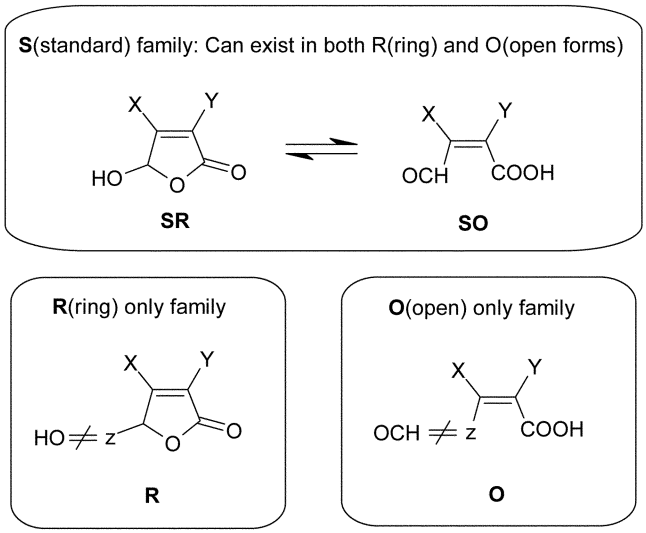

Figure 2. Three MX families.

compounds have unsaturated acidic moiety as the core structure. This structural resemblance may imply that these compounds induce mutagenicity with the same mechanism. At a glance, as the degree of chlorine or bromine substitution increases, the mutagenicity also increases. The compounds collected from the various reports[21-34] are categorized into three groups as shown in Figure 2. Compounds in standard family (**S**) contain the structure of 5-hydroxy-2(5H)-furanone. These compounds are capable of inter-conversion between hydroxyl ring form and aldehyde open form like MX. If an analog has a ring form and does not have 5-hydroxyl group, then it cannot be converted into corresponding open form. Therefore it belongs to ring family (**R**). On the other hand, if an MX analog is an open form and does not have aldehyde group, then it cannot be closed into corresponding ring form, belonging to open family (**O**). This open family has never been explicitly included in the previous QSAR studies. To use data in Table 1 for modeling, the compounds which belong to standard family must be represented either **SR** or **SO** form (Figure 2).

Methods

Quantitative structure-activity relationships (QSARs) are important tools to understand why the active compounds exhibit certain biochemical activities. The challenge now is to improve the accuracy and predictability of QSAR model by taking into account the structural and physicochemical features of the concerned compounds. One of the most widely used tools in 3D QSAR study is comparative molecular filed analysis (CoMFA).[40,41] CoMFA is based on the assumption that changes in the biological activity correlate with changes in the steric and electrostatic fields of molecules. However, it requires some knowledge or hypothesis regarding the functionally active conformations of the molecules and molecular superposition as a prerequisite for structural alignment. Moreover, care must be taken in constructing molecular alignments because slight differences in alignment can lead to wide variation in the resultant CoMFA model. In addition, this alignment process is very time consuming. In the study of MX and its analogs, CoMFA has been used as a tool, but the structures included

Table 1. The Mutagenicity of MX analogs

	X	Y	Z	ln(TA100)	N
Standard Family					
S1 (MX)	$CHCl_2$	Cl	OH	8.62	9
S2 (BMX2)	$CHBr_2$	Cl	OH	8.61	1
S3 (BMX3)	$CHBr_2$	Br	OH	6.41	2[a]
S4 (CMCF)	CH_2Cl	Cl	OH	6.37	5
S5 (BMBF)	CH_2Br	Br	OH	6.04	1
S6 (MCA)	Cl	Cl	OH	1.87	6[a]
S7 (MBA)	Br	Br	OH	1.71	1
S8	CH_2Cl	H	OH	1.35	3
S9 (MBF)	CH_3	Br	OH	0.41	1
S10 (MCF)	CH_3	Cl	OH	0.21	4
S11	H	Cl	OH	−1.61	1
S12 (MF)	CH_3	H	OH	−3.51	2
Ring Family					
R1	$CHBr_2$	Cl	OCH_3	8.65	1
R2	$CHCl_2$	Cl	OCH_3	8.65	1
R3	$CHBr_2$	Cl	H	5.20	1
R4	$CHBr_2$	Br	H	4.86	1
R5 (RMX)	$CHCl_2$	Cl	H	4.54	6
R6	CH_2Br	Br	H	2.11	1
R7	CH_2Cl	Cl	H	1.70	4
R8	CH_2Cl	Br	H	1.37	1
R9	CH_2Br	Cl	H	1.37	1
R10	Cl	Cl	OCH_3	0.99	1
R11	CH_3	Cl	OC_2H_5	0.74	1
R12	Br	Br	H	0.17	1
R13	H	Cl	OC_2H_5	−0.22	1
R14	CH_3	Cl	H	−0.78	2[b]
R15	Cl	Cl	H	−0.62	2
R16	CH_2Cl	H	H	−1.59	3[a]
R17	$CHCl_2$	H	H	−2.41	2[b]
Open Family					
O1 (BA-4)	$CHCl_2$	Cl	$CHCl_2$	7.11	1
O2 (BA-3)	CH_2Cl	Cl	$CHCl_2$	5.48	1
O3 (ox-mCMF)	CH_2Cl	H	COOH	0.47	1
O4 (ox-CMCF)	CH_2Cl	Cl	COOH	−0.92	2
O5 (BA-1)	CH_2Cl	H	$CHCl_2$	−1.20	1
O6 (BA-2)	$CHCl_2$	H	$CHCl_2$	−1.20	1
O7 (ox-MCF)	CH_3	Cl	COOH	−1.27	1
O8 (ox-MCA)	Cl	Cl	COOH	−2.12	1

