MOLECULAR & CELLULAR TOXICOLOGY, Vol. 1, No. 1, 46-51, March 2005

3D QSAR (3 Dimensional Structure Activity Relationship) Study of Mutagen X

Hae-seok Yoon[1] **& Seung Joo Cho**[1]

[1]Life Science Division, Korea Institute of Science and Technology, Seoul 130-650, Korea
Correspondence and requests for materials should be addressed to S.-J. Cho (chosj@kist.re.kr)

Accepted March 10, 2005

Abstract

Mutagen X (MX) exists in our drinking water as the bi-products of chlorine disinfection. Being one of the most potent mutagen, it attracted much attention from many researchers. MX and its analogs are tested and modeled by quantitative structure activity relationship (QSAR) methods. As a result, factors affecting this class of compounds have been found to be steric and electrostatic effects. We tried to collect all the data available from the literature. The quantitative structure-activity relationship of a set of 29 MX was analyzed using Molecular Field Analysis (MFA) and Receptor Surface Analysis (RSA). The best models gave $q^2 = 0.918$, $r^2 = 0.949$ for MFA and $q^2 = 0.893$, $r^2 = 0.954$ for RSA. The models indicate that an electronegative group at C6 position of the furanone ring increases mutagenicity.

Keywords: Mutagen X, QSAR, MFA, RSA, mutagenicity

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 (3-chloro-4-(dichloromethyl)-5-hydroxyl-2 (5H)-furanone) is produced in water is not clearly understood[2,3]. MX is a potent mutagen ever tested in Ames test with test strain TA100[4]. 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[5-8]. 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[9] 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[10]. It is also found to induce apoptosis of HL-60 cells[11]. A relatively large number of MX analogs have been synthesized, tested for mutagenicity[12], subject to many experimental studies. As a result, the resultant MX analogs span a wide range of mutagenicity[13-19].

They are modeled by structure-activity relationship methods[20]. 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. 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[37]. 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[20]. This may imply that MX acts as an electron acceptor. In particular, LUMO electron density and partial charge of the C3 correlated with mutagenicity.

Electron density near C3 also showed negative linear dependency by NMR study. In this study, we

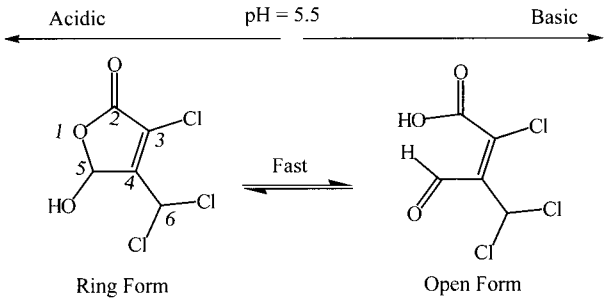

Fig. 1. Two Forms of MX in Equilibrium

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.

Table 1. The Mutagenicity of MX Analogs

	X	Y	Z	ln(TA100)	N
Standard Family					
S1 (MX)	$CHCl_2$	Cl		8.62	9
S2 (BMX2)	$CHBr_2$	Cl		8.61	1
S3 (BMX3)	$CHBr_2$	Br		6.41	2[a]
S4 (CMCF)	CH_2Cl	Cl		6.37	5
S5 (BMBF)	CH_2Br	Br		6.04	1
S6 (MCA)	Cl	Cl		1.87	6[a]
S7 (MBA)	Br	Br		1.71	1
S8	CH_2Cl	H		1.35	3
S9 (MBF)	CH_3	Br		0.41	1
S10 (MCF)	CH_3	Cl		0.21	4
S11	H	Cl		−1.61	1
S12 (MF)	CH_3	H		−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]

Data in this table comprise of 15 reports. N is the number of reports that have mutagenicity data. X, Y and Z are substituents for MX analogs as shown in Figure 2. ln (TA100) is the natural log for experimental values (rev/nm in Ames test). 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 indicates that the compound is not mutagenic and logarithms are taken for remaining value.

The compounds are collected from the available reports and categorized into two groups as shown in Figure 2. Compounds which belongs to standard family (S) contain the structure of 5-hydroxy-2 (5H)-furanone. These compounds are capable of 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 the corresponding open form. Therefore it belongs to ring family (R). The mutagenicity of MX is the average value of 9 different studies[13-19]. All the activity values are within the order of magnitude (3430-13800). Thus the average value can be considered highly reliable. The whole set comprises of 29 compounds. The range of activity is fairly well spread for any particular family as well as for the whole set. All the compounds have unsaturated acidic moiety as a common structure. This structural resemblance might imply that these compounds induce mutagenicity with the same mechanism.

