Bull. Environ. Contam. Toxicol. (1995) 55:796–801

Environmental Contamination and Toxicology

Discriminant Analysis of the Relationship Between Genotoxicity and Molecular Structure of Organochlorine Compounds

Q. Huang, X. Wang, Y. Liao, L. Kong, S. Han, L. Wang

Department of Environment Science and Engineering, Nanjing University, Nanjing 210093, People's Republic of China

Received: 23 May 1994/Accepted: 1 June 1995

Organochlorine compounds are widely employed and exist as residues in the environment. Some of the compounds are carcinogenic while others are not, even though they exhibit positive responses in short-term mutagenic assays. The main structural features affecting the carcinogenicity of organochlorine compounds are still unknown. Up to now, there is no effective way to predict whether or not organochlorine compounds are carcinogenic.

In this report, the genotoxicity of 41 organochlorine compounds are studied using discriminant analysis. We examined the experimental results of the Salmonella assay (Ames Test), and the carcinogenicity of these compounds to the mouse. Discriminant functions relating to the genotoxicity of organochlorine compounds were established based on molecular descriptors.

MATERIALS AND METHODS

The 41 compounds studied (Table 1), include chlorinated hydrocarbons, chlorinated aromatics and some pesticides, etc., which are extensively used. None of the 41 compounds contain amino or nitro groups. The mutagenicity and carcinogenicity data were taken from the U.S. National Toxicology Program (Ashby et al., 1991). Three kinds of genetoxicity indices were examined, the first involving the Salmonella assay (SA) , the second, mouse liver carcinogenicity (CL), and the third, whole mouse carcinogenicity (CW) (tumors at all tissue sites). The three genotoxicity indices involving the 41 compounds are given in Table 1. The compounds with positive effects were denoted as "+", otherwise as "-". Six of the 41 compounds studied were selected at random to be used for validation of the established discrimiant functions. The remaining 35 compounds were used as a training set, based on which discriminant functions were generated.

Correspondence to: L. Wang

Table 1. The genetoxicity and molecular descriptors of 41 compounds.

Chemicals	^1X	^2X	^3X_p	Env_1	Sub_2	Sub_3	Sub_4	Sam** E	Sam** D	CL** E	CL** D	CW** E	CW** D
*Dichloromethane	1.60	0.91	0.00	1.60	2	0	0	+	+	+	-	+	+
Bromo-dichloromethane	2.44	4.11	0.00	3.57	2	0	0	+	+	+	+	+	+
Chloro-dibromomethane	2.82	4.79	0.00	2.27	1	0	0	-	+	+	+	+	+
Di-methylvinylchloride	1.94	1.40	0.65	0.65	0	1	0	+	+	-	-	+	+
3-Chloromethylpyridine	2.92	1.76	1.19	0.80	1	0	0	+	+	-	-	+	-
1-Chloro-2,2-bromopropane	4.14	3.15	3.32	0.80	1	0	0	+	+	-	-	+	-
1,2-Dichloroethane	2.10	1.13	0.64	1.60	2	0	0	+	+	-	-	+	+
*3-Chloro-2-methylpropene	2.01	1.36	0.68	0.80	1	0	0	+	+	-	-	-	-
1,3-Dichloropropene	2.20	1.08	0.53	1.45	1	1	0	+	+	-	-	+	+
1,2-Dichloropropane	2.44	1.99	0.99	1.45	2	0	0	+	+	+	-	+	+
Chloroethane	1.51	0.80	0.00	0.80	1	0	0	+	+	-	-	+	+
Trichloroethane	2.07	1.62	0.74	1.79	0	3	0	-	-	+	+	+	+
1,1,2,2-Tetarchloroethane	2.95	2.99	1.71	2.62	4	0	0	-	-	+	+	+	+
1,1,1,2-Tetrachloroethane	2.85	3.52	1.36	2.50	4	0	0	-	-	+	+	+	+
*Pentachloroethane	3.29	4.30	2.22	3.01	5	0	0	-	-	+	+	+	+
n-Butylchloride	2.51	1.42	0.75	0.80	1	0	0	+	+	-	-	-	-
2-Chloromethylpyridine	3.26	1.72	1.14	0.80	1	0	0	+	+	-	-	-	-
Chloroalcohol	1.62	0.79	0.25	0.80	1	0	0	+	+	-	-	-	+
2,4-Dichlorophenol	3.09	2.44	1.45	1.13	0	0	2	-	-	-	-	-	+
1,2-Di-chlorobenzene	2.96	2.17	1.61	1.13	0	0	2	-	-	-	-	-	+
*Lindane	5.92	5.68	5.85	3.92	6	0	0	-	-	-	-	-	-