See reference 11 for the data in this table. Words in parenthesis are common names. X, Y and Z are substituents for MX analogs (Figure 2). ln(TA100) is the natural log for experimental values (rev/nm in Ames test). N is the number of reports that have mutagenicity data. When there are more than two reports, after the logarithms have been taken, the values are averaged, and the resultant values are listed in this table. a) The maximum value is more than one order of magnitude larger than the minimum value. b) One of the reports indicates that the compound is not mutagenic and logarithms are taken for remaining value. For S3, the reported values are 4.68 and 7.71. For S6, the average without maximum and minimum values is 1.41 (standard deviation is 1.24.).

Table 2. HQSAR results for various sets and methods

Data Set	q^2	r^2	Ave[a]	q^2	r^2	Ave	q^2	r^2	Ave	q^2	r^2	Ave	q^2	r^2	Ave
	A			**B**			**C**			**H**			**A, B**		
SR	0.812	0.918	0.788	0.911	0.967	0.900	0.955	0.981	0.943	*	*	*	0.799	0.911	0.753
SO	0.784	0.895	0.725	0.915	0.968	0.902	0.954	0.981	0.941	*	*	*	0.797	0.906	0.722
R	0.889	0.993	0.839	0.916	0.986	0.899	0.897	0.993	0.882	0.236	0.542	0.226	0.854	0.980	0.824
O	0.840	0.994	0.716	0.560	0.784	0.493	0.872	0.996	0.821	*	*	*	0.780	0.991	0.710
R, O	0.822	0.961	0.780	0.872	0.978	0.845	0.913	0.981	0.878	*	*	*	0.835	0.934	0.793
R, SR	0.865	0.937	0.853	0.903	0.929	0.895	0.903	0.968	0.882	0.102	0.344	0.088	0.858	0.939	0.837
O, SO	0.542	0.868	0.473	0.708	0.868	0.670	0.788	0.918	0.761	*	*	*	0.698	0.939	0.604
R, SR, O	0.821	0.899	0.799	0.857	0.891	0.848	0.881	0.965	0.860	0.079	0.195	0.074	0.833	0.920	0.808
R, SO, O	0.732	0.889	0.676	0.798	0.907	0.774	0.851	0.939	0.831	0.033	0.298	0.011	0.752	0.898	0.715
	A, C			**A, H**			**B, C**			**B, H**			**C, H**		
SR	0.785	0.875	0.721	0.921	0.992	0.816	0.952	0.981	0.943	*	*	*	*	*	*
SO	0.764	0.861	0.688	0.831	0.982	0.763	0.952	0.980	0.938	*	*	*	*	*	*
R	0.896	0.989	0.847	0.793	0.972	0.743	0.906	0.985	0.888	0.211	0.542	0.192	0.271	0.565	0.202
O	0.882	0.995	0.824	0.534	0.879	0.412	0.870	0.997	0.839	*	*	*	*	*	*
R, O	0.907	0.989	0.821	0.808	0.961	0.760	0.916	0.988	0.901	0.099	0.416	0.080	0.158	0.489	0.107
R, SR	0.854	0.951	0.842	0.881	0.962	0.853	0.891	0.965	0.877	0.101	0.340	0.062	0.106	0.355	0.082
O, SO	0.680	0.905	0.630	0.645	0.954	0.553	0.764	0.897	0.742	*	*	*	*	*	*
R, SR, O	0.882	0.966	0.824	0.840	0.932	0.800	0.901	0.971	0.869	0.132	0.238	0.100	0.099	0.369	0.076
R, SO, O	0.767	0.931	0.723	0.753	0.919	0.705	0.837	0.933	0.816	0.037	0.326	0.017	0.125	0.327	0.056
	A, B, C			**A, B, H**			**A, C, H**			**B, C, H**			**A, B, C, H**		
SR	0.801	0.888	0.715	0.865	0.996	0.783	0.839	0.980	0.761	*	*	*	0.808	0.988	0.768
SO	0.792	0.878	0.699	0.885	0.999	0.777	0.835	0.999	0.777	*	*	*	0.886	0.990	0.781
R	0.888	0.988	0.853	0.844	0.970	0.775	0.861	0.987	0.762	0.259	0.517	0.203	0.835	0.972	0.767
O	0.881	0.995	0.819	0.683	0.987	0.582	0.798	0.994	0.711	*	*	*	0.821	0.995	0.751
R, O	0.889	0.987	0.841	0.833	0.957	0.798	0.874	0.984	0.804	0.124	0.483	0.103	0.887	0.981	0.833
R, SR	0.857	0.935	0.836	0.882	0.950	0.850	0.875	0.963	0.839	0.094	0.312	0.072	0.896	0.966	0.848
O, SO	0.668	0.921	0.647	0.697	0.934	0.637	0.748	0.927	0.660	*	*	*	0.698	0.925	0.659
R, SR, O	0.853	0.947	0.820	0.872	0.944	0.817	0.873	0.958	0.826	0.113	0.228	0.084	0.890	0.968	0.855
R, SO, O	0.778	0.937	0.734	0.750	0.932	0.697	0.796	0.950	0.690	0.094	0.353	0.054	0.767	0.937	0.701

