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

Ikuo Moriguchi,1 Hiroyuki Hirano,2 and Shuichi Hirono1

1School of Pharmaceutical Sciences, Kitasato University, Tokyo, Japan; 2Zeria 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 correlating 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.

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:

\[ Z = w_0 + w_1 x_1 + w_2 x_2 + \ldots + w_p x_p \]  

In this equation, \( x_k \) is \( k \)th descriptor (\( k=1,2,\ldots,p \)) for structures, \( w_k \) (\( k=0,1,2,\ldots,p \)) = 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,

\[ M(Z) = \frac{1}{1 + \left( \frac{1}{|\text{Boundary}|} - 1 \right)^4} \]

when \( Z \leq \text{Boundary} + 0.1 \),

otherwise \( M(Z) = 1 \)  

For noncarcinogenic activity,

\[ M(Z) = \frac{1}{1 + \left( \frac{1}{|\text{Boundary} - Z|} - 1 \right)^4} \]

when \( Z \geq \text{Boundary} - 0.1 \),

otherwise \( M(Z) = 1 \)  

In these equations, \( \text{Boundary} \) 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 |
| Acetoxime       | Dexon          | Methyl parathion |
| Acetone         | 4,4'-Diamino-2,2'-stilbenedisulfonic acid | 3-Methyl-4-nitroquinoline-N-oxide |
| 4-Acetylaminofluorene | Diarylanilide Yellow | 2-Methylindole |
| Acrystalsalicylic acid | Diazon | 2-Naphthol |
| Acrylicamide    | Dibenylanthracene-5,6-oxide | N-(1-Naphthyl)ethanamide |
| Acridine        | Dibenzylnitrosamine | 2-Naphthylamine-1,5-disulfonic acid |
| Aldicarb        | 1,2-Dichlorobenzene | Nicotine |
| 2-Aminonaphthalene-1-sulfonic acid | Dibenzylaramine | 5-Nitro-2-furoic acid |
| Aminophenol     | 1,1-Dimethyl-4,4'-bipyridinium dichloride | 4-Nitro-o-phenylenediamine |
| 4-Aminoquinoline-1-oxide | Ethylene glycol | 4-Nitroanilinic acid |
| α-Amphetamine   | Ethylamine     | Nitrobenzene |
| Anthracene      | Ethylene glycol | 1-Nitronaphthalene |
| Anthranic acid  | Ethylene glycol | p-Nitrophenol |
| Anthrone        | Ethylene glycol | Nitrosoxyrene |
| l-Ascorbic acid | Ethylene glycol | Orange G |
| Azinphosmethyl  | Ethylene glycol | Groic acid |
| Benzantrone     | Ethylene glycol | Penicillin V |
| Benzoimidazole  | Ethylene glycol | Perylene |
| Benzofluorene   | Ethylene glycol | Phenacetin |
| Benzoic acid    | Ethylene glycol | Phenol |
| Benzoic acid    | Ethylene glycol | 1-Phenyl-3-methyl-5-pyrazolone |
| Butyl chloride  | Ethylene glycol | Phenylenepipine |
| γ-Butyrolactone | Ethylene glycol | o-Phenylphenol |
| Caffeine        | Ethylene glycol | Photodiethylin |
| Calmodine       | Ethylene glycol | Phthalaldehyde |
| Calmodine       | Ethylene glycol | Phthalic anhydride |
| Camphor         | Ethylene glycol | Piperonyl butoxide |
| Caprolactam     | Ethylene glycol | Ponceau SJ |
| Carbazole       | Ethylene glycol | Promethazine |
| Carbosil        | Ethylene glycol | Propyl p-hydroxybenzoate |
| Carbosil        | Ethylene glycol | Propylene |
| d-Carvone       | Ethylene glycol | Pyrazinamide |
| Chlormazine     | Ethylene glycol | Pyrene |
| Calcoisine      | Ethylene glycol | Pyrimethamine |
| Camphor         | Ethylene glycol | Quintozine |
| Caprolactam     | Ethylene glycol | Resorcinol |
| Carbazole       | Ethylene glycol | Riboflavin |
| Chlorobenzaldeene| Ethylene glycol | Succinic anhydride |
| Chloroanilide   | Ethylene glycol | Sulafaguanadin |
| Chloroacetic acid| Ethylene glycol | Sulfoxazide |
| Chloroacetylacetanilide | Ethylene glycol | 3-Sulfolene |
| Chloroanilide   | Ethylene glycol | Sunset yellow FCF |
| Chloroacetamide | Ethylene glycol | Tetracene |
| Chloroacetoxybenzenesulphonamide | Ethylene glycol | 2,3,5,6-Tetrachloro-4-nitroanisole |
| Chloroacetamide | Ethylene glycol | Tetracycline |
| Chloroacetoxybenzenesulphonamide | Ethylene glycol | Tolozamide |
| Chloroacetamide | Ethylene glycol | Tolbutamide |
| Chloroacetoxybenzenesulphonamide | Ethylene glycol | Toluen |
| Chloroacetoxybenzenesulphonamide | Ethylene glycol | Tryptophan |
| Chloroacetoxybenzenesulphonamide | Ethylene glycol | Vinyl chloride |
| Chloroacetoxybenzenesulphonamide | Ethylene glycol | M-Vinyltoluene |
| Chloroacetoxybenzenesulphonamide | Ethylene glycol | p-Vinyltoluene |
| Chloroacetoxybenzenesulphonamide | Ethylene glycol | m-Xylene |
| Chloroacetoxybenzenesulphonamide | Ethylene glycol | p-Xylene |
| Chloroacetoxybenzenesulphonamide | Ethylene glycol | (Continued)
Table 1. Continued.

