Prediction of the Acute Toxicity (96-h LC_{50}) of Organic Compounds to the Fathead Minnow (*Pimephales promelas*) Using a Group Contribution Method

Todd M. Martin† and Douglas M. Young*

U.S. EPA, National Risk Management Research Laboratory, Cincinnati, Ohio 45268

Received March 22, 2001

A group contribution method has been developed to correlate the acute toxicity (96-h LC_{50}) to the fathead minnow (*Pimephales promelas*) for 397 organic chemicals. Multilinear regression and computational neural networks (CNNs) were used for model building. The models were able to achieve a fairly good correlation of the data ($r^2 > 0.9$). The linear model, which included four specific interaction terms, provided a rapid means of predicting the toxicity of a compound. The CNN model was able to yield virtually the same predictions with or without the four interaction terms that were included in the multilinear model.

Introduction

Every year millions of chemicals are added to the Chemical Abstracts Service registry. Since experimental measurements of aquatic toxicity are extremely time-consuming and expensive, it is imperative to develop robust quantitative structure−activity relationships (QSARs) that can predict toxicity. The majority of QSARs correlate toxicity using the octanol−water partition coefficient (K_{ow}) and/or various physicochemical, topological, and quantum-chemical parameters (*1*−*6*). These correlations work well for compounds in the same chemical class but in general require different coefficients for each class. In addition, many compounds defy classification.

Hall and co-workers (*7*) used an additivity model to correlate the acute toxicity (96-h LC_{50}) of 105 substituted benzenes. Their additivity model is a slight variation on the Free-Wilson method (*8*). Their model can be written as follows:

$$-\log(LC_{50}) = \sum_i n_i \Delta T_i + T_0 \qquad (1)$$

where LC_{50} is the aqueous concentration that causes 50% mortality in the fathead minnow (units of mol/L are used throughout this paper); n_i is the number of groups of type i which have replaced hydrogen on benzene; ΔT_i is the change in toxicity caused by substituting a hydrogen with group type i; and T_0 is the toxicity of the parent compound, benzene.

Gao and co-workers (*9*) developed a more general group contribution approach to correlate the toxicity (96-h LC_{50}) of 130 mostly substituted benzene compounds. In this approach, chemical fragments, called groups, are used to describe a molecule. Each group has a unique contribution to the toxicity (*9*). This is in contrast to the Free-Wilson method where corrections are added when groups are substituted on the parent compound. The method of Gao and co-workers assumes that each atom in every structure must be accounted for. Using their group contribution approach, the toxicity of a compound is given by (*9*)

$$-\log(LC_{50}) = \sum_{i=1}^{ng} n_i \alpha_i \qquad (2)$$

where ng is the number of groups in the model; n_i is the number of groups of type i in the compound; and α_i is the toxicity contribution of group i.

An advantage of this method is that the independent variables (n_i) can be determined by visual inspection of a compound's structural formula. The toxicity can then simply be determined from eq 2.

Kaiser and Niculescu (*10*) developed a group contribution model to correlate the toxicity of 865 organic chemicals. Their model included groups to account for many different types of functional groups. Their model distinguished between groups attached to aliphatic and aromatic carbons. They used a probabilistic neural network to relate their groups to toxicity.

In the current study, the method of Gao et al. (*9*) was applied to a more diverse dataset of 379 chemicals. The basic groups in this study were similar to those used in the UNIFAC model for evaluating activity coefficients (*11*). Groups that were attached to aliphatic carbons were differentiated from groups attached to aromatic carbons (denoted by AC) to account for the electronic effects of the aromatic ring. Several special groups were included in this study because they were identified as being excessively toxic or operate by a specific mode of action (*12*, *13*). Multilinear regression and a computational neural network (CNN) were used for model building.

Experimental Section

Experimental Data. A large majority of the 96-h LC_{50} values for the fathead minnow were obtained using the median values given in the AQUIRE database (*14*). Additional chemicals were added from the literature (*3*, *7*, *15*−*19*). Some chemicals in the literature were not added to our data set due to a

* To whom correspondence should be addressed.
† ORISE Research Fellow at U.S. EPA.

10.1021/tx0155045 CCC: $20.00
Published on Web 09/11/2001

structural feature that was not present in more than one chemical (which would require the addition of a statistically insignificant parameter). The data set used in this study is given in Table 1.

Description of Groups. The groups utilized in the group contribution model are given in Table 2. The alkane fragments of the molecules were described using the basic UNIFAC groups (CH_3, CH_2, CH, and C). Describing alkenes using UNIFAC parameters would require five different parameters (CH_2═CH, CH_2═C, CH═CH, CH═C, and C═C). For this analysis, it was determined that the number of parameters could be reduced to two by using C═C and H. The H group is the total number of hydrogens attached to the two unsaturated carbons. It was found that these parameters allowed one to fit data for seven diene compounds and for three chlorinated ethene compounds.

Groups were also added to describe alkynes (C≡CH and C≡C). The chemicals that contained alkyne groups in our dataset were propargyl alcohols. Since propargyl alcohols are classified as being excessively toxic (*13*) these parameters should only be used for propargyl alcohols. It has been shown that if the hydroxyl group is attached to a tertiary (3°) carbon, the compound is not excessively toxic (*22*). Therefore, an additional C≡CH parameter was added to describe propargyl alcohols in which the hydroxyl group is attached to a tertiary carbon.

The following UNIFAC groups were also added: hydroxyl (OH), aldehyde (CHO), carboxyl (COOH), and bromo (Br). The diol (2 OH) parameter was used in place of the single alcohol parameter for compounds containing two hydroxy groups. This was done to account for the nonadditive reduced toxicity for compounds containing two hydroxy groups. The methacrylate [CH_2═C(CH_3)COO] group was also added to account for the nonadditive reduced toxicity of this group.

Several other groups (ether O, ketone, ester, nitrile, amines, and chloro) added were fragments of UNIFAC groups. The corresponding UNIFAC groups would have required three times as many parameters (i.e., CH_3X, CH_2X, and CHX as opposed to just X). For example for 2-octanone the parameters are as follows: $n_{CH_3} = 2$, $n_{CH_2} = 5$, $n_{C(=O)} = 1$, and all other parameters are zero. Special groups were added for oxime (C═NOH) and acrylate (CH_2═CHCOO) because they were characterized as being excessively reactive in the literature (*13*).

The LC_{50} values for aldehydes, diesters, and carboxylic acids were found to be relatively insensitive to the number of CH_2 groups present in these compounds. This can be partially explained by the fact that aldehydes and diesters are classified as being excessively reactive (*13*). Russom and co-workers (*13*) were unable to classify carboxylic acids by mode of action. To account for these inconsistencies a second-order interaction term was added for each of these groups. For example, for aldehydes the interaction term is as follows:

$$CH(=O)xCH_2 = n_{CH(=O)} n_{CH_2} \alpha_{CH(=O)xCH_2} \qquad (3)$$

For single aldehydes $n_{CH(=O)} = 1$ so that the net effect of adding the interaction term to the summation in eq 1 is that the toxicity contribution of the CH_2 groups is reduced (i.e., for negative α values). Note this correction term was not added for single esters. The correction term was added for the lone diacid in the data set (adipic acid).