The statistical details of the 3D QSAR models are given in table 2. The predicted activities obtained from MFA and RSA 3D-QSAR models. Scatter-plots of actual versus predicted activities from MFA and RSA are shown in Figure 4.

MFA

The best model was chosen based on the LOF values = 0.877 with smoothness value = 0.1. The regression analysis produced a best QSAR models and is shown in Table 3. Reasonably good values of $r^2 = 0.949$ and $q^2 = 0.918$ for the above QSAR equation explain satisfactorily the variances in the activity.

The steric (CH_3) and electrostatic (H^+) descriptors in the multiple model of MFA specify (Table 4) the regions where variations in the structural features (steric or electrostatic) lead to increased or decreased activities. In the multiple model, we assumed that

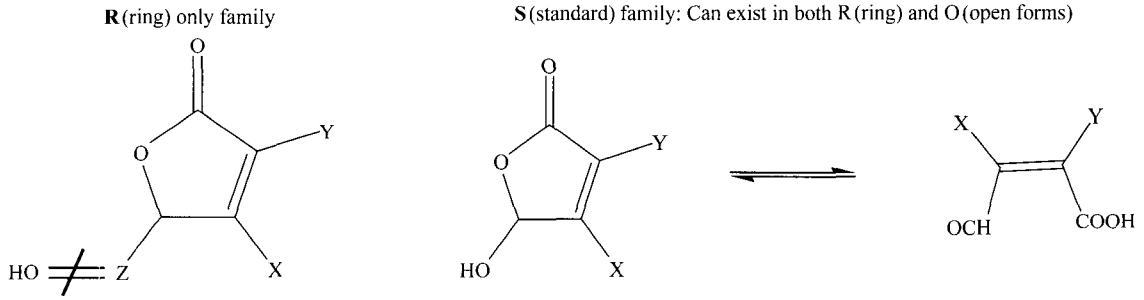

Fig. 2. Two Families of MX Analogs

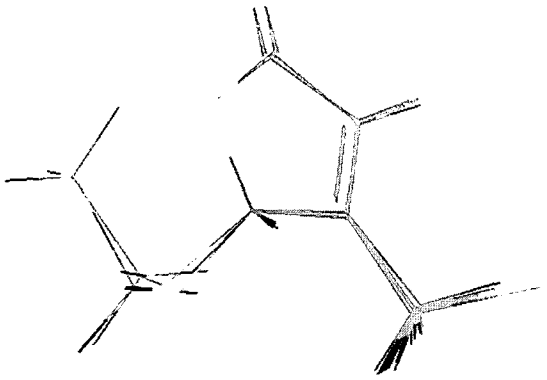

Fig. 3. Superposition of MX Analogs Used in MFA and RSA Studies

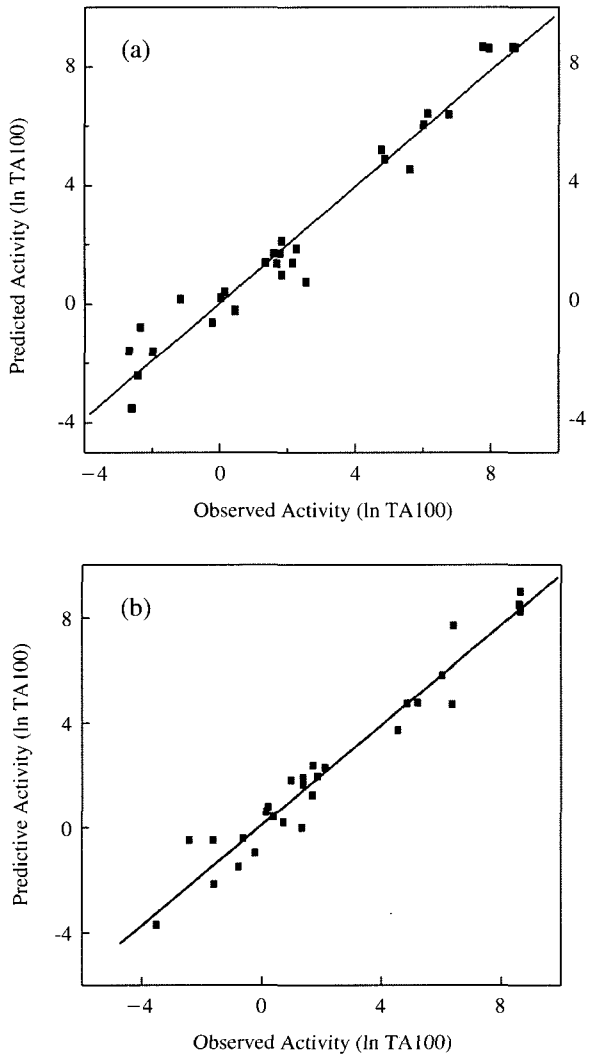

Fig. 4. Scatter-plots of Actual Verses Predicted Activity (a) MFA (b) RSA

common descriptors shown in figure 5, $H^+/137$ and $H^+/112$ most important to prediction. The negative coefficient of $H^+/137$ and $H^+/112$ descriptors indicates that the negative charge (likely fluoride) at C6 position of the furanone ring increases mutagenicity.

RSA

A Receptor Surface Model (RSA) with $r^2 = 0.954$

Table 2. Results of MFA and RSA

Structure	Activity	MFA	MFA residual	RSA	RSA residual
S1	8.620	8.159	0.461	9.201	−0.581
S2	8.610	7.915	0.695	8.448	0.162
S3	6.410	6.317	0.093	6.723	−0.313
S4	6.370	6.255	0.115	6.761	−0.391
S5	6.040	6.268	−0.228	5.020	1.020
S6	1.870	2.655	−0.785	2.503	−0.633
S7	1.710	2.070	−0.360	1.095	0.615
S8	1.350	1.211	0.139	1.066	0.284
S9	0.410	0.069	0.341	−0.865	1.275
S10	0.210	−0.219	0.429	1.045	−0.835
S11	−1.610	−1.323	−0.287	0.168	−1.778
S12	−3.510	−2.645	−0.865	−3.809	0.299