Data sets (SR: Standard family of ring form, SO: Standard family of open form, R: ring family, O: open family, R,SR: ring family and standard family of ring form, etc.), **Statistical parameters** (q^2: crossvalidation by LOO procedure, r^2: correlation, **Ave**: average value of Ensemble q^2). **Fragment Options** (**A**: atom information is considered, **B**: bond information, **C**: connectivity, H: Hydrogen, **A,B**: atom information and bond information, etc.). * means that data with no significant model, *i.e.*, q^2 is less than zero. Fingerprints were generated for all substructures between 4 and 7 atoms in size for all molecules.

in the previous studies only covers ring family and ring form of standard family. This might come from the ambiguity of conformation selection and alignment, *i.e.*, if open family is included in the data set, it will be more arbitrary to align them. On the other hand, hologram QSAR (HQSAR), a newly developed QSAR technique, relates biological activity to structural fragments. HQSAR eliminates the need for generation of 3D structures, putative binding conformations and molecular alignments. For standard family, we do not know either ring form or open form is responsible for the mutagenicity. Therefore we need to consider both cases of standard family. As explained previously, standard family of ring form is annotated **SR**, and of open form, **SO**. We also considered the various combinations of three families. For example, **R**, **SR** in Table 2 means that the data set is a union of ring family and standard family of open form. Naturally when we consider the three families altogether, the combination can be either **R**, **SR**, **O** or **R**, **SO**, **O**. Fingerprints were generated for all substructures between 4 and 7 atoms in size for all molecules. The substructure fingerprints were then hashed into hologram bins with lengths of 53, 59, 61, 71, 97, 151, 199, 275, 307, 353, 401, 997. These prime numbers were chosen to minimize the fragment collision problem. For each hologram length, various combinations of fragment distinction parameters and fragment generation parameter were considered and the results are listed in Table 2. Each molecule in the dataset is broken down into structural fragments. The parameters of the fragments are then hashed into Molecular Hologram. Unique fragments are always hashed into the same bin. Atom distinction parameter provides the ability to distinguish between fragments based on differences in their elemental types. *i.e.*, NH_2