Since groups have different properties depending on whether they are attached to aliphatic or aromatic carbons (AC), additional parameters were added for groups attached to aromatic carbons. The groups which are direct analogues of the aliphatic parameters are as follows: $ACCH_3$, $ACCH_2$, ACCH, ACC, ACC(═O), ACC(═O)O, ACCH(═O), ACO, ACC≡N, $ACNH_2$, ACNH, ACN, $ACNO_2$, ACOH, ACCl, ACBr, ACF, and ACH. The $ACCH_2$-CH_3 parameter was added to improve the accuracy of the predictions since it appeared in five compounds (as opposed to using $ACCH_2 + CH_3$). The ACCH═CH_2 group was added since there was insufficient data to describe other kinds of alkenes attached to aromatic carbons. The $ACCF_3$ group was added because there was insufficient data for fluoro groups attached to an aliphatic carbon. The $ACCH_2Cl$ group was added since benzylic halides were classified as being excessively toxic (*13*). For example, 4-methyl-2-nitroaniline is characterized as follows: $n_{ACH} = 3$, $n_{ACNH_2} = 1$, $n_{ACNO_2} = 1$, $n_{ACCH_3} = 1$, and all other parameters are zero. For the aliphatic diesters an interaction term was added between the C(═O)O group and the CH_2 group. A similar interaction term was added for the aromatic diesters (phthalates).

In this study, several parameters were added to describe compounds that contain more than one benzene ring. The AC parameter was added to describe compounds that contain two phenyl groups attached to the same molecular fragment. For these compounds the connecting molecular fragment can only be assigned to one of the aromatic carbons that it is attached to. The toxicity contribution of the other aromatic carbon attached to the fragment is given by AC. For example for diphenyl ether, $n_{ACO} = 1$, $n_{AC} = 1$, $n_{ACH} = 10$. The AC parameter is not to be used in the same fashion as the ArC parameter of Gao and co-workers. (*9*). For example, the toxicity of a chloro group attached to an aromatic carbon would be estimated by the ACCl group not by adding the value for ArC to the value for Cl (using the groups given by Gao and co-workers). The ACAC parameter was added to describe compounds that contain two phenyl groups directly attached to each other. For example for biphenyl, $n_{ACAC} = 1$, $n_{ACH} = 10$. The Ringfusion C parameter was added to describe compounds which contain carbons that are members of two aromatic rings. For example for naphthalene, $n_{ACH} = 8$, $n_{RFC} = 2$.

According to Hall and co-workers (*7*) the occurrence of two nitro groups ortho or para to each other increases the toxicity significantly. To account for this, correction parameters (o-, m-, and p-) were added for the number of nitro groups that are ortho, meta, or para to a arbitrarily chosen nitro group. For example for 1,3-dinitrobenzene there is one nitro group (on AC#3) that is meta to the nitro group on AC#1. Therefore the group values would be given as $n_o = 0$, $n_m = 1$, $n_p = 0$. Whereas for 1,3,5-trinitrobenzene, two nitro groups are meta to the nitro group on AC#1. Therefore, the group values would be given as $n_o = 0$, $n_m = 2$, $n_p = 0$. Finally the AN parameter was added to describe nitrogens in aromatic rings such as pyridine.

Multilinear Model. The basic equation for the multilinear model is given in eq 2. The multilinear model utilized all 58 of the parameters in Table 2 (including the interaction terms). The parameter values were determined using multilinear regression.

Neural Net Model. In the neural net model, the parameters (with the exception of the interaction terms) were submitted to a three-layer, fully connected, feed-forward computational neural network (CNN). Thus, 54 variables were used as input to the neural network model. The three layers consist of an input layer, a hidden layer, and an output layer. The interaction terms (with CH_2) were not added since the neural net should be able to account for the interactions between some of the groups and the CH_2 group. The adjustable weights were optimized using the back-propagation algorithm. The CNN used in this study is described in further detail in Chapter 3 of Wasserman (*20*).

Model Validation. To evaluate the predictive ability of a model, one must use it to predict values for chemicals not included in the training set (the prediction set). For a particular training trial, 80% of the original data set was placed in the training set, 10% was put in a validation set, and the remaining 10% was placed in the prediction set. Care was taken so that each group was manifested in at least two compounds in the training set. For the CNN model, training was concluded when the root-mean-squared error in the validation set was minimized. Following the completion of training, the fit error for all three data sets was calculated. To avoid overfitting the data, the ratio of the number of chemicals in the dataset to the number of adjustable parameters should be greater than two

Table 1. Experimental 96-h Fathead Minnow LC_{50} Data and Fitted Values Using the Multilinear Model