R1	8.650	7.228	1.422	7.554	1.096
R2	8.650	7.005	1.645	8.386	0.264
R3	5.200	5.697	−0.497	4.546	0.654
R4	4.860	5.641	0.781	4.522	0.338
R5	4.540	5.715	−1.175	4.638	−0.098
R6	2.110	2.576	−0.466	1.414	0.696
R7	1.700	2.044	−0.344	2.566	−0.866
R8	1.370	1.845	−0.475	1.320	−0.050
R9	1.370	2.756	−1.386	2.584	−1.214
R10	0.990	1.898	−0.908	0.254	0.736
R11	0.740	0.328	0.412	0.632	0.108
R12	0.170	−0.012	0.182	−0.397	−0.397
R13	−0.220	−1.246	1.026	0.417	−0.637
R14	−0.780	−1.870	1.090	0.405	−1.185
R15	−0.620	0.568	−1.188	−0.348	−0.272
R16	−1.590	−2.624	1.034	−2.656	1.066
R17	−2.410	−3.071	0.661	−1.983	−0.427

The Models of MFA and MSA used here are the best models.

Table 3. Statistical Results for MFA and RSA Models

	MFA	RSA
NC[a]	29	29
q^{2b}	0.918	0.893
r^{2c}	0.949	0.954
BS r^{2d}	0.949	0.955
PRESS[e]	29.260	38.233
Outliers[f]	3	4

[a], The number of compounds. [b], Squared correlation coefficients of cross-validated analysis. [c], Squared correlation coefficient of non-cross-validated analysis. [d], The average squared correlation coefficient calculated during the validation procedure [e], Predicted sum of square [f], A structure with a residual greater than two times the standard deviation

and $q^2 = 0.893$ was developed. The GFA results of the RSA best model are summarized in Table 3.

The QSAR equation generated by RSA is given in equation 1.

$$\text{Activity} = -17.197 - 3.901 \times \text{ELE}/1182 - 0.097 \times \text{ELE}/10802 + 41.146 \times \text{VDW}/8642 + 16.549 \times \text{VDW}/456 - 4.78081 \times (\text{ELE}/1277 + 2.132)^2 \quad (1)$$

The positioning of these descriptors, indicated by a number along with the descriptors, on the model explains the nature of the substituents required. The negative coefficients of (ELE/1182) and (ELE/1277) descriptors explain the possibilities of improving the activity with some electronegative substituents at C6 position (Figure 6).