(ammonia), PH_3 (phosphine) and CH_3 (methyl group) are distinguished upon fragmentation. Bond distinction parameter provides the ability to distinguish between fragments based on differences in their bond types, *i.e.*, C-C-H (in ethane), C=C-H (in ethylene). Connection parameter allows the holograms to retain information about the hybridization states of the atoms in the fragments, *i.e.*, in ethylene glycol ($OHCH_2CH_2OH$) the two carbons are sp3 hybridized, while in acetic acid ($CH_3C(=O)OH$) the first carbon is sp3 hybridized and the second is sp2 hybridized. The connection flag therefore allows hybridization information to be included in the hologram. Hydrogen parameter provides the ability to distinguish between fragments based on whether or not hydrogen atoms are included, *i.e.*, C_6H_6 (benzene) and C_5H_5N (pyridine) are identical if hydrogen atoms and atom distinctions are ignored. Fragment distinction parameters comprise of information on atoms (**A**), bonds (**B**), and connections (**C**). In generating fragments, both cases of hydrogen (**H**) atom inclusion and exclusion were studied. For each hologram length, there could be a model. The collection of these models comprises the ensemble. For statistical parameter, q^2 (crossvalidation by leave-one-out procedure), r^2, Ave (average value of the ensemble q^2) were considered. The symbol * was used when the best value of q^2 was less than zero for the ensemble. Thus the symbol * implies no significant model was found for the entire hologram lengths used.

Results and Discussion

All the models containing symbol * have the parameter H, without parameter **A**. This implies that we need to consider atom specification if we include hydrogen for fragment generation. This may come from the fact that the halogen atom substitution effects cannot be seen from the resultant fragments. If **A** is used along with **H**, the statistical parameters indicate reasonable models (q^2: 0.53-0.92). Among the four parameters (**A**, **B**, **C**, **H**), when considered alone, **C** gave the best results (q^2: 0.85-0.95). Therefore, connectivity is the most important factor. Adding parameter **B** gives only small improvement over some datasets. As previously noted, standard family can exist either ring or open form. All the individual families gave good statistical parameters. Because we don't know which form really represents standard family, we considered both forms (**SR** and **SO**). Both **SR** and **SO** gave highly predictive q^2 and r^2. The other datasets such as **R** and **O** also gave reasonable results, with somewhat smaller values of q^2 and r^2. When we combine the two sets, the values of q^2 went down dramatically with **O**, **SO** set. This implies that the union of standard family and open family is less homogeneous than the union of standard family and ring family. For the whole set, we have two possible combinations, and the set of **R**, **SR**, **O** gave better results than that of **R**, **SO**, **O** in most cases. The best statistics for **R**, **SR**, **O** is $q^2 = 0.90$ and $r^2 = 0.97$ (**B** and **C**), while for **R**, **SO**, **O** $q^2 = 0.85$ and $r^2 = 0.93$ (**B**). It is interesting to note that if parameter **A** is considered, the statistical parameter gets worse in general. In the previous study of CoMFA, steric parameter was important, which implies that bromine substitution should give higher mutagenicity than chlorine substitution. The substitution is important as reported in previous studies, but the results also implies that the atom specification is not so important. In other words, whether chlorine or bromine is substituted is not so important, rather the degree of halogen substitution is important. Fingerprints were generated for all structures between 4 and 7 atoms in size for all molecules. We have tried to find a better model by varying the range of fragment length. We have fully covered the fragment length (2-8), resulting in 28 combinations. In Table 3, the q^2 and r^2 values are listed along with the range of fragment length. We could not find better model using other fragment length range rather than default range (4-7). Therefore following discussion is based on default fragment lengths. In Table 4, the predicted values and crossvalidated predicted values are listed and compared with observed values. S3 gave the largest difference not only between observed and predicted values also between observed and crossvalidated predicted values. Actually the value of S3 is from two different studies. When we look into the data carefully, the two values (rev/nmol) are 2880 and 129. Therefore the two values differ significantly. If we remove the smaller value based on the model, then the observed value would be 7.97 which is closer to the predicted value and crossvalidated predicted value. If we use this value for S3 instead of the value in table 1, the $q^2 = 0.919$ and $r^2 = 0.980$, and ensemble average q^2 would be 0.888 which give good statistical parameters. The contribution to activity of each atom in a given molecule in

