Prediction of the Rodent Carcinogenicity of Organic Compounds from Their Chemical Structures Using the FALS Method

Ikuo Moriguchi,[1] Hiroyuki Hirano,[2] and Shuichi Hirono[1]

[1]School of Pharmaceutical Sciences, Kitasato University, Tokyo, Japan; [2]Zeria Pharmaceutical Co., Ltd., Tokyo, Japan

Fuzzy adaptive least-squares (FALS), a pattern recognition method recently developed in our laboratory for correlating structure with activity rating, was used to generate quantitative structure–activity relationship (QSAR) models on the carcinogenicity of organic compounds of several chemical classes. Using the predictive models obtained from the chemical class-based FALS QSAR approach, the rodent carcinogenicity or noncarcinogenicity of a group of organic chemicals currently being tested by the U.S. National Toxicology Program was estimated from their chemical structures. — Environ Health Perspect 104(Suppl 5):1051–1058 (1996)

Key words: QSAR, FALS, rodent carcinogenicity, predictive models

Introduction

The prediction of carcinogenicity has become a subject of great importance for regulatory perspectives and ecotoxicity assessments. Especially, prediction only from the chemical structure is desired, since it can be utilized even when a test compound is unavailable or does not exist. Approaches using some correlative methods for noncongeneric chemicals were reviewed by Richard (*1*), who found that published prediction accuracies were in excess of 90%, while prospective prediction accuracies were less than 70% in these approaches. Moreover, worse results were published for a prospective prediction of rodent carcinogenicity using a variety of quantitative structure–activity relationship (QSAR) approaches (*2*). Further studies are required to improve the predictive reliability.

We have recently developed fuzzy adaptive least-squares (FALS) (*3,4*), a pattern recognition method for correlating structure with activity rating, and applied the method to a noncongeneric structure–carcinogenicity correlation (*5*). Ideally, rational preclassification of compounds based on possible carcinogenic mechanisms should be extensively investigated to enhance the predictive accuracy of noncongeneric QSAR approaches. Unfortunately, for this purpose there is still not sufficient knowledge concerning molecular mechanisms of carcinogenicity. In this study, a rough chemical classification was adopted to generate the predictive models. Using data from the International Agency for Research on Cancer (IARC) (*6*) and the National Toxicology Program (NTP) (*7,8*) on carcinogenicity as training sets, FALS QSAR models for eight chemical classes were generated. Based on these models, prospective predictions of rodent carcinogenicity of 25 organic chemicals issued by the National Institute of Environmental Health Sciences (NIEHS) were accomplished.

This paper is part of the NIEHS Predictive-Toxicology Evaluation Project. Manuscript received 15 February 1996; manuscript accepted 30 May 1996.

Address correspondence to Dr. Ikuo Moriguchi, School of Pharmaceutical Sciences, Kitasato University, Shirokane, Minato-ku, Tokyo 108, Japan. Telephone: 03-3444-6161. Fax: 03-3440-5246.

Abbreviations used: FALS, fuzzy adaptive least squares; QSAR, quantitative structure–activity relationship; IARC, International Agency for Research on Cancer; NTP, National Toxicology Program; NIEHS, National Institute of Environmental Health Sciences.

Methods

FALS Methodology

FALS is a nonparametric pattern classifier. It formulates QSAR in a single discriminant function irrespective of the number of activity rating classes, as:

$$Z = w_0 + w_1x_1 + w_2x_2 + \ldots + w_px_p \quad [1]$$

In this equation, x_k = kth descriptor (k = 1,2,...,p) for structures, w_k (k = 0,1, 2,...,p) = weight coefficient, and Z = discriminant score. A novel feature of FALS is that the degree to which each compound belongs to its activity class is given by a fuzzy membership function (*9*). In FALS, a bell-shaped membership function for each activity class is assumed to give the membership grade for the class members.

In the simplest case, in which the number of activity rating classes is only two, e.g., carcinogenic/noncarcinogenic dichotomization as in this study, the membership function, $M(Z)$, for each activity class is given as:

For carcinogenic activity,

$$M(Z) = 1/[1+\{(Z - Boundary)/0.1 - 1\}^4] \quad \text{when } Z \le Boundary + 0.1,$$
$$\text{otherwise } M(Z) = 1 \quad [2]$$

For noncarcinogenic activity,

$$M(Z) = 1/[1+\{(Boundary - Z)/0.1 - 1\}^4] \quad \text{when } Z \ge Boundary - 0.1,$$
$$\text{otherwise } M(Z) = 1 \quad [3]$$

In these equations, *Boundary* takes the value of $(n_1 - n_2)/(n_1 + n_2)$, where n_1 and n_2 are the numbers of noncarcinogens and carcinogens, respectively, in the training set. The calculated value of $M(Z)$ is the membership grade.