| Carcinogens                                      | Carcinogens                                      | Carcinogens                                      |
|-------------------------------------------------|-------------------------------------------------|-------------------------------------------------|
| Allyl isovalerate                               | Chloric acid                                    | 1,3-Dichloropropene                             |
| 3-Amino-1H-1,2,4-triazole                       | Chloromadinone acetate                          | Dichlores                                       |
| 1-Amino-2-methylanthraquinone                   | Chlorophenazan                                  | Dicofol                                         |
| 4-Amino-2-nitrophenol                           | 3-Chloro-2-methylpropane                        | 1,2,3,4-Diepoxybutane                           |
| 3-Amino-4-ethoxyacetanilide                     | 4-Chloro-m-phenylenediamine                     | 1,2,7,8-Diepoxyoctane                           |
| 2-Amino-4-nitrophenol                           | 4-Chloro-p-phenylenediamine                     | Diethyl sulfate                                 |
| 2-Amino-5-(5-nitro-2-furyl)-1,3,4-thiadiazole   | 5-Chloro-o-toluidine                            | Diethylylthiobistrol                            |
| 2-Amino-5-nitrophenol                           | p-Chloroaniline                                 | Diethylthiobistrol dilupionate                  |
| 2-Amino-5-nitrothiazole                         | Chlorobenzilate                                 | Diglycycl resorcinol ether                      |
| 3-Amino-9-ethylcarbazole                        | Chlorodibromomethane                            | 7,8-Dihydrobenz[a]pyrene                        |
| 2-Aminooanthraquinone                           | Chloroether                                     | 3,4-Dihydrocumarin                             |
| 11-Aminoundecanoic acid                         | Chloroform                                      | Dihydrosrofol                                 |
| o-Anisidine                                     | 3-(Chloromethyl)pyridine                        | 3,3’-Dimethylbenzidine                         |
| Aramite                                         | 10-Chloromethyl-9-chloroanthracene              | 3,3’-Dimethoxybenzidine-4,4’-diisocyanate       |
| 5-Azacytidine                                   | 10-Chloromethyl-9-methylenthacene               | Dimethyl morpholonomorphosphoridate            |
| Azaserine                                       | 7-Chloromethylbenz[a]anthracene                 | Dimethyl sulfate                                |
| Azidine                                         | 10/Chlorophenyl-3,3-dimethylenthacene           | 1,2-Dimethyl-5-nitroimidazole                   |
| Aziridine                                       | Chlorothalonal                                  | 1,2-Dimethyl-5-nitroimidazole                   |
| Aziridyl benzoxquinone                          | Chrysene                                        | 1,2-Dimethyl-5-nitroimidazole                   |
| Azoxymethane                                    | Citrus red 2                                   | 9,10-Dimethylanthracene                        |
| Benz[a]anthracene                               | Comarin                                         | 7,12-Dimethylbenz[a]anthracene                 |
| Benz[a]anthracene-5,6-oxide                     | m-Cresidine                                     | 7,9-Dimethylbenz[j]c]icline                     |
| Benz[a]cridine                                  | p-Cresidine                                     | Dimethylcarbamyl chloride                      |
| Benzaldehyde                                    | Cupferron                                       | Dimethylvinyl chloride                         |
| Benzene                                         | Cynasin                                         | 2,4-Dinitrofluorobenzene                       |
| Benz[a]pyrene                                   | Cyclcamic acid                                  | 1,4-Dinitrosipiperazine                        |
| Benz[a]pyrene-4,5-oxide                         | Cyclchochlorine                                 | 2,4-Dinitrotoluene                             |
| Benz[b]fluoranthene                             | Cyclophosphamide                                | 1,4-Dioxane                                    |
| Benz[b]fluoranthene                             | Cytrhembena                                     | 1,1-Diphenyl-2-butynyl N-cyclohexy carbamate   |
| Benzofuran                                      | Dacarabene                                      | Diphenylhydantoin                              |
| N-Benzoxoxygen-4-methylenoazobenzene            | Daminozide                                      | Direct black 38                                 |
| Benzyl acetate                                  | Dapsone                                         | Direct blue 6                                   |