Table 1. (continued)

Chemicals	^1X	^2X	^3X_p	Env_1	Sub_2	Sub_3	Sub_4	Sam**		CL**		CW**	
								E	D	E	D	E	D
Chlorobenzene	2.48	1.62	0.99	0.57	0	0	1	-	+	-	-	-	-
1,1,2-Trichloroethane	2.52	2.13	1.05	2.11	3	0	0	-	+	+	-	+	+
Pentachlorophenol	4.55	3.80	3.06	2.83	0	0	5	-	-	+	+	+	+
Hexachloroethane	3.65	5.55	2.89	3.40	6	0	0	-	-	+	+	+	+
Tetrachloroethene	2.52	2.41	1.28	2.27	0	4	0	-	-	+	+	+	+
1,4-Dichlorobenzene	3.13	2.31	1.31	1.13	0	0	2	-	-	+	-	+	+
2,4,6-Trichlorophenol	3.58	2.97	1.78	1.70	0	0	3	-	-	+	+	+	+
*Tetrachlorodibenzo-p-dioxin	6.39	5.05	3.85	2.27	0	0	4	-	-	+	+	+	-
1,1-Dichloroethene	1.49	1.44	0.00	1.13	0	2	0	-	-	+	+	+	+
Hexachlorodibenzo-p-dioxin	7.52	6.18	5.31	3.40	0	0	6	-	-	+	+	+	+
Chlorendic acid	6.10	8.16	4.86	2.27	2	2	0	-	-	+	+	+	+
Chlordane	8.35	8.72	8.75	4.70	6	2	0	-	-	+	+	+	+
Dichlorodiphenylchloroethylen	6.66	6.04	3.05	2.27	0	2	2	-	-	+	+	+	+
Aldrin	6.34	9.02	7.27	3.57	6	0	0	-	-	+	+	+	+
*Dicofol	7.45	7.37	3.69	2.83	3	0	2	-	-	+	+	+	-
Tetrachlorodiphenylethane	6.99	5.08	3.38	2.44	2	0	2	-	-	-	-	-	-
Anilazine	5.37	3.50	2.37	1.70	2	0	1	-	-	-	-	-	-
Endrin	8.51	9.01	9.18	3.40	4	2	0	-	-	-	-	-	-
Methoxychlor	7.43	5.91	3.55	1.70	3	0	0	-	-	-	-	-	-
Heptachlor	7.04	7.69	9.36	4.05	5	2	0	-	-	+	+	+	+

* Reserved for validation

** "E" is experimental results from reference, "D" is Discriminated results through function 1,2 and 3.

The Molecular Connectivity Index (MCI) was applied here to describe the molecular structure (Kier et al., 1976). MCI values, first-order to third-order, for the 41 compounds(^1X, ^2X, ^3X_p) were calculated. MCI reflects molecular size, degree of branching and bond type, etc..

To describe the number and position of the chlorine atom, we selected four substructures: (1) Cl–, (2) Cl–C<, (3) Cl–C=, (4) Cl–C (aromatic carbon). The substructure descriptors are referred to as Sub_1, Sub_2, Sub_3 and Sub_4 respectively.