The weight coefficients in the discriminant function are generated so as to maximize the sum of the membership grade over the set of compounds by an adaptive least-squares iteration. The resultant discriminant functions that have various descriptors are validated by the leave-one-out prediction. The discriminant function with a scientifically reasonable set of structural descriptors giving the best leave-one-out prediction is finally adopted as the QSAR model. The FALS methodology has been described on a number of occasions (*3–5*).

Database and Chemical Classes

A database including a total of 586 compounds listed in Table 1 was used for the training sets. The compounds had been designated as carcinogenic or noncarcinogenic by IARC (*6*) and/or NTP (*7,8*) based upon evaluation of rodent test data. If the two agencies' carcinogenicity/noncarcinogenicity assignments differed for any given compound, the NTP designation was adopted. Compounds giving equivocal

Table 1. Training set compounds.

Noncarcinogens	Noncarcinogens	Noncarcinogens
Acetanilide	D&C red 9	Methyl bromide
Acetic acid	DDT	Methyl methacrylate
Acetohexamide	Dexon	Methyl parathion
Acetone	4,4′-Diamino-2,2′-stilbenedisulfonic acid	3-Methyl-4-nitroquinoline-*N*-oxide
4-Acetylaminofluorene	Diarylanilide Yellow	2-Methylindole
Acetylsalicylic acid	Diazinon	2-Naphthol
Acridine	Dibenz[*a*,*h*]anthracene-5,6-oxide	*N*-(1-Naphthyl)ethylenediamine
Aldicarb	Dibenzo-*p*-dioxin	2-Naphthylamine-1,5-disulfonic acid
2-Aminonaphthalene-1-sulfonic acid	Dibenzylnitrosamine	Nicotine
Aminophenol	1,2-Dichlorobenzene	5-Nitro-2-furoic acid
4-Aminoquinoline-1-oxide	1,1-Dichloroethane	4-Nitro-*o*-phenylenediamine
d,*l*-Amphetamine	2,4-Dichlorophenol	4-Nitroanthranilic acid
Anilazine	Dimethoate	Nitrobenzene
Anthracene	2,4-Dimethoxyaniline	1-Nitronaphthalene
Anthranilic acid	Dimethyl sulfoxide	*p*-Nitrophenol
Anthrone	1,1-Dimethyl-4,4′-bipyridinium dichloride	Nitrosostyrene
L-Ascorbic acid	Dimethylamine	Orange G
Azinphosmethyl	*p*-Dimethylaminobenzaldehyde	Orotic acid
Benzanthrone	Dimethylformamide	Penicillin V
Benzimidazole	2,3-Dimethylquinoxaline	Perylene
Benzo[*e*]pyrene	*m*-Dinitrobenzene	Phenanthrene
Benzoic acid	2,4-Dinitrophenol	Phenol
Benzoin	Dinitrosopentamethylenetetramine	1-Phenyl-3-methyl-5-pyrazolone
1*H* Benzotriazole	Dioxathion	Phenylephrine
Benzyl alcohol	Endrin	*o*-Phenylphenol
BHT	Ephedrin	Photodieldrin
Biphenyl	Ethanol	Phthalamide
5-Bromodeoxyuridine	Ethionamide	Phthalic anhydride
n-Butyl chloride	Ethylene glycol	Piperonyl butoxide
Butylurea	1,1′-Ethylene-2,2′-bipyridinium dibromide	Ponceau SX
γ-Butyrolactone	Ethylenediaminetetraacetic acid	Promethazine
Caffeine	Fluorene	Propyl *p*-hydroxybenzoate
Calmagite	5-Fluorodeoxyuridine	Propylene
Calmoisine	Folpet	Pyrazinamide
Camphor	Geranyl acetate	Pyrene
Caprolactam	Gibberellic acid	Pyrimethamine
Carbazole	Glycidyl stearate	Quintozene
Carbromal	HC blue 2	Resorcinol
d-Carvone	Hexacarbate	Riboflavin
Chloramine	Hexachlorocyclopentadiene-1,3	Succinic anhydride
Chloro-*p*-phenylenediamine	Hydrocortisone	Sulfaguanidine
3-Chloro-*p*-toluidine	1-Hydroxy-2-acetylamino fluorene	Sulfisoxazole
Chloroacetic acid	3-Hydroxy-2-acetylamino fluorene	3-Sulfolene
4-Chloroacetylacetanilide	5-Hydroxy-2-acetylamino fluorene	Sunset yellow FCF
o-Chloroaniline	7-Hydroxy-2-acetylamino fluorene	Tetracene
o-Chlorobenzalmelanonitrile	8-Hydroxyquinoline	2,3,5,6-Tetrachloro-4-nitroanisole
2-Chloroethanol	Indole	Tetracycline
Chloroethyl trimethyl ammonium	5-Iododeoxyuridine	Tolazamide
2-(Chloromethyl)pyridine	Iodoform	Tolbutamide
Chlorophenylamine	Lithocholic acid	Toluene
Chloropropham	Malaoxon	L-Tryptophan
2-Chloroquinoline	Malathion	Vinylidene chloride
Chlorpropamide	Maleic hydrazide	*m*-Vinyltoluene
Cis-9,10-epoxystearic acid	D-Mannitol	*p*-Vinyltoluene
Clonitralid	*d*,*l*-Menthol	*m*-Xylene
Colchicine	Methanol	*o*-Xylene
Coumaphos	Methionine	*p*-Xylene
Cyclohexylamine	Methoxychlor	