| Benzyl chloride                                 | Daunorubicin                                    | Direct brown 99                                 |
| Benzyl violet 4B                                 | p.p’-DDE                                        | Epichlorhydrin                                 |
| o-Benzyl-p-chlorophenol                         | Decabromodiphenyl oxide                         | 3,4’-Epoxo-6-methyl-cyclohexymethyl 3,4-epoxy-6-|
| Bis(1-aziridinyl)-morpholophosphorine           | β-Deoxy-6-thioguanosine                         | methyl-cyclohexy carboxylate                    |
| Bis[2-chloro-1-methyl]ether                     | Di(2-ethylhexyl) adipate                        | 1,2-Epoxybutane                                |
| Bis[2-chloroethyl]ether                         | Di(2-ethylhexyl) pthalate                      | 1-Epoxyethyl-3,4-epoxy-cyclohexane             |
| Bis[2-hydroxymethyl]dithiocarbamate             | Bis[2,6-Dichloroethyl] ether                    | Estradiol 3-benzoate                           |
| 1,2-Bis(chloromethylethylene)                  | bis[2,4-Diminoanisole]                         | Estradiol dipropionate                         |
| 1,4-Bis[4-hexoxygen]-methylene/benzene         | bis[2,4-Diaminoanisole]                        | Estradiol mustard                              |
| Bischloromethyleneether                         | bis[2,4-Diminoanisole]                        | 17β-Estradiol                                  |
| 2,2-Bis[chloroxygen]-propanoic diglycidyl ether| bis[2,4-Diminoanisole]                        | Estradiol                                     |
| Blue VRS                                        | bis[2,4-Diminoanisole]                        | Estrone                                        |
| Brilliant blue FCF diammonium                  | bis[2,4-Diminoanisole]                        | Estrone benzate                                |
| Bromochloromethane                              | bis[2,4-Diminoanisole]                        | Ethylhexiastriol                               |
| Bromoethane                                     | bis[2,4-Diminoanisole]                        | Ethionine                                      |
| 7-Bromomethyl-12-methylbenz[a]anthracene       | bis[2,4-Diminoanisole]                        | Ethyl acetate                                  |
| 1,3-Butadiene                                   | bis[2,4-Diminoanisole]                        | Ethyl bromoacetate                             |
| 1,4-Butane sultone                             | bis[2,4-Diminoanisole]                        | Ethyl methanesulfonate                         |
| Butyl benzyl phthalate                          | bis[2,4-Diminoanisole]                        | Ethyl p-toluenesulfonate                       |
| β-Butyroactone                                  | bis[2,4-Diminoanisole]                        | N-ethyl-N’-nitro-N-nitroso-quinidine           |
| C.I. Acid Orange 3                              | bis[2,4-Diminoanisole]                        | Ethanlene dibromide                            |
| C.I. Acid Red 114                              | bis[2,4-Diminoanisole]                        | Ethanol oxide                                  |
| C.I. Direct Blue 15                             | bis[2,4-Diminoanisole]                        | Ethylene sulphonide                            |
| C.I. Direct Blue 218                           | bis[2,4-Diminoanisole]                        | Ethylene thiourea                              |
| C.I. Disperse Blue 1                            | bis[2,4-Diminoanisole]                        | Ethynolodial diacete                           |
| C.I. Disperse Yellow 3                          | bis[2,4-Diminoanisole]                        | Evans blue                                     |
| C.I. Pigment Red 3                              | bis[2,4-Diminoanisole]                        | Fanft                                          |
| C.I. VAT Yellow 4                               | bis[2,4-Diminoanisole]                        | Fast green FCF                                 |
| Canthinardan                                    | bis[2,4-Diminoanisole]                        | Formaldehyde                                  |
| Captan                                          | bis[2,4-Diminoanisole]                        | 2-(Formylhydrazino)-4-(5-nitro-2-furyl)-thiazole|
| Carbon tetrachloride                            | bis[2,4-Diminoanisole]                        | Furaladone                                    |
| Chloramben                                      | bis[2,4-Diminoanisole]                        | Furan                                         |
| Chlorambucil                                    | bis[2,4-Diminoanisole]                        | Furural                                        |
| Chloride                                        | bis[2,4-Diminoanisole]                        | Furosemide                                    |
| Chlordecone (kepone)                           | bis[2,4-Diminoanisole]                        | Glycidaldehyde                                |