		−log(LC50)				−log(LC50)	
CAS no.	name	exp	fit	CAS no.	name	exp	fit
110-82-7	cyclohexane	2.96	2.68	96-18-4	1,2,3-trichloropropane	3.41	3.74
67-64-1	acetone	0.85	1.23	110-56-5	1,4-dichlorobutane	3.39	3.58
78-93-3	2-butanone	1.35	1.68	628-76-2	1,5-dichloropentane	3.75	4.03
107-87-9	2-pentanone	1.84	2.13	627-30-5	3-chloro-1-propanol	2.07	2.02
591-78-6	2-hexanone	2.37	2.57	115-20-8	2,2,2-trichloroethanol	2.80	2.74
110-43-0	2-heptanone	2.94	3.02	57-15-8	1,1,1-trichloro-2-methyl-2-propanol	3.12	3.34
111-13-7	2-octanone	3.45	3.47	106-94-5	1-bromopropane	3.26	3.29
821-55-6	2-nonanone	3.97	3.91	109-65-9	1-bromobutane	3.57	3.74
693-54-9	2-decanone	4.50	4.36	111-25-1	1-bromohexane	4.68	4.63
6175-49-1	2-dodecanone	5.19	5.25	629-04-9	1-bromoheptane	5.09	5.08
563-80-4	3-methyl-2-butanone	2.00	2.00	111-83-1	1-bromooctane	5.36	5.52
108-10-1	4-methyl-2-pentanone	2.27	2.45	109-64-8	1,3-dibromopropane	5.05	4.90
75-97-8	3,3-dimethyl-2-butanone	3.06	2.28	107-10-8	*n*-propylamine	2.28	2.02
96-22-0	3-pentanone	1.75	2.13	109-73-9	*n*-butylamine	2.44	2.47
110-12-3	5-methyl-2-hexanone	2.86	2.90	33966-50-6	*sec*-butylamine	2.42	2.34
502-56-7	5-nonanone	3.66	3.91	110-58-7	*n*-pentylamine	2.69	2.91
108-94-1	cyclohexanone	2.19	2.23	111-26-2	*n*-hexylamine	3.25	3.36
76-22-2	camphor	3.14	2.97	111-68-2	*N*-heptylamine	3.72	3.81
67-56-1	methanol	0.05	0.40	111-86-4	*n*-octylamine	4.40	4.25
64-17-5	ethanol	0.52	0.85	112-20-9	*n*-nonylamine	4.82	4.70
71-23-8	1-propanol	1.12	1.30	2016-57-1	*n*-decylamine	5.18	5.15
71-36-3	1-butanol	1.59	1.74	7307-55-3	n-undecylamine	5.91	5.59
71-41-0	1-pentanol	2.21	2.19	124-22-1	n-dodecylamine	6.26	6.04
111-27-3	1-hexanol	2.94	2.64	2869-34-3	n-tridecylamine	6.48	6.49
111-70-6	1-heptanol	3.51	3.08	13952-84-6	2-butanamine	2.42	2.34
111-87-5	1-octanol	3.98	3.53	598-74-3	1,2-dimethylpropylamine	2.49	2.67
143-08-8	1-nonanol	4.41	3.98	5813-64-9	2,2-dimethyl-1-propylamine	2.26	2.62
112-30-1	1-decanol	4.82	4.42	15673-00-4	3,3-dimethylbutylamine	2.23	3.07
112-42-5	1-undecanol	5.22	4.87	107-45-9	*tert*-octylamine	3.72	3.67
112-53-8	1-dodecanol	5.27	5.32	693-16-3	1-methylheptylamine	4.40	4.13
67-63-0	2-propanol	0.78	1.17	141-43-5	2-aminoethanol	1.47	1.19
78-92-2	2-butanol	1.31	1.62	111-42-2	diethanolamine	1.85	0.78
104-76-7	2-ethyl-1-hexanol	3.66	3.41	78-96-6	1-amino-2-propanol	1.47	1.51
78-83-1	2-methyl-1-propanol	1.69	1.62	109-85-3	2-methoxyethylamine	2.16	1.78
75-65-0	2-methyl-2-propanol	1.06	1.45	109-89-7	diethylamine	1.93	2.01
600-36-2	2,4-dimethyl-3-pentanol	2.85	2.71	143-16-8	di-*n*-hexylamine	5.38	5.58
77-74-7	3-methyl-3-pentanol	2.18	2.35	110-73-6	2-(ethylamino) ethanol	1.78	1.63
108-93-0	cyclohexanol	2.06	2.17	100-37-8	*N*,*N*-diethylethanolamine	1.82	2.10
111-46-6	diethylene glycol	0.15	0.66	96-80-0	2-(diisopropylamino)-ethanol	2.86	2.75
107-21-1	ethylene glycol	0.04	0.00	105-14-6	5-(diethylamino)-2-pentanone	2.67	3.38
107-41-5	hexylene glycol	1.13	1.38	22104-62-7	4-(dimethylamino)-3-methyl-2-butanone	4.18	2.36
57-55-6	1,2-propylene glycol	0.13	0.33	102-69-2	tripropylamine	3.45	3.83
2216-51-5	L-menthol	3.92	3.59	91-65-6	*N*,*N*-diethylcyclohexylamine	3.86	3.81
75-09-2	dichloromethane	2.42	2.24	103-76-4	1-(2-hydroxyethyl)piperazine	1.31	1.65
67-66-3	chloroform	3.06	2.84	140-31-8	*N*-aminoethyl piperazine	1.77	2.37
56-23-5	carbon tetrachloride	3.56	3.40	78-90-0	1,2-propanediamine	1.81	2.24
107-06-2	1,2-dichloroethane	2.90	2.69	107-15-3	ethylenediamine	2.55	1.91
71-55-6	1,1,1-trichloroethane	3.40	3.12	96-29-7	methyl ethyl ketoxime	2.01	2.25
79-00-5	1,1,2-trichloroethane	3.21	3.29	127-06-0	acetone oxime	2.12	1.81
79-34-5	1,1,2,2-tetrachloroethane	3.92	3.89	100-64-1	cyclohexanone oxime	2.74	2.81
76-01-7	pentachloroethane	4.43	4.45	1634-04-4	methyl *tert*-butyl ether	2.12	2.04
67-72-1	hexachloroethane	5.23	5.01	142-96-1	di-*n*-butyl ether	3.61	3.67
58-89-9	hexachlorocyclohexane	6.52	6.30	60-29-7	diethyl ether	1.46	1.89
78-87-5	1,2-dichloropropane	2.93	3.01	108-20-3	diisopropyl ether	2.11	2.53
142-28-9	1,3-dichloropropane	2.94	3.13	693-65-2	di-*n*-pentyl ether	4.71	4.57
109-87-5	dimethoxymethane	1.04	1.20	110-65-6	2-butyne-1,4,-diol	3.21	2.61
123-91-1	1,4-dioxane	0.93	1.31	764-01-2	2-butyn-1-ol	3.84	3.46
110-88-3	trioxane	1.18	0.62	818-72-4	1-octyn-3-ol	5.49	5.74
109-99-9	tetrahydrofuran	1.52	1.55	2028-63-9	3-butyn-2-ol	3.78	3.96
470-82-6	1,8-epoxy-p-menthane	3.18	3.18	7383-19-9	1-heptyn-3-ol	4.80	5.30
64-19-7	acetic acid	2.85	2.88	115-19-5	2-methyl-3-butyn-2-ol	1.41	1.48
109-52-4	*n*-pentanoic acid	3.12	3.01	78-27-3	1-ethynyl-1-cyclohexanol	2.69	2.48
142-62-1	n-hexanoic acid	2.76	3.05	77-75-8	3-methyl-1-pentyn-3-ol	1.91	1.93
112-05-0	*n*-nonanoic acid	3.18	3.18	127-66-2	2-phenyl-3-butyn-2-ol	3.11	3.48
124-04-9	adipic acid	3.18	3.08	19549-98-5	3,6-dimethyl-1-heptyn-3-ol	3.46	3.15
75-07-0	acetaldehyde	3.11	3.17	3923-52-2	1,1-diphenyl-2-propyn-1-ol	4.27	4.81
123-72-8	butanal	3.65	3.45	1482-15-1	3,4-dimethyl-1-pentyn-3-ol	2.74	2.25
110-62-3	pentanal	3.82	3.59	95-63-6	1,2,4-trimethylbenzene	4.19	3.94
123-15-9	2-methylvaleraldehyde	3.73	4.22	2416-94-6	2,3,6-trimethylphenol	4.22	4.10
66-25-1	hexanal	3.66	3.73	527-60-6	2,4,6-trimethylphenol	4.02	4.10
590-86-3	3-methyl-butanal	4.42	4.08	25167-83-3	tetrachlorophenol	6.13	5.48
96-17-3	2-methylbutyraldehyde	3.94	4.08	91-23-6	*o*-nitroanisole	2.90	3.25
141-78-6	ethyl acetate	2.58	2.63	98-86-2	acetophenone	2.87	3.03
79-20-9	methyl acetate	2.27	2.18	100-51-6	benzyl alcohol	2.37	2.49
109-21-7	*n*-butyl *n*-butyrate	4.09	4.41	623-25-6	1,4-bis(chloromethyl) benzene	6.65	6.46

Table 1 (Continued)