To study the feasibility of this proposition, we replaced C6 group with fluoride, bromide. And predicted with MFA and RSA models (Table 5). With more electronegative group at C6 group would enhance the mutagenicity.

The quantitative structure-activity relationship of a set of 29 MX was analyzed using molecular-field analysis (MFA) and receptor surface analysis (RSA) provides similar information about the structural requirement. The MFA and RSA models infer similar information about the physical requirements for mutagenicity. In previous study, positive steric contribution emphasizes the importance of substitution at the C6 position.

Our MFA and RSA models for mutagenic activity of MX compounds provides as a result of electrostatic factor more important than steric one at the C6 position. Biggest mutagenic compound, 3-Cloro-4-dibromometyl-5-methoxy-furanone, was modified by substitution by Fluoride at C6. This was predicted more mutagenic. So we came to the conclusion that C6 substitution with more negative charge group can enhance the mutagenicity.

Table 4. Multiple Models by MFA

Model	Equation
Model 1	Activ ty = $-25.9401 - 0.608\ x\ H^+/137 + 0.475\ x\ H^+/119 - 0.358\ x\ H^+/112 - 37.239$ x $CH_3/124$
Model 2	Activity = $-24.4134 - 0.658\ x\ H^+/137 + 0.511\ x\ H^+/119 - 0.200\ x\ H^+/112 - 92.847$ x $CH_3/125$
Model 3	Activity = $-33.788 - 113.624$ x $CH_3/125 - 1.279$ x $H^+/148 + 0.969\ x\ H^+/119 - 0.267\ x\ H^+/112$
Model 4	Activity = $-24.74 - 0.405\ x\ H^+/137 + 0.319$ x $H^+/94 - 97.493$ x $CH_3/125 - 0.174\ x\ H^+/112$
Model 5	Activity = $-29.7718 - 73.996$ x $CH_3/125 - 12.152$ x $CH_3/95 - 0.471\ x\ H^+/137 + 0.366\ x\ H^+/119$

Italics indicate common descriptors

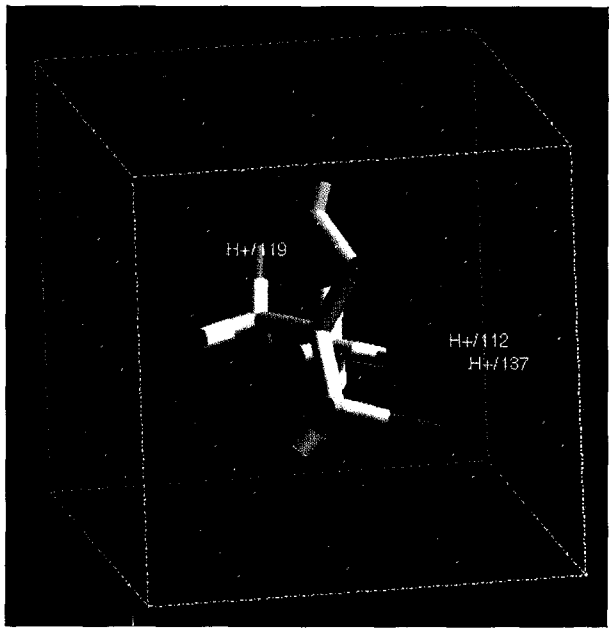

Fig. 5. Common Descriptors, ($H^+/112$) and ($H^+/137$)-Elec-Trostatic Factor, are Important

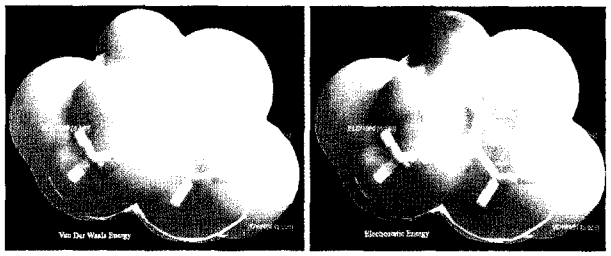

Fig. 6. The Violet and Green Colors in This Figure Indicate the Favorable and Unfavorable Interactions Respectively, Between the Molecules and the Receptor Surface

Table 5. Prediction of Activity According to New Substitution

	X	Y	Z	ln (TA100)	Prediction
R1 (Reference)	$CHBr_2$	Cl	OCH_3	8.62	7.228
R1-1	CHBrF(S)	Cl	OCH_3		10.279
R1-2	CHBrF(R)	Cl	OCH_3		4.486
R1-3	$CHBr_2$	F	OCH_3		6.404

It is show more mutagenicity C6 substitution than C3 one, specially substitute S-form for X.