(Continued)
| Carcinogens                                      | Carcinogens                                      | Carcinogens                                      |
|-------------------------------------------------|-------------------------------------------------|-------------------------------------------------|
| Glycidol                                        | Nithiazide                                      | Phenothiazine                                    |
| Griseofulvin                                    | Nitrogen mustard N-oxide                       | Piperonyl sulfoxide                             |
| Guinea green B                                  | Nitrilotriacetic acid                          | Pivaliactone                                    |
| HC Blue 1                                       | N-(5-Nitro-2-furyl)-2-thiazolyl)acetamide       | Polybrominated biphenyl                         |
| Heptachlor                                       | 5-Nitro-o-aminidine                            | Ponceau 3R                                      |
| Hexachlorobenzene                               | 5-Nitro-o-tolidine                             | Ponceau MX                                      |
| Hexachlorobutadiene                             | 3-Nitro-p-acetoephenthenetet                     | Progestrone                                      |
| Hexachlorodibenzo-p-dioxins                     | 2-Nitro-p-phenylenediamine                     | Pronamid                                        |
| Hexachloroethane                                | 5-Nitroacenaphthene                            | Pronetol                                        |
| Hydroxymethyl-12-methylbenz[a]anthracene        | o-Nitroanisole                                  | 1,3-Propane sultone                             |
| Hydroquinone                                    | 5-Nitrobenzimidazole                           | p-Propioloactone                                |
| N-Hydroxy-2-acetylaminofluorene                 | 6-Nitrobenzimidazole                           | N-Propyl carbamate                               |
| N-Hydroxyethylhydrazine                        | 2-Nitrobiphenyl                                | Propylene oxide                                 |
| 4-Hydroxymethylquinoline-1-oxide                | 4-Nitrophenyl                                  | Propyleneimine                                  |
| 6-Hydroxymethylbenz[a]pyrene                    | Nitrofen                                       | Reserpine                                       |
| N-Hydroxyphenacetin                             | 2-Nitrofluorenone                              | Retorsine                                       |
| 1'-Hydroxyasafrole                              | Nitrofurantoin                                 | Rhodamine 6G                                    |
| N-Hydroxyurethane                              | Nitrofurazone                                  | Rhodamine B                                     |
| ICR-10 (quinacrine mustard)                    | (N-nitrofuranylidenyl)amine-2-imidazolidinone  | Saccharin                                       |
| ICR-170                                        | Nitrogen mustard                               | Safrole                                          |
| ICRF-159                                       | 2-Nitronapthalene                              | Semicarbazide                                   |
| Indeno[1,2,3-CD]pyrene                          | 2-Nitropropane                                 | Shikimic acid                                   |
| Indomethacin                                    | 8-Nitroquinoline                               | Sterigmatocystin                                |
| L-aspartic acid                                 | N-Nitrosodimethylamine                         | Streptorotocin                                  |
| L-aspartic acid                                 | N-Nitrosodimethylamine                         | Styrene                                         |
| L-aspartic acid                                 | N-Nitrosodimethylamine                         | Sudan 1                                         |
| Light green SF                                  | N-Nitroso-N-methylurethane                     | Sudan 1                                         |
| Methylene alcohol                               | N-Nitrosobutyrate                              | Sulfafate                                       |
| Methylamine                                     | N-Nitrosobutyrate                              | Sulfamethoxazole                                |
| Methylamine                                     | N-Nitrosobutyrate                              | Testosterone                                    |
| 2-Mercaptobenzothiazole                         | N-Nitrosodihydroxyamine                       | Testosterone propionate                          |
| Mestranol                                       | N-Nitrosodihydroxyamine                       | 2,3,7,8-Tetrachlorodibenzo-p-dioxin             |