Table 3. The Effect of Fragment Length Variation

Short Long	2	3	4	5	6	7	8
2	0.696(0.821)						
3	0.737(0.861)	0.738(0.886)					
4	0.766(0.903)	0.763(0.910)	0.744(0.899)				
5	0.859(0.955)	0.865(0.957)	0.878(0.961)	0.884(0.962)			
6	0.881(0.961)	0.884(0.962)	0.894(0.966)	0.890(0.963)	0.891(0.967)		
7	0.898(0.971)	0.897(0.970)	0.901(0.971)	0.900(0.971)	0.889(0.935)	0.893(0.952)	
8	0.889(0.939)	0.888(0.944)	0.889(0.945)	0.890(0.948)	0.896(0.957)	0.896(0.948)	0.898(0.947)

Fragment length variation was performed on the whole set (**R**, **SR**, **O**) for the condition that gave the best statistical result (Table 2). The range of fragment length is from short to long. The best predictive value (q^2) for each fragmentation method is listed. Data in the parenthesis are r^2.

Table 4. Residuals for the best model for the whole set (**R**, **SR**, **O**)

	Observed	Predicted	Residual	cv predicted	cv residual	N
Standard Family						
S1 (MX)	8.62	8.33	−0.29	8.30	−0.32	9
S2 (BMX2)	8.61	8.33	−0.28	8.30	−0.31	1
S3 (BMX3)	6.41	8.33	1.92	9.05	2.64	2[a]
S4 (CMCF)	6.37	5.34	−1.03	4.95	−1.42	5
S5 (BMBF)	6.04	5.34	−0.70	5.06	−0.98	1
S6 (MCA)	1.87	1.24	−0.63	1.00	−0.87	6[a]
S7 (MBA)	1.71	1.24	−0.47	1.08	−0.63	1
S8	1.35	0.32	−1.03	−0.74	−2.09	3
S9 (MBF)	0.41	0.40	−0.01	0.60	0.19	1
S10 (MCF)	0.21	0.40	0.19	0.68	0.47	4
S11	−1.61	−0.96	0.65	0.12	1.73	1
S12 (MF)	−3.51	−2.55	0.96	−1.09	2.42	2
Ring Family						
R1	8.65	8.78	0.13	8.74	0.09	1
R2	8.65	8.78	0.13	8.74	0.09	1
R3	5.20	4.29	−0.91	3.99	−1.21	1
R4	4.86	4.29	−0.57	4.10	−0.76	1
R5 (RMX)	4.54	4.29	−0.25	4.19	−0.35	6
R6	2.11	2.14	0.03	2.12	0.01	1
R7	1.70	2.14	0.44	2.20	0.50	4
R8	1.37	2.14	0.77	2.27	0.90	1
R9	1.37	2.14	0.77	2.27	0.90	1
R10	0.99	1.55	0.56	2.29	1.30	1
R11	0.74	0.56	−0.18	0.27	−0.47	1
R12	0.17	−0.04	−0.21	−0.57	−0.74	1
R13	−0.22	−0.30	−0.08	−0.98	−0.76	1
R14	−0.78	−1.53	−0.75	−1.79	−1.01	2[b]
R15	−0.62	−0.04	0.58	0.07	0.69	2
R16	−1.59	−1.73	−0.14	−1.85	−0.26	3[a]
R17	−2.41	−1.66	0.75	−1.02	1.39	2[b]
Open Family						
O1 (BA-4)	7.11	7.20	0.09	7.72	0.61	1
O2 (BA-3)	5.48	5.56	0.08	5.14	−0.34	1
O3 (ox-mCMF)	0.47	0.53	0.06	−2.12	−2.59	1
O4 (ox-CMCF)	−0.92	−0.54	0.38	0.53	1.45	2
O5 (BA-1)	−1.20	−1.20	0.00	−1.32	−0.12	1
O6 (BA-2)	−1.20	−1.60	−0.40	−1.17	0.03	1
O7 (ox-MCF)	−1.27	−1.59	−0.32	−1.89	−0.62	1
O8 (ox-MCA)	−2.12	−2.36	−0.24	−1.41	0.71	1

N is the number of reports that have mutagenicity data. When there are more than two reports, after the logarithms have been taken, the values are averaged, and the resultant values are listed in this table. [a]The maximum value is more than one order larger than the minimum value in magnitude. [b]One of the reports indicate that the compound is not mutagenic and logarithms are taken for remaining value.

the dataset is calculated as follows: The contribution to activity of each atom in the fragment is taken as the partial least squares (PLS) coefficient for that fragment divided by the number of atoms in the fragment. Thus, all atoms are assumed to contribute equally to the activity of a given fragment. If a fragment is found twice it is counted twice. The total contribution to activity of a selected atom is obtained by summation of the individual atomic contributions from the fragments containing that atom.