Carcinogens	Carcinogens	Carcinogens
Acetamide	Acronycine	Aflatoxin G_1
N-Acetoxy-2-acetylaminofluorene	Actinomycin D	Aflatoxin M_1
1′-Acetoxysafrole	AF-2	Aldrin
4-Acetylaminobiphenyl	Aflatoxin B_1	Allyl chloride
2-Acetylaminofluorene	Aflatoxin B_2	Allyl glycidyl ether

(Continued)

Table 1. *Continued.*

Carcinogens	Carcinogens	Carcinogens
Allyl isovalerate	Chlorendic acid	1,3-Dichloropropene
3-Amino-1*H*-1,2,4-triazole	Chlormadinone acetate	Dichlorvos
1-Amino-2-methylanthraquinone	Chlornaphazen	Dicofol
4-Amino-2-nitrophenol	3-Chloro-2-methylpropene	1,2,3,4-Diepoxybutane
3-Amino-4-ethoxyacetanilide	4-Chloro-*m*-phenylenediamine	1,2,7,8-Diepoxyoctane
2-Amino-4-nitrophenol	4-Chloro-*o*-phenylenediamine	Diethyl sulfate
2-Amino-5-(5-nitro-2-furyl)-1,3,4-thiadiazole	5-Chloro-*o*-toluidine	Diethylstilbestrol
2-Amino-5-nitrophenol	*p*-Chloroaniline	Diethylstilbestrol dipropionate
2-Amino-5-nitrothiazole	Chlorobenzilate	Diglycidyl resorcinol ether
3-Amino-9-ethylcarbazole	Chlorodibromomethane	7,8-Dihydrobenzo[*a*]pyrene
2-Aminoanthraquinone	Chloroethane	3,4-Dihydrocoumarin
11-Aminoundecanoic acid	Chloroform	Dihydrosafrole
o-Anisidine	3-(Chloromethyl)pyridine	3,3'-Dimethoxybenzidine
Aramite	10-Chloromethyl-9-chloroanthracene	3,3'-Dimethoxybenzidine-4,4'-diisocyanate
5-Azacytidine	10-Chloromethyl-9-methylanthracene	Dimethyl morpholonophosphoramidate
Azaserine	7-Chloromethylbenz[*a*]anthracene	Dimethyl sulfate
Aziridine	1(4-Chlorophenyl)-3,3-dimethyltriazine	1,2-Dimethyl-5-nitroimidazole
2-(1-Aziridinyl)ethanol	Chlorothalonil	((Dimethylamino)-methyleneimino)-5-(2-(5-nitro-2-furan)vinyl)oxadiazole
Aziridyl benzoquinone	Chrysene	
Azoxymethane	Cinnamyl anthranirate	9,10-Dimethylanthracene
Benz[*a*]anthracene	Citrus red 2	7,12-Dimethylbenz[*a*]anthracene
Benz[*a*]anthracene-5,6-oxide	Coumarin	7,9-Dimethylbenz[*c*]acridine
Benz[*c*]acridine	*m*-Cresidine	Dimethylcarbamyl chloride
Benzaldehyde	*p*-Cresidine	Dimethylvinyl chloride
Benzene	Cupferron	2,4-Dinitrofluorobenzene
Benzo[*a*]pyrene	Cycasin	1,4-Dinitrosopiperazine
Benzo[*a*]pyrene-4,5-oxide	Cyclamaic acid	2,4-Dinitrotoluene
Benzo[*b*]fluoranthene	Cyclochlorotine	1,4-Dioxane
Benzo[*j*]fluoranthene	Cyclophosphamide	1,1-Diphenyl-2-butynyl *N*-cyclohexylcarbamate
Benzofuran	Cytembena	Diphenylhydantoin