| Methylyprylene                                  | N-Nitrosodihydroxyamine                       | 1,1,1,2-Tetrachloroethane                       |
| Meexprogestrone acetate                         | N-Nitrosodihydroxyamine                       | 1,1,2,2-Tetrachloroethane                       |
| Megestrol acetate                               | Nitrosoethylamyl                               | Tetrachloroethylene                              |
| Metabolite                                      | Nitrosoethylamyl                               | Tetrachloroethoxy                                |
| Melphalan                                       | N-Nitrososarcosine                             | Tetrachlorohydroxy                              |
| Melphalan                                       | N-Nitrososarcosine                             | Tetrachlorohydroxy                              |
| 2-Mercaptobenzothiazole                         | N-Nitrosomorphine                              | Tetrachlorohydroxy                              |
| Mestranol                                       | 2-Nitrosonaphthyl                              | Tetranitromethane                                |
| Methacrylurean                                  | N'-Nitrosomononcicilene                        | Thiopeno                                        |
| 3-Methoxy-4-aminooazobenzene                   | N'-Nitrosopiperidine                           | Thiouarsial                                     |
| Methanol                                       | N'-Nitrosopropionamide                         | o-Toluenesulfonamide                            |
| Methylamid                                       | N'-Nitrosopropionamide                         | Toxaphene                                       |
| Methylzid                                       | N'-Nitrosopropionamide                         | Trenimon (tris aziridyl-p-benzoziquine)          |
| 2-Methyl-1-nitroanthraquinone                   | Norethisterone                                 | Trinitrofluoren-9-one                            |
| N-Methyl-N'-nitro-N-nitrosoguanidine            | Norethisterone                                 | Triamterene                                     |
| N-Methyl-N'-nitro-N-nitrosoguanidine            | Norethisterone                                 | Tribromomethane                                 |
| N-Methyl-N'-nitro-N-nitrosoguanidine            | Norethisterone                                 | 2,4,6-Trichloroaniline                          |
| N-Methyl-N'-nitro-N-nitrosoguanidine            | Norethisterone                                 | 1,1,2-Trichloroethane                           |
| N-Methyl-N'-nitro-N-nitrosoguanidine            | Norethisterone                                 | Trichloroethylene                                |
| N-Methyl-N'-nitro-N-nitrosoguanidine            | Norethisterone                                 | 2,4,6-Trichlorophenol                           |
| N-Methyl-N'-nitro-N-nitrosoguanidine            | Norethisterone                                 | 1,2,3-Trichloropropan                           |
| N-Methyl-N'-nitro-N-nitrosoguanidine            | Norethisterone                                 | Triethylene glycol diglicidyl ether             |
| N-Methyl-N'-nitro-N-nitrosoguanidine            | Norethisterone                                 | Trifluraline                                    |
| Methyliumricol                                  | Pentachloroanisole                             | Tris (1-aziridyl-s triazine)                     |
| Metabolite                                      | Pentachloroanisole                             | Tris(2,3-dibromopropyl)phosphate                |
| Michler's ketone                                | Pentachloroanisole                             | Tris(2-chloro-1-methyl)ethylether               |
| Mirex                                          | N-Pentyl-N'-nitro-N-nitrosoguanidine           | 1,2,3-Trichloromethoxypropyl                    |
| Mitomycin C                                    | Peroxycetic acid                               | Trypan blue                                     |
| Monocrotaline                                   | Phenacetin                                     | Uraclid mustard                                 |
| Monorax                                        | Phenazonipidyl                                 | Urethane                                        |
| Mustard                                        | Phenesterin                                   | Vinyl chloride                                  |
| Nalidixic acid                                  | Phenicarbazide                                 | 4-Vinylcyclohexene                              |
| Naphthalene                                     | Phenbarbital                                  | Zearalenone                                     |
| Natulan (procarbazine)                         | Phenoxynbenzamine                              |                                                |
| Nitrilazole                                     | 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.