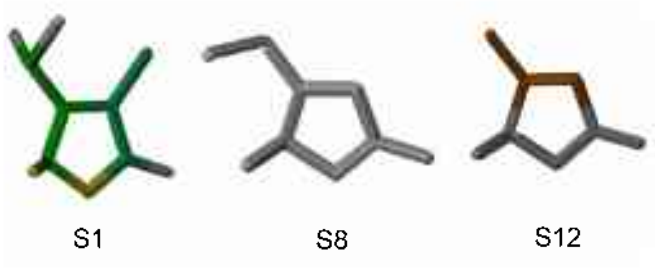

Figure 3. Atomic contributions to mutagenicity: Green color denotes the greatest contribution on mutagenicity while red signifies least contribution.

Using the corrected model, we plot the atom contribution of S1, S8, and S12 in Figure 3. The green color means high mutagenicity and yellow means somewhat less high mutagenicity. Gray color signifies average contribution on mutagenicity. Red color indicates negative contribution. As the degree of substitution decreases, the mutagenicity also decreases (*i.e.*, the colors of C3, C4, C6 change from green to gray to red.). This phenomenon was generally found for the whole set. In particular, halogen substitution on the particular positions (C3, C4, C6) would increase mutagenicity.

Conclusion

The mutagenicity of MX analogs was previously reported to negatively correlate with the energy level of LUMO (17 compounds).[37] It seems reasonable to consider MX analogs as electrophiles, thus reacting with electron-rich DNA, then inducing mutagenesis. Steric factor was the most important with CoMFA (21 compounds).[39] The importance of steric factor might indicate the degree of halogen substitution, since halogen atoms are much bigger than hydrogen. In this work, the most important parameter is connectivity (39 compounds). The degree of halogen substitution must be related with this connectivity parameter. Although the generation methods for these descriptors are different (Quantum mechanical, 3D Lennard Jones potential, 2D connectivity), they gave reliable statistical parameters. Moreover, physical origin of these parameters might be the same. That is to say, as the degree of halogen substitution increases, the connectivity increases (connectivity parameter of HQSAR), the volume increases (steric factor of CoMFA), the molecule gets more electronegative (LUMO), and as a result, the molecules gets more reactive against electron-rich DNA. Because other descriptors (LUMO energy level and CoMFA steric field parameter) are conformation dependent, the connectivity parameter in HQSAR can be used most conveniently.