On the basis of the four described substructures, we computed the environment descriptors, which describe the immediate surroundings the substructures are imbedded in (Yuta, 1981). Environment descriptors were calculated as first order MCI values on atoms comprising the substructures, and the first nearest-neighbor atoms. If no presence was found for a specific substructure, the descriptor was given the value of zero. The environment descriptors related to the four substructures described above were labeled as Env_1, Env_2, Env_3 and Env_4 respectively.

The canonical discriminant analysis program in "Statgraphics" software (STSC, Inc., 1987) was used for statistic analysis. The program determines the discriminant function based on the difference of covariance for different groups.

RESULTS AND DISCUSSION

For each genotoxicity index, all combinations of the molecular descriptors were tested as discrimination variables. The discrimination functions performing the best classification results are listed as follows.

For the SA:

$$D_1 = -0.163^2X + 0.384Env_1 - 0.572Sub_2 - 0.840Sub_3 - 0.910Sub_4 + 2.064 \quad (1)$$

Midpoint is 0.385

For the CL:

$$D_2 = -1.02^1X + 0.941^2X - 0.305^3X_p + 0.824Env_1 + 0.612Sub_3 - 0.598Sub_4 - 1.149 \quad (2)$$

Midpoint is -0.112

For the CW:

$$D_3 = -1.165^1X + 0.684^2X + 0.390Env_1 + 0.508Sub_3$$

$$+ 0.418Sub_4 + 0.707 \quad (3)$$

Midpoint is -0.354

Seven variables (^1X, ^2X, ^3X_p, Env_1, Sub_2, Sub_3, and Sub_4) are used in the above functions. The seven molecular descriptors for the 41 studied compounds are displayed in Table 1.

Using function 1, we may predict the mutagenic activities of the chemicals to the Salmonella based on molecular descriptors. According to discriminant analysis method, if the calculated discriminant value D_1 is bigger than the midpoint 0.385, the compound is ascribed as active, otherwise as inactive. Following the analogue procedure, the carcinogenicity to mouse liver and that to the whole mouse may be predicted through Function 2 and Function 3 respectively. The predicted activities of the three genotoxicity indices for the 41 compounds studied are also listed in Table 1. In the 35 compounds of training set, 32 ones are correctly classified for the SA, 32 compounds are correctly ascribed for the CL, and 30 compounds for the CW. The correct discriminant percent for SA, CL, and CW are 91.4%, 91.4%, and 85.7% respectively.

The derived discriminant functions were thus used to predict the genotoxicity of the 6 compounds in the validation set, all the 6 compounds are correctly predicted for the SA, 5 compounds are classified correctly for the CL, and 4 compounds for the CW. The predictions are good.

Former studies suggested that the mutagenic responses to Salmonella and the carcinogenic responses to the mouse of organochlorine compounds weren't concordant in most cases (Ashby et al., 1991). Many efforts to predict the carcinogenicity of organochlorine compounds through their mutagenicity to Salmonella resulted in failure. From the experiment data listed in Table 1, we can also see that some active compounds to Salmonella are inactive to the mouse, while many active compounds to the mouse are inactive to Salmonella. Despite the lack of concordance between the mutagenic and carcinogenic responses, all the discriminant functions for SA, CL and CW proposed in this paper exhibit high discriminant ability. Present studies indicate that discriminant analysis based on appropriate molecular descriptors may provide an effective method to predict the mutagenicity and carcinogenicity of the organic pollutants. Among the three genotoxicity indices studied, the lowest discriminant ability was observed for the CW of Function 3, which may be due to different carcinogenic mechanisms at various tissue sites in the mouse.

Acknowledgments. This paper is funded by the National Natural Science Foundation of P.R. China.

REFERENCES

Ashby J, Tennant RW (1991) Definitive relationships among chemical structure, carcinogenicity and mutagenicity for 301 chemicals tested by the U.S.NTP. Mut Res 257:229-306

Kier LB, Hall LH (1976) Molecular connectivity in chemistry and drug research. Academic Press, New York, pp 46-48

Yuta K, Jurs PC (1981) Computer-assisted structure-activity studies of chemical carcinogens, aromatic amines. J Med Chem 24:241-251