Methods

Quantitative structure-activity relationships (QSARs) correlate, within congeneric series of compounds, affinities of ligands inhibition constants, rate constants, and other biological activities. 3D QSAR models were developed using molecular field analysis (MFA) and receptor surface analysis (RSA) methodologies using Cerius^2 suite of programs on a series of Mutagen X. The three-dimensional molecular structures of the compounds in the data sets were fully optimized and atomic charges were calculated with AM1 (Austin Model 1) Hamiltonian. The resultant charges were used for electrostatic parameter calculations. All the possible conformations were generated and selected based on the minimum energy. Then the chosen conformers were superimposed as shown in Figure 3 by matching corresponding atoms in the 5-membered ring.

The MFA formalism calculates probe interaction energies on a rectangular grid and a bundle of molecules. Molecular field values were generated for all the aligned molecules using CH_3 and H^+ for steric and electrostatic interactions with default grid values of 2.0Å and partial atomic charges were determined by Gasteiger method. RSA attempts to postulate and to represent the essential features of a receptor site itself, rather than common features of the molecules that bind to it. The aligned molecular aggregate was reconsidered for the generation of receptor surface, which in principle represent a virtual active site of the target. The receptor surface was generated with weights based on the biological activity data. The interaction energies of all the molecules were evaluated within this receptor surface. The receptor surface descriptors, expressed as 3D-field descriptors, derived from the Van Der Waals (VDW) and electrostatic interaction energies (ELE) between molecule-receptor model interaction energies on a receptor surface. By using Genetic Function Algorithm (GFA) regression method, multiple QSAR equations were generated. The GFA was initially conceived by taking inspiration from two seemingly disparate algorithms: genetic algorithm (GA) and multivariate adaptive regression splines (MARS) algorithm. GAs are derived from an analogy with the evolution of DNA[38]. There are especially good at searching problem spaces with a large number of dimensions, as they conduct a very efficient directed sampling of the large space of possibilities. MARS algorithm is a statistical technique for modeling data. It provides an error measure, called the lack of fit (LOF) score, which automatically penalizes models with too many features.