N-Benzoyloxy-4-methylaminoazobenzene	Dacarbazine	Direct black 38
Benzyl acetate	Daminozide	Direct blue 6
Benzyl chloride	Dapsone	Direct brown 95
Benzyl violet 4B	Daunorubicin	Epichlorohydrin
o-Benzyl-*p*-chlorophenol	*p,p'*-DDE	3',4'-Epoxy-6'-methyl-cyclohexylmethyl 3,4-epoxy-6-methyl-cyclohexylcarboxylate
Bis(1-aziridinyl)-morpholinophosphine	Decabromodiphenyl oxide	
Bis(2-chloro-1-methylethyl)ether	β-Deoxy-6-thioguanosine	1,2-Epoxybutane
Bis(2-chloroethyl)ether	Di(2-ethylhexyl) adipate	1-Epoxyethyl-3,4-epoxycyclohexane
Bis(2-hydroxymethyl)dithiocarbamate	Di(2-ethylhexyl) phthalate	Estradiol 3-benzoate
1,2-Bis(chloromethoxy)ethane	*N,N'*-Diacetylbenzidine	Estradiol dipropionate
1,4-Bis(chloromethoxymethyl)benzene	Diallate	Estradiol mustard
Bis(chloromethyl)ether	2,4-Diaminoanisole	17β-Estradiol
2,2-Bis(*p*-hydroxyphenyl)propane diglycidyl ether	2,4-Diaminophenol	Estriol
Blue VRS	Diazoacetylglycine hydrazide	Estrone
Brilliant blue FCF diammonium	Diazoacetylglycineamide	Estrone benzoate
Bromochloromethane	Dibenz[*a,c*]anthracene	Ethinylestradiol
Bromoethane	Dibenz[*a,h*]acridine	Ethionine
7-Bromomethyl-12-methylbenz[*a*]anthracene	Dibenz[*a,h*]anthracene	Ethyl acrylate
1,3-Butadiene	Dibenz[*a,j*]acridine	Ethyl bromoacetate
1,4-Butane sultone	Dibenzo[*a,e*]pyrene	Ethyl methanesulfonate
Butyl benzyl phthalate	Dibenzo[*a,h*]pyrene	Ethyl *p*-toluenesulfonate
β-Butyrolactone	Dibenzo[*a,i*]pyrene	*N*-Ethyl-*N'*-nitro-*N*-nitrosoguanidine
C.I. Acid Orange 3	Dibenzo[*a,l*]pyrene	Ethylene dibromide
C.I. Acid Red 114	7*H*-Dibenzo[*c,g*]carbazole	Ethylene oxide
C.I. Direct Blue 15	Dibenzo[*h,rst*]pentaphene	Ethylene sulphide
C.I. Direct Blue 218	2,3-Dibromo-1-propanol	Ethylene thiourea
C.I. Disperse Blue 1	1,2-Dibromo-3-chloropropane	Ethynodiol diacetate
C.I. Disperse Yellow 3	1,1-Dibromo-3-chloropropane	Evans blue
C.I. Pigment Red 3	Dibromomannitol	Fanft
C.I. VAT Yellow 4	3,3'-Dichloro-4,4'-diaminodiphenyl ether	Fast green FCF
Cantharidin	2,6-Dichloro-*p*-phenylenediamine	Formaldehyde
Captan	1,4-Dichlorobenzene	2-(Formylhydrazino)-4-(5-nitro-2-furyl)-thiazole
Carbon tetrachloride	3,3'-Dichlorobenzidine	Furaltadone
Chloramben	1,2-Dichloroethane	Furan
Chlorambucil	Dichloromethane	Furfural
Chlordane	9,10-Dichloromethylanthracene	Furosemide
Chlordecone (kepone)	1,2-Dichloropropane	Glycidaldehyde