| Descriptor* | Class 1 | Class 2 | Class 3 | Class 4 | Class 5 | Class 6 | Class 7 | Class 8 |
|-------------|---------|---------|---------|---------|---------|---------|---------|---------|
| Log Pcalc   | 0.703   | 0.800   | 0.050   | 0.130   | -0.255  |         |         |         |
| (Log Pcalc)^2 | 0.112   |         |         |         |         |         |         |         |
| (mw/100)^1/2 | -0.005  |         |         |         |         |         |         |         |
| C, sp^3     | -1.438  | -0.147  | 0.056   | 0.470   | -0.781  |         |         |         |
| C, sp^2     | 0.099   | 0.124   |         |         |         |         |         |         |
| H(-C)       | 0.057   | -0.136  | 0.206   |         |         |         |         |         |
| H(-C)       | -0.040  |         |         |         |         |         |         |         |
| H(-Cl)      | 0.062   |         |         |         |         |         |         |         |
| H(-Cl)      | -0.203  |         |         |         |         |         |         |         |
| H(-Br)      | -0.598  |         |         |         |         |         |         |         |
| H(-S)       | 0.929   |         |         |         |         |         |         |         |
| S           | -0.147  | -0.283  | -0.208  |         |         |         |         |         |
| C           | 0.881   |         |         |         |         |         |         |         |
| Cl          | -0.032  |         |         |         |         |         |         |         |