CAS no.	name	−log(LC50) exp	−log(LC50) fit	CAS no.	name	−log(LC50) exp	−log(LC50) fit
123-86-4	*n*-butyl acetate	3.81	3.52	100-44-7	benzyl chloride	4.40	4.85
540-88-5	*tert*-butyl acetate	2.55	3.23	100-46-9	benzylamine	3.02	3.21
142-92-7	*n*-hexyl acetate	4.56	4.41	100-52-7	benzaldehyde	3.93	4.02
123-66-0	ethyl hexanoate	4.21	4.41	100-10-7	4-(dimethylamino)-benzaldehyde	3.51	3.89
109-60-4	*n*-propyl acetate	3.23	3.07	122-03-2	4-isopropyl benzaldehyde	4.35	4.77
111-15-9	2-ethoxyethyl acetate	3.50	3.28	446-52-6	2-fluorobenzaldehyde	4.96	4.22
108-59-8	dimethyl malonate	4.03	3.29	6361-21-3	2-chloro-5-nitrobenzaldehyde	4.67	4.69
87-91-2	diethyl L-(+)-tartrate	2.50	2.88	104-88-1	4-chlorobenzaldehyde	4.81	4.54
105-53-3	diethyl malonate	4.01	3.63	552-89-6	2-nitrobenzaldehyde	4.01	4.17
123-25-1	diethyl succinate	3.09	3.80	555-16-8	4-nitrobenzaldehyde	4.18	4.17
141-28-6	diethyl adipate	4.05	4.13	613-45-6	2,4-dimethoxybenzaldehyde	3.92	3.73
110-40-7	diethyl sebacate	4.98	4.80	1761-61-1	2-hydroxy-5-bromobenzaldehyde	5.19	4.83
105-99-7	dibutyl adipate	4.85	4.80	635-93-8	2-hydroxy-5-chlorobenzaldehyde	5.31	4.70
818-61-1	2-hydroxyethyl acrylate	4.38	4.13	90-02-8	2-hydroxybenzaldehyde	4.73	4.17
140-88-5	ethyl acrylate	4.60	4.51	121-33-5	3-methoxy-4-hydroxybenzaldehyde	3.12	4.03
106-63-8	isobutyl acrylate	4.79	5.28	708-76-9	2-hydroxy-4,6-dimethoxybenzaldehyde	4.83	3.89
999-61-1	2-hydroxypropyl acrylate	4.59	4.45	653-37-2	pentafluorobenzaldehyde	5.25	5.05
868-77-9	2-hydroxyethyl methacrylate	2.76	2.68	387-45-1	2-chloro-6-fluorobenzaldehyde	4.23	4.75
80-62-6	methyl methacrylate	2.50	2.62	874-42-0	2,4-dichlorobenzaldehyde	4.99	5.06
2370-63-0	2-Ethoxyethyl methacrylate	3.76	3.72	58-90-2	2,3,4,6-tetrachlorophenol	5.35	5.48
75-05-8	acetonitrile	1.49	1.26	4901-51-3	2,3,4,5-tetrachlorophenol	5.74	5.48
107-12-0	propionitrile	1.56	1.71	3481-20-7	2,3,5,6-Tetrachloroaniline	5.93	5.14
2243-27-8	*n*-octyl cyanide	4.30	4.39	732-26-3	2,4,6-tri-*tert*-butylphenol	6.63	6.52
764-13-6	2,5-dimethyl-2,4-hexadiene	4.46	4.39	150-76-5	4-methoxyphenol	3.11	3.25
513-81-5	2,3-dimethyl-1,3-butadiene	4.08	3.60	120-07-0	*N*-phenyldiethanolamine	2.39	2.77
5194-50-3	2,4-hexadiene	3.61	3.60	103-83-3	*N*,*N*-dimethylbenzylamine	3.55	3.23
5989-27-5	d-limonene	5.29	4.87	150-19-6	1-hydroxy-3-methoxybenzene	3.22	3.25
77-73-6	dicyclopentadiene	3.63	3.87	150-78-7	1,4-dimethoxybenzene	3.07	2.95
1647-16-1	1,9-decadiene	5.68	5.49	5673-07-4	2,6-Dimethoxytoluene	3.88	3.19
78-79-5	isoprene	2.95	3.21	13608-87-2	2,3,4-trichloroacetophenone	5.05	4.59
75-35-4	1,1-dichloroethylene	2.84	2.98	95-95-4	2,4,5-trichlorophenol	5.34	4.96
79-01-6	trichloroethylene	3.47	3.65	88-06-2	2,4,6-trichlorophenol	4.64	4.96
127-18-4	tetrachloroethylene	3.91	4.33	937-20-2	2,4-dichloroacetophenone	4.21	4.07
4117-14-0	2-decyn-1-ol	5.16	6.14	70-69-9	4-aminopropiophenone	3.01	3.29
107-19-7	propargyl alcohol	4.56	3.63	102-27-2	*N*-ethyl-m-toluidine	3.44	3.72
100-61-8	*N*-methylaniline	3.03	3.03	618-87-1	3,5-dinitroaniline	3.93	3.91
121-69-7	*N*,*N*-dimethylaniline	3.27	3.11	35572-78-2	2-methyl-3,5-dinitroaniline	4.12	4.15
91-66-7	*N*,*N*-diethylaniline	3.96	4.00	56207-39-7	2-methyl-3,6-dinitroaniline	5.37	5.26
121-87-9	2-chloro-4-nitroaniline	3.93	3.72	10202-92-3	3-methyl-2,4-dinitroaniline	4.25	4.15
59-50-7	p-chloro-*m*-cresol	4.40	4.15	70343-06-5	3-methyl-2,6-dinitroaniline	4.21	4.15
573-56-8	2,6-dinitrophenol	3.67	4.26	19406-51-0	4-amino-2,6-dinitrotoluene	4.46	4.15
24544-04-5	2,6-diisopropylaniline	4.10	4.55	616-73-9	2,4-dinitro-5-methylphenol	4.79	4.49
10031-82-0	4-ethoxybenzaldehyde	3.74	4.32	6284-83-9	1,3,5-trichloro-2,4-dinitrobenzene	6.09	5.67
97-02-9	2,4-dinitroaniline	4.08	3.91	88-73-3	1-chloro-2-nitrobenzene	3.73	3.91
6393-42-6	2,6-dinitro-4-methylaniline	4.12	4.15	108-90-7	chlorobenzene	3.70	3.76
634-93-5	2,4,6-trichloroaniline	4.59	4.61	108-41-8	1-chloro-3-methylbenzene	3.84	3.99
634-67-3	2,3,4-trichloroaniline	4.74	4.61	106-43-4	1-chloro-4-methylbenzene	4.33	3.99
615-65-6	2-chloro-4-methylaniline	3.59	3.81	121-73-3	*m*-chloronitrobenzene	3.93	3.91
71-43-2	benzene	3.50	3.24	95-50-1	o-dichlorobenzene	4.19	4.28
1746-23-2	p-*tert*-butylstyrene	5.51	5.06	541-73-1	m-dichlorobenzene	4.27	4.28
100-42-5	styrene	3.51	4.02	106-46-7	p-dichlorobenzene	4.27	4.28
30030-25-2	chloromethyl styrene	5.69	5.63	95-49-8	2-chlorotoluene	4.23	3.99
1745-81-9	2-allyl phenol	3.95	4.26	95-73-8	2,4-dichlorotoluene	4.54	4.52
97-53-0	eugenol	3.84	4.12	120-82-1	1,2,4-trichlorobenzene	4.80	4.80
89-62-3	4-methyl-2-nitroaniline	3.79	3.44	87-61-6	1,2,3-trichlorobenzene	4.89	4.80
108-86-1	bromobenzene	4.45	3.89	108-70-3	1,3,5-trichlorobenzene	4.74	4.80
608-71-9	pentabromophenol	6.72	6.66	95-94-3	1,2,4,5-tetrachlorobenzene	5.83	5.32
106-37-6	1,4-dibromobenzene	5.28	4.54	634-66-2	1,2,3,4-tetrachlorobenzene	5.29	5.32
118-79-6	1,3,5-tribromo-2-hydroxybenzene	4.71	5.35	87-86-5	pentachlorophenol	6.04	6.00
106-40-1	4-bromoaniline	3.56	3.70	771-60-8	pentafluoroaniline	3.69	4.08
98-95-3	nitrobenzene	3.02	3.39	350-46-9	1-fluoro-4-nitrobenzene	3.70	3.59
528-29-0	o-dinitrobenzene	5.45	5.02	371-40-4	1-amino-4-fluorobenzene	3.82	3.25
99-65-0	m-dinitrobenzene	4.12	4.10	128-37-0	2,6-di-*tert*-butyl-4-methylphenol	5.78	5.71
100-25-4	p-dinitrobenzene	5.39	5.21	108-95-2	phenol	3.50	3.39
108-88-3	toluene	3.42	3.47	105-67-9	2,4-xylenol	3.86	3.86
88-72-2	*o*-nitrotoluene	3.56	3.62	95-65-8	3,4-xylenol	3.94	3.86
99-08-1	*m*-nitrotoluene	3.64	3.62	95-75-0	3,4-dichlorotoluene	4.74	4.52
99-99-0	*p*-nitrotoluene	3.44	3.62	1126-79-0	*n*-butyl phenyl ether	4.42	4.43
119-34-6	4-amino-2-nitrophenol	3.63	3.36	39905-57-2	4-hexyloxyaniline	4.78	5.14
4920-77-8	3-methyl-2-nitrophenol	3.52	3.78	122-99-6	2-phenoxyethanol	2.60	3.16
700-38-9	5-methyl-2-nitrophenol	3.51	3.78	95-48-7	*o*-cresol	3.90	3.63
88-75-5	2-nitrophenol	2.94	3.54	108-39-4	*m*-cresol	3.29	3.63
100-02-7	4-nitrophenol	3.53	3.54	106-44-5	*p*-cresol	3.76	3.63
602-01-7	2,3-dinitrotoluene	4.99	5.25	95-57-8	*o*-chlorophenol	4.00	3.91
121-14-2	2,4-dinitrotoluene	3.77	4.34	106-48-9	*p*-chlorophenol	4.46	3.91