(Continued)

Table 1. *Continued.*

Carcinogens	Carcinogens	Carcinogens
Glycidol	Nithiazide	Phenytoin
Griseofulvin	Nitrogen mustard *N*-oxide	Piperonyl sulfoxide
Guinea green B	Nitrilotriacetic acid	Pivalolactone
HC Blue 1	*N*-(4-(5-Nitro-2-furyl)-2-thiazolyl)acetamide	Polybrominated biphenyl
Heptachlor	5-Nitro-*o*-anisidine	Ponceau 3R
Hexachlorbenzene	5-Nitro-*o*-toluidine	Ponceau MX
Hexachlorobutadiene	3-Nitro-*p*-acetophenetide	Progesterone
Hexachlorodibenzo-*p*-dioxins	2-Nitro-*p*-phenylenediamine	Pronamide
Hexachloroethane	5-Nitroacenaphthene	Pronetalol
Hydorxymethyl-12-methylbenz[*a*]anthracene	*o*-Nitroanisole	1,3-Propane sultone
Hydroquinone	5-Nitrobenzimidazole	β-Propiolactone
N-Hydroxy-2-acetylaminofluorene	6-Nitrobenzimidazole	*N*-Propyl carbamate
N-(2-Hydroxyethyl)hydrazine	2-Nitrobiphenyl	*N*-Propyl-*N*′-nitro-*N*-nitrosoguanidine
4-Hydroxylaminoquinoline-1-oxide	4-Nitrobiphenyl	Propylene oxide
6-Hydroxymethylbenzo[*a*]pyrene	Nitrofen	Propyleneimine
N-Hydroxyphenacetin	2-Nitrofluorene	Propylthiouracil
1′-Hydroxysafrole	Nitrofurantoin	Quinoline
N-Hydroxyurethane	Nitrofurazone	Reserpine
ICR-10 (Quinacrine mustard)	((Nitrofurfurylidene)amino)-2-imidazolidinone	Retrorsine
ICR-170	Nitrogen mustard	Rhodamine 6G
ICRF-159	2-Nitronaphthalene	Rhodamine B
Indeno(1,2,3-CD)pyrene	2-Nitropropane	Saccharin
Iodinated glycerol	8-Nitroquinoline	Safrole
Isatidine	4-Nitroquinoline-1-oxide	Semicarbazide
N-Isobutyl-*N*′-nitro-*N*-nitrosoguanidine	*N*-Nitroso-*N*-methylurethane	Shikimic acid
Isonicotinic acid hydrazide	*N*-Nitrosobutylurea	Sterigmatocystin
Isophorone	*N*-Nitrosodibutylamine	Streptozotocin
Isophosphamide	*N*-Nitrosodiethanolamine	Styrene
Isosafrole	*N*-Nitrosodiethylamine	Sudan I
Lasiocarpine	*N*-Nitrosodimethylamine	Sudan II
Light green SF	4-Nitrosodimethylaniline	Sulfallate
d-Limonene	*N*-Nitrosodipentylamine	Sulfamethoxazole
Luteoskyrin	*p*-Nitrosodiphenylamine	Testosterone
Mannomustine	*N*-Nitrosodipropylamine	Testosterone propionate
Medroxyprogesterone acetate	*N*-Nitrosoethylurea	2,3,7,8-Tetrachlorodibenzo-*p*-dioxin
Megestrol acetate	Nitrosomethyl ethylamine	1,1,1,2-Tetrachloroethane
Melamine	*N*-Nitrosomethyl vinylamine	1,1,2,2-Tetrachloroethane
Melphalan	*N*-Nitrosomethylurea	Tetrachloroethylene
2-Mercaptobenzothiazole	*N*-Nitrosomorphine	Tetrachlorvinphos
Mestranol	2-Nitrosonaphthylene	Tetranitromethane
Methapyrilene	*N*′-Nitrosonornicotine	Thiotepa
3-Methoxy-4-aminoazobenzene	*N*-Nitrosopiperidine	Thiouracil
Methyl carbamate	*N*-Nitrosopyrrolidine	*o*-Toluenesulfonamide
Methyl iodide	*N*-Nitrososarcosine	Toxaphene
Methyl methanesulfonate	Norethisterone	Trenimon (tris) aziridinyl-*p*-benzoquinone
2-Methyl-1-nitroanthraquinone	Norethisterone acetate	Trinitrofluoren-9-one
N-Methyl-*N*′-nitro-*N*-nitrosoguanidine	Norethynodrel	Triamterene
N-Methyl-*N*-nitrosoaniline	Ochratoxin A	Tribromomethane
Methylazoxymethanol	Oil orange SS	2,4,6-Trichloroaniline
Methylazoxymethanol acetate	Orange I	1,1,2-Trichloroethane
7-Methylbenz[*a*]anthracene	Oxazepam	Trichloroethylene
α-Methylbenzyl alcohol	4,4′-Oxydianiline	2,4,6-Trichlorophenol
3-Methylcholanthrene	Parascorbic acid	1,2,3-Trichloropropane
4,4′-Methylene bis(2-chloroaniline)	Patulin	Triethylene glycol diglycidyl ether
N-Methylolacrylamide	Penicillic acid	Trifluraline
Methylthiouracil	Pentachloroanisole	Tris (1-aziridinyl-*s*-triazine)
Metronidazole	Pentachloroethane	Tris(2,3-dibromopropyl)phosphate
Michler's ketone	Pentachlorophenol	Tris(2-chloro-1-methylethyl)ether
Mirex	*N*-Pentyl-*N*′-nitro-*N*-nitrosoguanidine	1,2,3-Tris(chloromethoxy)propane
Mitomycin C	Peroxyacetic acid	Trypan blue
Monocrotaline	Phenacetin	Uracil mustard
Monuron	Phenazopyridine	Urethane
Mustard	Phenesterin	Vinyl chloride
Nalidixic acid	Phenicarbazide	4-Vinylcyclohexene
Naphthalene	Phenobarbital	Zearalenone
Natulan (procarbazine)	Phenoxybenzamine	
Niridazole	Phenylbutazone	

evidence of carcinogenicity were not used. Inorganic and metallo-organic chemicals, polymers, and mixtures were also excluded from the training sets.