References

1. Moudgal, C.J., Lipscomb, J.C. & Bruce, R.M. Potential health effects of drinking water disinfection by-products using quantitative structure toxicity relationship. *Toxicology.* **147**, 109 (2000).
2. Junhe, L., Huixian, Z., Chengyong, Y., Zirui, Y. & Jinqi, Z. A possible new disinfection by-product--2-chloro-5-oxo-3-hexene diacyl chloride (COHC)--in formation of MX by chlorinating model compounds. *Wat. Res.* **36**, 970 (2002).
3. Meier, J.R., Blazak, W.F. & Knohl, R.B. Mutagenic and clastogenic properties of 3-chloro-4-(dichloromethyl)-5-hydroxy-2 (5H)-furanone: a potent bacterial mutagen in drinking water. *Environ. Mol. Mut.* **10**, 411 (1987).
4. Maron, D.M. & Ames, B.N. Revised methods for the Salmonella mutagenicity test. *Mut. Res.* **113**, 173 (1983).
5. Tikkanen, L. & Kronberg, L. Genotoxic effects of various chlorinated butenoic acids identified in chlorinated drinking water. *Mut. Res.* **240 (2)**, 109 (1990).
6. Matsumura, H., Watanabe, M., Matsumoto, K. & Ohta, T. 3-Chloro-4-(dichloromethyl)-5-hydroxy-2 (5H)-furanone (MX) induces gene mutations and inhibits metabolic cooperation in cultured Chinese hamster cells. *J. Tox. Environ. Health* **43**, 65 (1994).
7. LaLonde, R.T. Xie, S., Chamulitrat, W. & Mason, R. P. Oxidation and radical intermediates associated with the glutathione conjugation of mucochloric acid. *Chem. Res. Toxicol.* **7**, 482 (1994).
8. Munter, T., Curieux, F.L., Sjoholm, R. & Kronberg, L. Reaction of the potent bacterial mutagen 3-chloro-4- (dichloromethyl)-5-hydroxy-2 (5H)-furanone (MX) with 2′-deoxyadenosine and calf thymus DNA: identification of fluorescent propenoformyl derivatives. *Chem. Res. Toxicol.* **11**, 226 (1998).
9. Munter, T., Curieux, F.L., Sjoholm, R. & Kronberg, L. Identification of an ethenoformyl adduct formed in the reaction of the potent bacterial mutagen 3-chloro-4- (dichloromethyl)-5-hydroxy-2 (5H)-furanone with guanosine. *Chem. Res. Toxicol.* **12**, 46 (1999).
10. Meier, J.R., Monarca, S., Patterson, K.S., Villarini, M., Daniel, F.B., Moretti, M. & Pasquini, R. Urine mutagenicity and biochemical effects of the drinking water mutagen, 3-chloro-4-(dichloromethyl)-5-hydroxy-2[5H]-furanone (MX), following repeated oral administration to mice and rats. *Toxicology.* **110**, 59 (1996).
11. Marsteinstredet, U., Wiger, R., Brunborg, G. Homgslo, J. K. & Holme, J. A. Apoptosis in HL-60 cells induced by 3-chloro-4-(dichloromethyl)-5-hydroxy-2[5H]-furanone (MX). *Chemico-Biological Interactions*, **106**, 89 (1997).
12. Marsteinstredet, U., Brunborg, G., Bjoras, M. Soderlund, E., Seeberg, E., Kronberg, L. & Holme, J.A.

DNA damage induced by 3-chloro-4-(dichloromethyl)-5-hydroxy-2[5H]-furanone (MX) in HL-60 cells and purified DNA in vitro. *Mut. Res.* **390**, 171 (1997).
13. Franzen, R. Goto, S., Tanabe, K. & Morita, M. Genotoxic activity of chlorinated butenoic acids in Salmonella typhimurium strains TA98, TA100 and TA104. *Mut. Res.* **417**, 31 (1998).
14. Ishiguro, Y., Santodonato, J. & Neal, M.W. Mutagenic potency of chlorofuranones and related compounds in Salmonella. *Environ. Mol. Mutagenesis*, **11**, 225 (1988).
15. LaLonde. R.T., Cook, G.P., Perakyla, H. & Bu, L. Structure-activity relationships of bacterial mutagens related to 3-chloro-4-(dichloromethyl)-5-hydroxy-2 (5H)-furanone: an emphasis on the effect of stepwise removal of chlorine from the dichloromethyl group. *Chem. Res. Toxicol.* **4(5)**, 540 (1991).
16. Kronberg, L. & Christman, R.F. Chemistry of mutagenic by-products of water chlorination. *Sci. Total Environ.* **81**, 219 (1989).
17. LaLonde, R.T., Xie, S. & Bu, L. Test of chiral recognition in the Salmonella typhimurium (TA100) mutagenicity of mucochloric acid-cysteine adducts. *Environ. Mol. Mut.* **22**, 181 (1993).
18. Meier, J.R., Knohl, R.B., Coleman, W.E., Ringhand, H.P., Munch, J.W., Kaylor, W.H., Stereicher, R.P. & Kopfler, F.C. Studies on the potent bacterial mutagen, 3-chloro-4-(dichloromethyl)-5-hydroxy-2(5H)-furanone: aqueous stability, XAD recovery and analytical determination in drinking water and in chlorinated humic acid solutions. *Mut. Res.* **189**, 363 (1987).
19. LaLonde, R.T., Cook, G.P., Perakyla, H. Dence, C. W. & Babish, J.G. Salmonella typhimurium (TA100) mutagenicity of 3-chloro-4-(dichloromethyl)-5-hydroxy-2 (5H)-furanone and its open- and closed-ring analogs. *Environ. Mol. Mut.* **17**, 40 (1991).
20. Tuppurainen, K. Lotjonen, S., Laatikainen R. & Varitiainen, T. Structural and electronic properties of MX compounds related to TA100 mutagenicity. A semi-empirical molecular orbital QSAR study. *Mut. Res.* **266(2)**, 181 (1992).