References

1. Moudgal, C. J.; Lipscomb, J. C.; Bruce, R. M. *Toxicology* **2000**,

147, 109-131.
2. Junhe, L.; Huixian, Z.; Chengyong, Y.; Zirui, Y.; Jinqi, Z. *Wat. Res.* **2002**, *36*, 970-974.
3. Meier, J. R.; Blazak, W. F.; Knohl, R. B. *Environ. Mol. Mut.* **1987**, *10*, 411-424.
4. Halonen, I.; Tarhanen, J.; Ollikainen, S.; Ruokojarvi, P.; Tuppurainen, K.; Ruuskanen, J. *Chemosphere* **1994**, *28*, 2129-2138.
5. Maron, D. M.; Ames, B. N. *Mut. Res.* **1983**, *113*, 173-215.
6. Tikkanen, L.; Kronberg, L. *Mut. Res.* **1990**, *240*, 109-116.
7. Kronberg, L.; Franzen, R. *Environ. Sci. Technol.* **1993**, *27*, 1811-1818.
8. Matsumura, H.; Watanabe, M.; Matsumoto, K.; Ohta, T. *J. Tox. Environ. Health* **1994**, *43*, 65-72.
9. LaLonde, R. T.; Xie, S.; Chamulitrat, W.; Mason, R. P. *Chem. Res. Toxicol.* **1994**, 7, 482-486.
10. Munter, T.; Curieux, F. L.; Sjoholm, R.; Kronberg, L. *Chem. Res. Toxicol.* **1998**, *11*, 226-233.
11. Miettinen, I.; Martikinen, P.; Vartiainen, T.; Lotjonen, S. *Chemosphere* **1993**, *27*, 1707-1718.
12. Munter, T.; Curieux, F. L.; Sjoholm, R.; Kronberg, L. *Chem. Res. Toxicol.* **1999**, *12*, 40-52.
13. Munter, T.; Curieux, F. L.; Sjoholm, R.; Kronberg, L. *Chem. Res. Toxicol.* **1998**, *11*, 226-233.
14. Meier, J. R.; Monarca, S.; Patterson, K. S.; Villarini, M.; Daniel, F. B.; Moretti, M.; Pasquini, R. *Toxicology* **1996**, *110*, 59-70.
15. Marsteinstredet, U.; Wiger, R.; Brunborg, G.; Homgslo, J. K.; Holme, J. A. *Chemico-Biological Interactions* **1997**, *106*, 89-107.
16. Mowry, D. T. *J. Am. Chem. Soc.* **1950**, *72*, 2535-2537.
17. Nawrocki, J.; Andrzejewski, P.; Zelen, H.; Wasowicz, E. *Wat. Res.* **2001**, *35*, 1891-1896.
18. Marsteinstredet, U.; Brunborg, G.; Bjoras, M.; Soderlund, E.; Seeberg, E.; Kronberg, L.; Holme, J. A. *Mut. Res.* **1997**, *390*, 171-178.
19. Tikkanen, L.; Kronberg, L. *Mut. Res.* **1990**, *240*, 109-116.
20. M.T.Hyttinen, J.; Myohanen, S.; Jansson, K. *Carcinogenesis* **1996**, *17*, 1179-1181.
21. Kronberg, L.; Franzen, R. *Environ. Sci. Technol.* **1993**, *27*, 1811-1818.
22. Franzen, R.; Goto, S.; Tanabe, K.; Morita, M. *Mut. Res.* **1998**, *417*, 31-37.
23. Ishiguro, Y.; LaLonde, T.; Dence, C. W. *Environ. Tox. Chem.* **1987**, *6*, 935-946.
24. Ishiguro, Y.; Santodonato, J.; Neal, M. W. *Environ. Mol. Mutagenesis* **1988**, *11*, 225-234.
25. LaLonde, R. T.; Bu, L.; Henwood, A.; Fiumano, J.; Zhang, L. *Chem. Res. Toxicol.* **1997**, *10*, 1427-1436.
26. LaLonde, R. T.; Cook, G. P.; Perakyla, H.; Bu, L. *Chem. Res. Toxicol.* **1991**, *4*, 540-545.
27. LaLonde, R. T.; Xie, S. *Chem. Res. Toxicol.* **1992**, *5*, 618-624.
28. Kronberg, L.; Christman, R. F. *Sci. Total Environ.* **1989**, *81*, 219-230.
29. Kronberg, L.; Christman, R. F.; Singh, R.; Ball, L. M. *Environ. Sci. Technol.* **1991**, *25*, 99-104.
30. LaLonde, R. T.; Xie, S.; Bu, L. *Environ. Mol. Mut.* **1993**, *22*, 181-187.
31. LaLonde, R. T.; Cook, G. P.; Perakyla, H.; Dence, C. W. *Chem. Res. Toxicol.* **1991**, *4*, 35-40.
32. Meier, J. R.; Knohl, R. B.; Coleman, W. E.; Ringhand, H. P.; Munch, J. W.; Kaylor, W. H.; Stereicher, R. P.; Kopfler, F. C. *Mut. Res.* **1987**, *189*, 363-373.
33. LaLonde, R. T.; Lee, H. R. *Chem. Res. Toxicol.* **1994**, *7*, 779-783.
34. LaLonde, R. T.; Cook, G. P.; Perakyla, H.; Dence, C. W.; Babish, J. G. *Environ. Mol. Mut.* **1991**, *17*, 40-48.
35. Cho, S. J. *Bull. Korean Chem. Soc.* **2002**, *23*, 929-930.
36. Cho, S. J. *Bull. Korean Chem. Soc.* **2003**, *24*, 731-732.
37. Tuppurainen, K.; Lotjonen, S.; Laatikainen, R.; Varitiainen, T. *Mut. Res.* **1992**, 181-188.
38. Kronberg, L; Christman, R. F. *Sci. Total. Environ.* **1989**, *81*, 219.
39. Poso, A.; Tuppurainen, K.; Gynther, J. *J. Mol. Struct. (THEOCHEM)* **1994**, *304*, 255-260.
40. Cramer, R. D.; Petterson, D. E.; Bunce, J. D. *J. Am. Chem. Soc.* **1988**, *110*, 5959-5967.
41. Klebe, G.; Abraham, U. *J. Comput.-Aided Mol. Design* **1999**, *13*, 1-10.