The chemical classification was designed to be broad enough to permit a reasonable number of training compounds to fall into each class for generation of statistically significant QSAR models. With a special reference to the chemical features of the compounds to be predicted, the following eight chemical classes were investigated: class 1, hydrocarbons (39 compounds); class 2, heterocyclics (185 compounds); class 3, nitro and nitroso compounds and *N*-oxides (98 compounds); class 4, halides (152 compounds); class 5, alcohols, phenols, and ethers (160 compounds); class 6, carbonyl compounds (205 compounds); class 7, nonaromatic amines (25 compounds); and class 8, oxygenated sulfur compounds (52 compounds). An individual compound can appear in several classes according to its chemical structure. 2,3,5,6-Tetrachloro-4-nitroanisole, for example, appears in classes 3, 4, and 5.

Structural Descriptors

Three kinds of variables—continuous variables, discrete variables, and indicator variables—were investigated as candidate descriptors. Molecular weight, hydrophobic constant (log P), and its squared value were used as continuous variables. The log P (octanol/water) values used were calculated using the revised version (*10*) of our simple method (*11,12*). Discrete variables were defined as the number of specific atoms, bonds, functional groups, and specific ring and chain structures. The upper values of the discrete variables other than the number of specific atoms and bonds were empirically set at 3.0 so as to avoid possible overestimation for polyfunctional structures. Indicator variables were defined as 1 for the presence and 0 for the absence of any kind of structural or physicochemical features considered to be contributing to carcinogenicity.

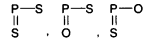
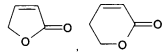
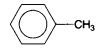
Results and Discussion

Generation of Predictive Models

The FALS analyses were performed for carcinogenic/noncarcinogenic dichotomization using eight sets of data for the various chemical classes. As a result, the eight satisfactory equations including from 5 to 25 descriptors (Moriguchi et al., unpublished data) were derived. They are listed in Table 2.

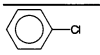
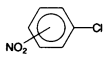
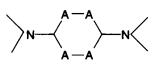
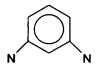
Descriptors with positive coefficients are usually considered to contribute in a

Table 2. FALS QSAR models for predicting carcinogenicity.

Descriptor[a]	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7	Class 8
Log P_{calc}			−0.703	0.080	0.050		0.130	−0.255
$(\text{Log } P_{calc})^2$		−0.005	0.112					
$(mw/100)^{1/2}$		−1.438			−0.147			2.152
C, sp^3		0.099			0.124		0.056	
C, sp^2				−0.057	−0.136		0.470	−0.781
C, sp						0.206		
H(–C)			0.062		−0.040			
H(–N)		−0.203			−0.138			
H(–O)			−0.589					−0.106
H(–S)		0.929						
S			−0.147		−0.283	−0.208		
P		0.881						
Cl		−0.032					0.491	
I				−0.048				
Number of unsat. bonds		0.125		−0.055	0.061			−0.048
C=C in chains					1.267	0.252		
QN[b]					2.017	1.514		
POL[b]		0.223	0.155	0.140				
				−0.359	−0.167	0.129		
				0.674				
>N–N< outside rings		1.038			1.290	0.995		
>C=O (keto)				0.222	−0.083	−0.237		
>C=O (quinones)		0.478			0.258	0.474		
(Sat. C)–OH		0.085		0.169	−0.137	−0.149		
(Unsat. C)–OH		−0.431			−0.178	−0.508		
(N)–OH			1.623					
(S)–OH								−0.365
–COOH, $(–CO)_2O$		−0.642		0.764		−0.310		
>N→O			−0.656					
$–NO_2$			0.225					
>N–CS–N<		1.102						
		−0.661						
$–CH_2–Cl$, $–CH_2–Br$		0.470		0.401				
>C=C–C–Cl, >C=C–Cl				0.332				
$–CHCl_2$, >CCl–CCl< and/or –CCl=CCl– outside rings		0.774						
$Q–CH_2–CH_2–Cl$					0.328			
$Q(CH_2–CH_2–Cl)_2$					−0.196	0.477		
$–CX_n$, n≥2				0.045				
$CO–NH_2$, $CS–NH_2$						−0.049		
$(\text{Aliph.C})_2NH$		0.500				0.036		
$(\text{Aliph.C})_3N$				0.607			−0.794	
$Ar–NH_2$		0.306		−0.010	0.668	0.200		
Ar–NH–				−0.474	0.497			
$Ar–N(CH_3)_2$						0.113		
Ar–N=N–Ar					0.272	0.752		
$Ar–NO_2$		0.578		0.313				
Ar–N=O			0.905					
Ar–NH–CO–					−0.408			
Ar–OH						0.398		
Ar–Cl		−0.341						
						−0.281		
		0.518						
	−0.705					−0.215		

(*Continued*)