Table 1 (Continued)

CAS no.	name	−log(LC50) exp	−log(LC50) fit	CAS no.	name	−log(LC50) exp	−log(LC50) fit
619-15-8	2,5-dinitrotoluene	5.15	5.44	120-83-2	2,4-dichlorophenol	4.30	4.43
606-20-2	2,6-dinitrotoluene	3.98	4.34	62-53-3	aniline	3.03	3.05
610-39-9	3,4-dinitrotoluene	5.08	5.25	95-80-7	2,4-diaminotoluene	1.93	3.10
618-85-9	3,5-dinitrotoluene	3.91	4.34	3698-83-7	1,3-dichloro-4,6-dinitro benzene	3.78	5.14
534-52-1	2-methyl-4,6-dinitrophenol	5.01	4.49	121-32-4	3-ethoxy-4-hydroxybenzaldehyde	3.28	4.48
616-72-8	1,5-dimethyl-2,4-dinitrobenzene	4.39	4.57	118-74-1	hexachlorobenzene	4.11	6.37
51-28-5	1,3-dinitro-4-hydroxybenzene	4.22	4.26	100-01-6	*p*-nitroaniline	3.08	3.20
99-35-4	1,3,5-trinitrobenzene	5.32	4.81	95-51-2	*o*-chloroaniline	4.35	3.57
118-96-7	2,4,6-trinitrotoluene	4.98	5.05	106-47-8	*p*-chloroaniline	3.62	3.57
603-83-8	2-amino-6-nitrotoluene	3.48	3.44	95-76-1	3,4-dichloroaniline	4.32	4.09
578-46-1	3-amino-4-nitrotoluene	3.79	3.44	554-00-7	2,4-dichloroaniline	4.07	4.09
119-32-4	4-amino-2-nitrotoluene	3.77	3.44	106-49-0	4-methylaniline	2.83	3.28
99-52-5	2-methyl-4-nitroaniline	3.24	3.44	95-47-6	*o*-xylene	3.81	3.71
99-55-8	2-methyl-5-nitroaniline	3.34	3.44	108-38-3	*m*-xylene	3.82	3.71
570-24-1	2-methyl-6-nitroaniline	3.80	3.44	106-42-3	*p*-xylene	4.21	3.71
1689-84-5	1-cyano-3,5-dibromo-4-hydroxybenzene	4.30	4.61	2176-62-7	pentachloropyridine	5.73	5.15
5922-60-1	1-cyano-2-amino-5-chlorobenzene	3.73	3.48	939-23-1	4-phenyl pyridine	3.98	3.87
6575-09-3	2-chloro-6-methyl-benzonitrile	4.00	3.91	4214-79-3	5-chloro-2-pyridinol	2.06	3.22
529-19-1	1-cyano-2-methylbenzene	3.42	3.39	91-20-3	naphthalene	4.32	4.34
619-24-9	3-nitrobenzonitrile	3.39	3.30	90-15-3	1-naphthalenol	4.54	4.49
619-72-7	4-nitrobenzonitrile	3.78	3.30	135-19-3	2-naphthol	4.62	4.49
100-47-0	benzonitrile	2.98	3.15	90-12-0	1-methylnaphthalene	4.20	4.57
100-41-4	ethylbenzene	3.59	3.75	86-57-7	1-nitronaphthalene	4.46	4.49
141-93-5	m-diethylbenzene	4.51	4.27	90-13-1	1-chloronaphthalene	4.85	4.86
123-07-9	*p*-ethylphenol	4.07	3.91	91-22-5	quinoline	3.45	3.65
589-16-2	4-ethylaniline	3.22	3.57	260-94-6	acridine	4.89	4.75
104-13-2	4-butylaniline	4.17	4.02	253-52-1	phthalazine	3.11	2.95
16245-79-7	4-octylaniline	6.24	5.81	1129-35-7	methyl 4-cyanobenzoate	3.54	3.45
37529-30-9	4-decylaniline	6.57	6.70	94-09-7	ethyl 4-aminobenzoate	3.67	3.80
88-85-7	2-*sec*-butyl-4,6-dinitrophenol	6.19	5.45	1126-46-1	methyl 4-chlorobenzoate	4.15	4.06
80-46-6	p-*tert*-amylphenol	4.80	4.88	368-77-4	3-(trifluoromethyl)benzonitrile;	3.56	3.81
98-54-4	p-*tert*-butylphenol	4.47	4.43	88-30-2	3-(trifluoromethyl)-4-nitrophenol	4.36	4.20
89-83-8	thymol	4.67	4.38	1582-09-8	2,6-dinitro-*N*,*N*-dipropyl-4-(trifluoro-methyl)benzeneamine	6.50	6.42
98-82-8	cumene	4.28	3.99				
538-68-1	*n*-pentylbenzene	4.94	4.65	831-82-3	4-phenoxyphenol	4.58	4.58
25154-52-3	nonylphenol	6.22	6.60	119-61-9	benzophenone	4.11	4.36
131-11-3	dimethyl phthalate	3.70	3.84	101-84-8	diphenyl ether	4.63	4.42
84-66-2	diethyl phthalate	4.12	4.28	14548-46-0	4-benzoyl pyridine	3.25	3.67
84-69-5	diisobutyl phthalate	5.49	5.82	122-39-4	diphenylamine	4.65	4.37
84-74-2	dibutyl phthalate	5.33	5.17	118-55-8	phenyl salicylate	5.27	5.02
84-62-8	diphenyl phthalate	6.60	6.50	620-88-2	1-nitro-4-phenoxybenzene	4.91	4.58
93-89-0	ethyl benzoate	4.23	3.98	97-23-4	2,2′-methylenebis(4-chloro)phenol	5.94	6.00
619-50-1	methyl 4-nitrobenzoate	3.89	3.69	91-94-1	3,3′-dichlorobenzidine	5.09	5.23
90-43-7	*o*-phenylphenol	4.50	4.71	92-52-4	biphenyl	4.80	4.56
109-06-8	2-methylpyridine	2.02	2.78	104-90-5	5-ethyl-2-methyl pyridine	3.17	3.30
108-99-6	3-methylpyridine	2.81	2.78	1122-54-9	4-acetylpyridine	2.86	2.34
108-89-4	4-methylpyridine	2.36	2.78	5683-33-0	2-(dimethylamino)pyridine	2.98	2.42
100-70-9	2-pyridinecarbonitrile	2.16	2.46	110-86-1	pyridine	2.90	2.54
500-22-1	3-pyridinecarboxaldehyde	3.81	3.33	2859-67-8	3-pyridinepropanol	2.96	2.69

(*6*). Therefore, the maximum number of nodes in the hidden layer is three. For the multilinear model the validation set is equivalent to another prediction set.