Number of unsat. bonds 0.125, C in chains 1.267, cinnamaldehyde 2.017, C=C 1.514, \( \text{P}_{\text{cal}} \)

Table 2. FALS QSAR models for predicting carcinogenicity.

(Continued)
Table 2. Continued.

| Descriptor | Class 1 | Class 2 | Class 3 | Class 4 | Class 5 | Class 6 | Class 7 | Class 8 |
|------------|---------|---------|---------|---------|---------|---------|---------|---------|
| N–CO–N, N–CS–N in rings | 1.232 | 1.436 | -0.092 | -0.256 | -0.180 |
| M–1 | 1.232 | 1.436 | 1.141 | 0.769 | 1.436 | -0.092 | 0.433 | 1.046 |
| M–2 | -0.092 | 0.433 | -0.375 | 0.959 |

*Q, heteroatoms; A, C and/or heteroatoms; X, halogen atoms; Ar, aromatic rings; M-1, 9-Me-anthracene moiety; M-2, fluoroanthene moiety; M-3, benzo (a dihydrobenzo) & b phenanthrene moiety. M-1 and thereafter are indicator variables. *Source: MORIGUCHI ET AL.

Table 3. Reliability of the QSAR models.

| Class | Set of data | No. of descriptors | False negative | False positive | MMG | False negative | False positive | MMG |
|-------|-------------|--------------------|----------------|----------------|-----|----------------|----------------|-----|
| 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%          | 0.922 |
| 5     | 111/49      | 22                 | 2.5%           | 6.9%           | 0.878 | 7.5%           | 8.6%           | 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 |

*Number of training compounds: carcinogens/noncarcinogens. *Mean membership grade.

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 (oxymethalone), 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 No. | Name                        | Class | Z     | Membership grade | Overall prediction¹ |
|--------------|-----------------------------|-------|-------|------------------|---------------------|
| 1            | Scopolamine                 | 2     | -0.190| 1.000            | 0.009               |
| 2            | Codeine                     | 2     | 0.225 | 1.000            | 0.000               |
| 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               |
| 6            | 1-Butylhydroquinone         | 5     | -0.461| 0.999            | 0.995               |
| 7            | Ethylbenzene                | 5     | -0.505| 0.001            | 1.000               |
| 8            | Chloroprene                 | 2     | -0.013| 1.000            | 0.000               |
| 9            | D&C Yellow No. 11           | 2     | 0.126 | 1.000            | 0.001               |
| 10           | Isobutryraldehyde           | 6     | -0.179| 1.000            | 0.013               |
| 11           | 1-Chloro-2-propanol         | 4     | -0.118| 1.000            | 0.002               |
| 12           | Diethanolamine              | 5     | -0.660| 0.005            | 1.000               |
| 13           | Phenolphthalein             | 6     | -0.374| 0.001            | 1.000               |
| 14           | Pyridine                    | 2     | -0.158| 1.000            | 0.006               |
| 15           | Xylenesulfonic acid         | 8     | -0.763| 0.001            | 1.000               |
| 16           | Furfuryl alcohol            | 2     | -0.244| 1.000            | 0.018               |
| 17           | Primacine                   | 5     | -0.480| 0.068            | 1.000               |
| 18           | Ethylene glycol monobutyl ether | 2 | 0.225 | 1.000            | 0.000               |
| 19           | Isobutene                   | 1     | -0.667| 0.001            | 1.000               |
| 20           | Methyleugenol               | 5     | 0.752 | 1.000            | 0.000               |
| 21           | Oxyymetholone               | 5     | -0.188| 1.000            | 0.012               |
| 22           | Anthraquinone               | 6     | 0.641 | 1.000            | 0.000               |
| 23           | Emolin                      | 5     | -0.105| 1.000            | 0.005               |
| 24           | Citral                      | 6     | 0.196 | 1.000            | 0.000               |
| 25           | Cinnamaldehyde              | 6     | -0.056| 1.000            | 0.003               |

*+, 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):33–37 (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, Stone 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).