Table 2. *Continued.*

Descriptor[a]	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7	Class 8
						0.171		
			−0.401	−0.611				
N–CO–N, N–CS–N in rings						−0.258		
N–CO–N outside rings						−0.180		
M–1	1.232							
M–2	1.436							
M–3	1.411							
CH_2=C–C=C	0.789							
CH_2=C–C–C–C=C	1.436							
RNG[b]				−0.092		0.433		
				1.046				
						−0.375		
								0.595
Constant	−0.667	0.746	−0.412	−0.751	−0.227	−0.308	−0.468	−2.600
Boundary	−0.231	−0.416	−0.633	−0.500	−0.388	−0.376	0.040	−0.269

[a]Q, heteroatoms; A, C and/or heteroatoms; X, halogen atoms; Ar, aromatic rings; M-1, 9-Me-anthracene moiety; M-2, fluoranthene moiety; M-3, benzo (or dihydrobenzo)[*a* or *b*]phenanthrene moiety. M-1 and thereafter are indicator variables. [b]Source: Moriguchi (*10*).

Table 3. Reliability of the QSAR models.

			Recognition			Leave-one-out		
Class	Set of data[a]	No. of descriptors	False negative	False positive	MMG[b]	False negative	False positive	MMG[b]
1	24/15	6	5.1%	0.0%	0.949	7.7%	0.0%	0.923
2	131/54	23	1.6	8.1	0.890	5.4	9.7	0.845
3	80/18	11	2.0	3.1	0.922	3.1	6.1	0.899
4	114/38	20	2.0	5.9	0.899	7.9	11.2	0.822
5	111/49	22	2.5	6.9	0.878	7.5	8.8	0.827
6	141/64	25	2.0	10.7	0.860	4.4	12.7	0.802
7	12/13	5	4.0	0.0	0.935	8.0	12.0	0.783
8	33/19	7	5.8	3.8	0.877	5.8	7.7	0.831

[a]Number of training compounds: carcinogens/noncarcinogens. [b]Mean membership grade.

positive way to the estimate of carcinogenicity, whereas descriptors with negative coefficients contribute in a negative way. However, this is not always valid beyond the chemical classes. Moreover, strictly speaking, these coefficients cannot be used to make general inferences about the contribution of each fragment within a variety of structures. They are valid only when used in the context of the present multidimensional model within each chemical class.

The results of recognition and leave-one-out prediction of the eight QSAR models are shown in Table 3. The values of the mean membership grade were fairly good, from 0.860 to 0.949 in the recognition and from 0.783 to 0.923 in the leave-one-out prediction. The false negative was from 1.6 to 5.8% in the recognition and from 3.1 to 8.0% in the leave-one-out prediction. These equations were then used for the carcinogenicity prediction of 25 organic chemicals.

Prospective Prediction of the Organic Chemicals

The second NIEHS Predictive-Toxicology Evaluation Project involves the rodent carcinogenicity of 30 chemicals consisting of 25 organic and 5 inorganic compounds. The five inorganic compounds were omitted from our FALS prediction because sufficient carcinogenicity data for inorganic chemicals were not available for generating predictive QSAR models. The prediction of the 25 organic compounds was performed using the QSAR models for the eight chemical classes listed in Table 2. Salts such as scopolamine hydrobromide trihydrate and sodium xylenesulfonate were treated as undissociated forms. The results are shown in Table 4.

From the chemical features, compounds 1 (scopolamine) and 2 (codeine) fall into three chemical classes, and compounds 5 (tetrahydrofuran), 10 (D&C Yellow No. 11), 13 (1-chloro-2-propanol), 14 (diethanolamine), 15 (phenolphthalein), 18 (furfuryl alcohol), 19 (primaclone), 24 (oxymetholone), and 26 (emodin) fall into two chemical classes. When there were discrepancies between the estimates by two or three QSAR models, we evaluated them as "equivocal." Among the 25 organic chemicals, 14 showed positive, 5 showed equivocal, and 6 showed negative carcinogenicity. Further detailed predictions by the correlative method are thought to be unreliable, since there are not sufficient data concerning mechanisms and sites of tumor formation with a wide variety of chemicals for the generation of statistically significant QSAR models.

In these predictions, the mutagenicity and subchronic toxicity test data were not considered. The prediction based on the QSAR models can be performed in a very short time at a very low cost, and it can be utilized even when the test compound does not exist. Unfortunately, the first round of this exercise showed that the results by the correlative methods were not very good (*2*). It is considered that the predictive power of correlative methods significantly depends upon the quality and quantity of the training set data used. Sufficient high-quality data covering a large variety of chemical structures, as well as the use of mechanism-based descriptors, will enhance the prospective prediction accuracies of the QSAR approaches.

Table 4. Prediction results of 25 organic chemicals.