Results and Discussion

Results Using Entire Data Set. *Multilinear Model.* The multilinear model was first fit to the entire dataset of 397 chemicals. The results of the fit are shown in Figure 1 and in Table 1. The correlation coefficient (r^2) was 0.91, and the RMSE was 0.37. If the interaction terms with CH_2 are not included, the correlation coefficient was 0.89 and the RMSE was 0.42.

The coefficients of the multilinear model are given in Table 2. The aliphatic hydrocarbon parameters (CH_3, CH_2, CH, and C) were fairly well behaved in that their values decreased monotonically from 0.62 to −0.19. In the CLOGP method, the aliphatic functional groups are described using a C and an H parameter. This would indicate that the aliphatic parameters should be a linear function of the number of hydrogen atoms attached. The aliphatic parameters were approximately a linear function of the number of hydrogen atoms ($r^2 = 0.980$). The aromatic hydrocarbon parameters ($ACCH_2CH_3$, $ACCH_3$, $ACCH_2$, ACCH, and ACC) also decreased monotonically in their values from 1.06 to −0.27 with the exception of the parameter for $ACCH_2$. The parameter for $ACCH_2$ may have been near zero due to benzylic interaction with another functional groups (such as in the case of benzylic halides which required the addition of another parameter). The parameters for the aliphatic and aromatic amines varied linearly with the number of hydrogens ($r^2 = 0.9997$ and 0.9980, respectively).

In this study, different modes of action have been accounted for through the addition of special structural fragments (such as the acrylate group). However, complex compounds such as phosphates and some polycyclic compounds could not be described using a simple set of parameters. However, in most cases, it was unclear

Table 2. Parameters Utilized in the Group Contribution Model

group	description	value	SE[a]	t_o[b]	no.[c]	f[d]
CH_3	methyl	0.6172	0.0344	17.96	185	340
CH_2		0.4464	0.0104	42.77	163	604
CH		0.1522	0.0458	3.32	45	64
C		-0.1861	0.0978	1.90	22	27
C═C	ethenyl	0.7417	0.2212	3.35	12	19
H	no. of H on ethenyl	0.2212	0.1039	2.13	11	37
C≡CH	C≡CH (1°/2° OH)	3.4004	0.2114	16.08	4	4
C≡C	C≡C (1°/2° OH)	2.6072	0.2414	10.80	3	3
C≡CH	C≡CH (3° OH)	0.6475	0.1792	3.61	7	7
C═NOH	oxime	0.5733	0.2342	2.45	3	3
CH═O	aldehyde	2.5539	0.2747	9.30	7	7
C═O	ketone	0.0004	0.1142	0.00	19	19
C≡N	nitrile	0.6447	0.233	2.77	3	3
COOH	carboxyl	2.2621	0.3281	6.90	5	6
COO	ester	0.9461	0.1076	8.80	16	23
acrylate	acrylate	3.445	0.2027	17.00	4	4
methacrylate	methacrylate	1.9986	0.2325	8.59	3	3
O	ether O	−0.2392	0.0804	2.98	14	18
NH_2	primary amine	0.511	0.0816	6.26	25	27
NH	secondary amine	-0.1157	0.1757	0.66	6	6
N	tertiary amine	-0.7043	0.1481	4.75	9	9
Br	bromine	1.7817	0.1371	12.99	6	7
Cl	chlorine	0.897	0.0342	26.22	21	65
OH	hydroxyl	-0.2125	0.0786	2.70	49	49
2OH	diol	-0.8889	0.1534	5.79	8	8
CH═O × CH_2	interaction between aldehyde and CH2	-0.3057	0.1213	2.52	6	13
COOH × CH_2	interaction between carboxyl and CH2	-0.4038	0.0794	5.09	4	22
COO × CH_2	interaction between diester and CH2	-0.1393	0.0183	7.61	7	72
$ACCH_2CH_3$	AC + ethyl	1.0578	0.1464	7.22	5	6
$ACCH_3$	AC+ methyl	0.7755	0.0481	16.12	60	72
$ACCH_2$		0.002	0.1329	0.01	12	12
ACCH		0.0552	0.1606	0.34	5	6
ACC		-0.2708	0.1405	1.93	7	10
AC	see exp. section	-0.7484	0.1537	4.87	10	11
$ACCH═CH_2$	AC + ethenyl	1.3239	0.2435	5.44	3	3
$ACCF_3$	AC + CF3	1.196	0.2452	4.88	3	3
ACC═O	AC + ketone	-0.2842	0.1683	1.69	7	7
ACCOO	AC + ester	0.2243	0.1159	1.93	11	16
ACCH═O	AC + aldehyde	1.3221	0.1024	12.91	20	20
ACO	AC + ether O	-0.2208	0.0894	2.47	17	21
ACC≡N	AC + nitrile	0.4533	0.1352	3.35	10	10
ACNH2	AC + primary amine	0.3522	0.0682	5.17	42	44
ACNH	AC + secondary amine	-0.2793	0.2448	1.14	3	3
ACN	AC + tertiary amine	-0.8208	0.1892	4.34	6	6
AN	N in aromatic ring	-0.1519	0.1002	1.51	17	18
$ACNO_2$	AC + nitro	0.6918	0.0823	8.40	59	89
ACOH	AC + hydroxyl	0.6942	0.0651	10.66	56	57
ACH_2Cl	benzylic chloride	2.1514	0.1705	12.62	3	4
ACCl	AC + chlorine	1.0609	0.0238	44.61	53	115
ACBr	AC + bromine	1.1932	0.0619	19.28	7	15
ACF	AC + florine	0.7447	0.0561	13.26	6	14
ACH	unsubstituted AC	0.5393	0.0146	36.90	216	872
o	see Experimental Section	1.4769	0.2762	5.35	3	3
m	see Experimental Section	0.5607	0.1603	3.50	22	24
p	see Experimental Section	1.6663	0.274	6.08	3	3
RINGFUSION_C	ACs that are members of >1 rings	0.0114	0.0725	0.16	9	20
ACAC	AC bonded to AC	-0.8331	0.2313	3.60	4	4
ACCOO × CH_2	interaction between ar-diester and CH2	-0.1126	0.0385	2.93	3	20

[a] Standard error of a parameter. [b] t statistic = parameter value/standard error (for 95% confidence in a parameter the t statistic must be greater than 1.96). [c] Number of chemicals containing a given parameter. [d] Frequency (total number of times a group appears).

whether it was a failure of the model or due to a lack of sufficient consistent data.

The multilinear model successfully fit the toxicities of 2,2,2-trichloroethanol and 1,1,1-trichloro-2-methyl-2-propanol which are β-halogenated alcohols. β-Halogenated alcohols have been classified as being excessively toxic (*13*). The presence of three chlorines in the β position probably makes these compounds unreactive.

In most cases, the relative error in the parameters was fairly small. In the cases where a group failed the t-test (where the t_0 value was less than 1.96), the absolute error was fairly small and the value of the parameter was near zero. Thus, the uncertainty in the prediction for compounds containing these groups will also be reasonably small. For groups that appear in less than six chemicals in the data set there is a greater chance that experimental coincidence may yield an inaccurate parameter value. These groups have been included to allow the model to predict toxicities for a wider variety of organic chemicals. Toxicities estimated using these parameters should be

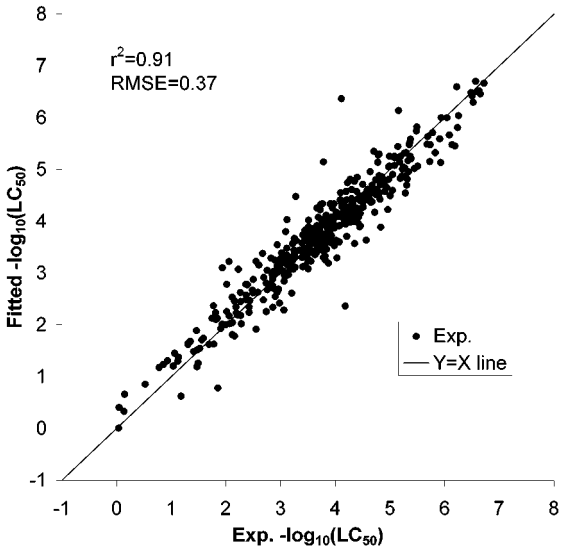

Figure 1. Fitted $-\log(LC_{50})$ vs experimental $-\log(LC_{50})$ using the multilinear model.

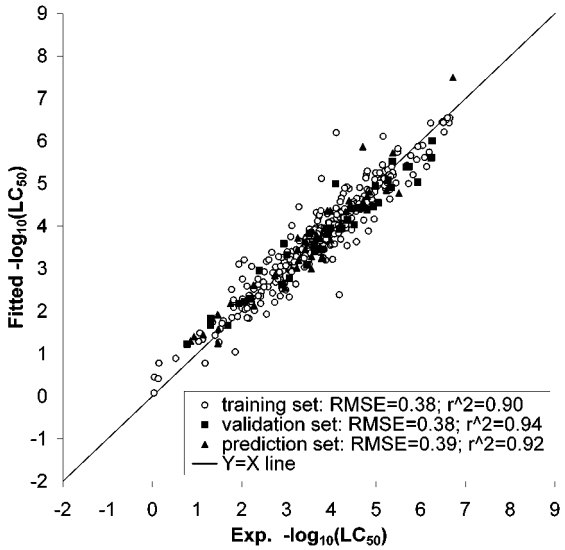

Figure 2. Fitted/predicted $-\log(LC_{50})$ vs experimental $-\log(LC_{50})$ using the multilinear model.

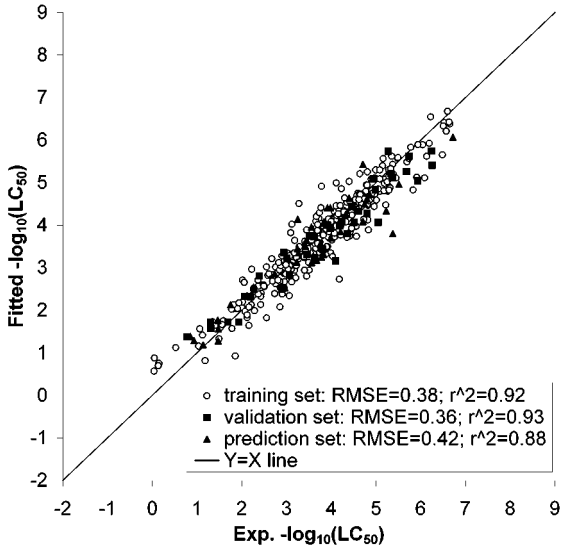

Figure 3. Fitted/predicted $-\log(LC_{50})$ vs experimental $-\log(LC_{50})$ using the CNN model.

used with caution until additional experimental data becomes available to reinforce the validity of these parameters.

The interaction between CH_2 and aldehydes, carboxylic acids, aliphatic diesters, and aromatic diesters effectively reduced the value for the CH_2 parameter from 0.446 to 0.141, 0.043, 0.168, and 0.221, respectively.

Several compounds deviated significantly from the multilinear model predictions. The experimental value for 4-(dimethylamino)-3-methyl-2-butanone is possibly in error since its toxicity $[-\log(LC_{50})]$ is 1.5 log units higher than 5-diethylamino-2-pentanone. The prediction for diethanolamine may be in error since the 2 OH parameter may not be valid in the presence of an amine group. The experimental value for 5-chloro-2-pyridinol (2.06) was about one log unit lower than the other pyridines in the data set. One would have expected that the chloro group would have made it more toxic compared to the other pyridine compounds in the data set. The toxicity of 2,4-diaminotoluene (1.93) was very low compared to other substituted benzene compounds. Its lower toxicity can possibly be attributed to two amino groups on the same benzene ring. The toxicity of 1,3-dichloro-4,6-dinitro benzene (3.78) was much less than the value for 1,3,5-trichloro-2,4-dinitrobenzene (6.09). One would expect that the removal of one chlorine group would not result in that great of a reduction (more than 2 orders of magnitude) in the toxicity. The toxicities of 1-octyn-3-ol and 2-decyn-1-ol were over predicted by 0.6 and 1.7 log units, respectively. This can explained by the fact that as chain length increases a molecule becomes more hydrophobic and less reactive. Thus, physical toxicity begins to predominate over chemical toxicity (*22*). The toxicity of hexachlorobenzene (4.11) should have been higher than pentachlorophenol (6.04) and 1,2,4,5-tetrachlorobenzene (5.83). The fitted error values of 3-ethoxy-4-hydroxybenzaldehyde (−1.20), 3-methoxy-4-hydroxybenzaldehyde (−0.91), and 2-hydroxy-4,6-dimethoxybenzaldehyde (0.94) were inconsistent.

Neural Network Model. Since neural nets are trained by monitoring the error in an external validation set, the neural net model was not fit to the entire data set.

Model Validation. The predictions for the training, validation, and prediction sets for the multilinear model and the neural network model are shown in Figures 2 and 3, respectively. The linear model performed slightly better than the CNN model.

The predictions shown in Figures 2 and 3 are for a single training and validation set. When the process of selecting these sets is repeated 100 times, again the linear model performs slightly better (see Figure 4). Using two nodes in the hidden layer yielded the same results for the prediction set (average RMSE = 0.52). Including the interaction terms in the CNN model only decreases the average RMSE to 0.51. For the multilinear model, removing the interaction terms increases the average RMSE for the prediction set from 0.45 to 0.49.