Chemical				Membership grade		Overall
No.	Name	Class	Z	Carcinogen	Noncarcinogen	prediction[a]
1	Scopolamine	2	−0.190	1.000	0.009	
		5	−0.157	1.000	0.008	+
		6	−0.176	1.000	0.012	
2	Codeine	2	0.225	1.000	0.000	
		5	−0.034	1.000	0.002	E
		7	−0.505	0.001	1.000	
3	1,2-Dihydro-2,2,4-trimethylquinoline	2	−0.268	1.000	0.026	+
4	Nitromethane	3	0.179	1.000	0.000	+
5	Tetrahydrofuran	2	−0.079	1.000	0.003	+
		5	−0.153	1.000	0.008	
6	*t*-Butylhydroquinone	5	−0.461	0.099	0.995	−
7	Ethylbenzene	1	−0.667	0.001	1.000	−
8	Chloroprene	4	−0.597	0.063	1.000	−
10	D&C Yellow No. 11	2	0.126	1.000	0.001	E
		6	−0.653	0.005	1.000	
11	Isobutyraldehyde	6	−0.179	1.000	0.013	+
13	1-Chloro-2-propanol	4	−0.118	1.000	0.002	+
		5	−0.004	1.000	0.002	
14	Diethanolamine	5	−0.660	0.005	1.000	−
		7	−0.374	0.001	1.000	
15	Phenolphthalein	5	−0.420	0.248	0.824	E
		6	−0.094	1.000	0.005	
16	Pyridine	2	−0.158	1.000	0.006	+
17	Xylenesulfonic acid	8	−0.763	0.001	1.000	−
18	Furfuryl alcohol	2	−0.244	1.000	0.018	E
		5	−0.480	0.068	1.000	
19	Primaclone	2	0.225	1.000	0.000	+
		6	0.197	1.000	0.000	
20	Ethylene glycol monobutyl ether	5	−0.262	1.000	0.037	+
22	Isobutene	1	−0.667	0.001	1.000	−
23	Methyleugenol	5	0.752	1.000	0.000	+
24	Oxymetholone	5	−0.188	1.000	0.012	E
		6	−0.382	0.440	0.563	
25	Anthraquinone	6	0.641	1.000	0.000	+
26	Emodin	5	−0.105	1.000	0.005	+
		6	0.098	1.000	0.001	
27	Citral	6	0.196	1.000	0.000	+
29	Cinnamaldehyde	6	−0.056	1.000	0.003	+

[a] +, Carcinogenic; −, noncarcinogenic; E, equivocal.

REFERENCES

1. Richard AM. Application of SAR methods to non-congeneric data bases associated with carcinogenicity and mutagenicity: issues and approaches. Mutat Res 305:73–97 (1994).
2. Hileman B. "Expert intuition" tops in test of carcinogenicity prediction. Chem Eng News 71(25):35–37 (1993).
3. Moriguchi I, Hirono S, Matsushita Y, Liu Q, Nakagome I. Fuzzy adaptive least squares applied to structure-activity and structure-toxicity correlations. Chem Pharm Bull 40:930–934 (1992).
4. Moriguchi I, Hirono S, Liu Q, Nakagome I. Fuzzy adaptive least squares and its application to structure-activity studies. Quant Struct-Act Relat 11:325–331 (1992).
5. Moriguchi I, Liu Q, Hirano H, Hirono S. Noncongeneric structure-toxicity correlation using fuzzy adaptive least-squares. In: Classical and Three-Dimensional QSAR in Agrochemistry, ACS Symposium Series 606 (Hansch C, Fujita T, eds). Washington:American Chemical Society Books, 1995;141–152.
6. Soderman J, ed. CRC Handbook of Identified Carcinogens and Noncarcinogens, Vol. 1. Boca Raton, FL:CRC Press, 1982.
7. Ashby J, Tennant RW. Definitive relationships among chemical structure, carcinogenicity and mutagenicity for 301 chemicals tested by the U.S. NTP. Mutat Res 257:229–306 (1991).
8. Gold LS, Manley NB, Slone TH, Garfinkel GB, Ames BN,

Rohrbach L, Stern BR, Chow K. Sixth plot of the carcinogenic potency database: results of animal bioassays published in the general literature 1989 to 1990 and by the National Toxicology Program 1990 to 1993. Environ Health Perspect 103(Suppl 8):3–122 (1995).

9. Novak V. Fuzzy Sets and Their Applications, Bristol:Adam Hilger, 1989;222–234.
10. Moriguchi I. Development of fuzzy adaptive least-squares and its uses in quantitative structure-activity relationships. Yakugaku Zasshi 115:805–822 (1995).
11. Moriguchi I, Hirono S, Liu Q, Nakagome I, Matsushita Y. Simple method of calculating octanol/water partition coefficient. Chem Pharm Bull 40:127–130 (1992).
12. Moriguchi I. Hirono S, Nakagome I, Hirano H. Comparison of reliability of log P values for drugs calculated by several methods. Chem Pharm Bull 42:976–978 (1994).