Eldred and co-workers (*6*) also observed that a CNN model did not significantly improve predictions over a linear toxicity model. A possible explanation for the lack of improvement in this study is that the large number of parameters makes it difficult to capture all the interactions between the group variables. To improve the CNN

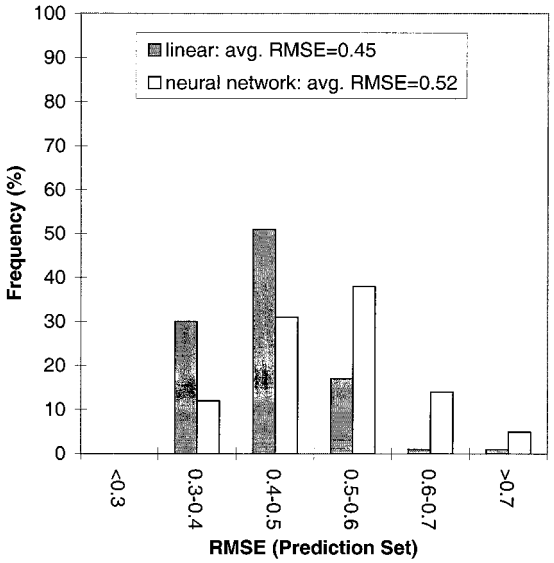

Figure 4. Distribution of validation set RMSE's for the multlinear and CNN models (obtained from 100 selections of the validation and training sets).

model, one would need to add more experimental data so that more nodes could be added to the model. The advantage of the neural network, however, is that it yields the same results with or without the interaction parameters for various groups with the CH_2 group. An advantage of the linear model is that the predictions are much easier to perform (all that is needed is eq 2 and Table 2).

Conclusions

A group contribution model based on 2-D structural descriptors was developed to correlate the aquatic toxicity of 397 organic chemicals. The structural descriptors were related to toxicity using a multilinear model and a computational neural network model. Both models fit the experimental data fairly well ($r^2 > 0.9$). This study illustrated that acute toxicity to the fathead minnow can be successfully correlated using only a compounds' 2-D molecular structure for a large number of chemicals.

Acknowledgment. Todd M. Martin was supported by an appointment to the Postdoctoral Research Program at the National Risk Management Research Laboratory administered by the Oak Ridge Institute for Science and Education (ORISE) *through an interagency agreement between the U.S. Department of Energy and the U.S. Environmental Protection Agency*. This article may or may not reflect the views of the supporting agencies.

References

(1) Konemann, H. (1981) Quantitative structure−activity relationships in fish toxicity Studies. *Toxicology* **19**, 209−221.
(2) Nendza, M., and Russom, C. L. (1991) QSAR modeling of the ERL-D Fathead Minnow Acute Toxicity Database. *Xenobiotica* **21**, 147−170.
(3) Newsome, L. D., Johnson, D. E., Lipnick, R. L., Broderius, S. J., and Russom, C. L. (1991) A QSAR study of the toxicity of amines to the fathead minnow. *Sci. Total Environ.* **109/110**, 537−551.
(4) Schultz, T. W., Wilke, T. S., Bryant, S. E., and Hosein, L. M. (1991) QSARS for selected aliphatic and aromatic amines. *Sci. Total. Environ.* **109/110**, 581−587.
(5) Karabunarliev, S., Mekenyan, O. G., Karcher, W., Russom, C. L., and Bradbury, S. P. (1996) Quantum-chemical descriptors for estimating the acute toxicity of substituted benzenes to the guppy (*Poecilia reticulata*) and fathead minnow (*Pimephales promelas*). *Quant. Struct.-Act Relat.* **15**, 311−320.
(6) Eldred, D. V., Weikel, C. L., Jurs, P. C., and Kaiser, K. L. E. (1999) Prediction of fathead minnow acute toxicity of organic compounds from molecular structure. *Chem Res. Toxicol.* **12**, 670−678.
(7) Hall, L. H, Maynard E. L., and Kier, L. B. (1989) Structure−activity relationship studies on the toxicity of benzene derivatives: III. Predictions and extensions to new substituents. *Environ. Toxicol. Chem.* **8**, 431−436.
(8) Free, S. M., and Wilson, J. W. (1964) A mathematical contribution to structure−activity studies. *J. Med. Chem.* **7**, 395−399.
(9) Gao, C., Govind, R., and Tabak, H. H. (1992) Application of the group contribution method for predicting the toxicity of organic chemicals. *Environ. Toxicol. Chem.* **11**, 631−636.
(10) Kaiser, K. L. E., and Niculescu, S. P. (1999) Using probabilistic neural networks to model the toxicity of chemicals to the fathead minnow (*Pimephales Promelas*): A study based on 965 compounds. *Chemosphere* **38**, 3237−3245.
(11) Fredenslund, Aa., Jones, R. L., and Prausnitz, J. M. (1975) Group-contribution estimation of activity coefficients in nonideal liquid mixtures. *AIChE J.* **21**, 1086−1099.
(12) Verhaar, J. M., Leeuwen, C. J., and Hermens, J. L. M. (1992) Classifying environmental pollutants. 1: Structure−activity relationships for prediction of aquatic toxicity. *Chemosphere* **25**, 471−491.
(13) Russom, C. L., Bradbury, S. P., Borderius, S. J., Hammermeister, D. E., and Drummond, R. A. (1997) Predicting modes of toxic action from chemical structure: Acute toxicity in the fathead minnow (*Pimephales promelas*). *Environ. Toxicol. Chem.* **16**, 948−967.
(14) U.S. Environmental Protection Agency (2001) AQUIRE database http://www.epa.gov/ecotox/.
(15) Brooke, L. T., Call, D. J., Geiger, D. L., and Northcott, C. E., Eds. (1984) *Acute Toxicities of Organic Chemicals to Fathead Minnows (Pimephales promelas)*, Vol. 1, Center for Lake Superior Environmental Studies, University of Wisconsin, Superior, WI.
(16) Geiger, D. L., Northcott, C. E., Call, D. J., and Brooke, L. T., Eds. (1985) *Acute Toxicities of Organic Chemicals to Fathead Minnows (Pimephales promelas)*, Vol. 2, Center for Lake Superior Environmental Studies, University of Wisconsin, Superior, WI.
(17) Geiger, D. L., Poirier, S. H., Brooke, L. T., and Call, D. J., Eds. (1986) *Acute Toxicities of Organic Chemicals to Fathead Minnows (Pimephales promelas)*, Vol. 3, Center for Lake Superior Environmental Studies, University of Wisconsin, Superior, WI.
(18) Geiger, D. L., Call, D. J., and Brooke, L. T., Eds. (1988) *Acute Toxicities of Organic Chemicals to Fathead Minnows (Pimephales promelas)*, Vol. 4, Center for Lake Superior Environmental Studies, University of Wisconsin, Superior, WI.
(19) Geiger, D. L., Brooke, L. T., and Call, D. J., Eds. (1990) *Acute Toxicities of Organic Chemicals to Fathead Minnows (Pimephales promelas)*, Vol. 5, Center for Lake Superior Environmental Studies, University of Wisconsin, Superior, WI.
(20) Wasserman, P. D. (1989) *Neural Computing: Theory and Practice*, Van Nostrand Reinhold, New York.
(21) Hansch, C., and Leo, A. (1979) *Substituent Constants for Correlational Analysis in Chemistry and Biology*, John Wiley and Sons, New York.
(22) Veith, G. D., Lipnick, R. L., and Russom, C. L. The toxicity of acetylenic alcohols to the fathead minnow, *Pimephales promelas*: narcosis and proelectrophile activation. *Xenobiotica* **19**, 555−565.

TX0155045