

Prediction of IC₅₀ Values for ACAT Inhibitors from Molecular Structure

S. J. Patankar and P. C. Jurs*

Department of Chemistry, 152 Davey Laboratory, Penn State University, University Park, Pennsylvania 16802

Received September 16, 1999

A quantitative structure–activity study is performed on several series of compounds derived from *N*-chlorosulfonyl isocyanate to develop models that relate their structures to IC₅₀ activity for inhibition of acyl-CoA:cholesterol *O*-acyltransferase (ACAT). Numerical descriptors are used to encode topological, electronic, and geometric information from the molecular structures of the inhibitors. A data set of 157 compounds showing triglyceride- and cholesterol-lowering activity is used to develop successful linear regression models and nonlinear computational neural network models. The models are validated using an external prediction set.

INTRODUCTION

The reduction of triglyceride and cholesterol levels by inhibiting acyl-CoA:cholesterol *O*-acyltransferase (ACAT) has given rise to extensive efforts in identifying effective ACAT inhibitors.^{1–3} Several different classes of inhibitors have been reported in recent years.^{4–6} The focus has shifted from intestinal inhibition to inhibition in the liver and arterial wall due to clinical failures of a number of different compounds.^{7–9} Lipophilicity and ionization potential are important factors that contribute to activity after oral dosing.¹⁰

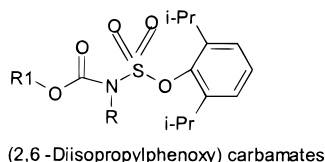
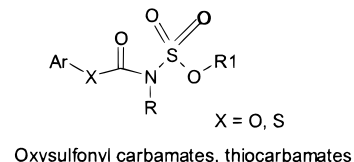
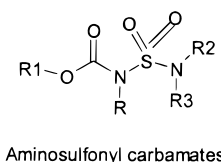
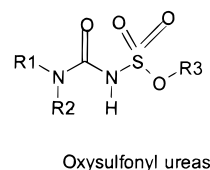
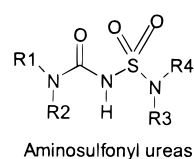
Many structure–activity relationships and in vitro determinations of IC₅₀ values based on biological assays have been reported for ACAT inhibitors.^{11–13} However, no work has been done using quantitative structure–activity relationship (QSAR) methods. It could be advantageous to predict IC₅₀ values rather than measure them with biological assays due to cost, time, safety, and sample availability. QSAR methods have been used in our group for prediction of many properties such as aqueous solubility,¹⁴ mammalian toxicity,^{15–17} boiling points,¹⁸ infinite dilution activity coefficient,¹⁹ and human intestinal absorption.²⁰ In this study molecular structure was used as a sole basis for prediction of IC₅₀ values of substituted *N*-carbonyl-functionalized ureas, carbamates, and thiocarbamates. Models were built using subsets of descriptors chosen by feature selection routines from a large pool of calculated molecular descriptors. Nonlinear computational neural network models (CNN)²¹ further improved upon the multiple linear regression models.

EXPERIMENTAL AND METHODOLOGY

This study was performed on a set of compounds reported by Picard et al.²² The in vitro IC₅₀ values were determined by measuring the incorporation of [1-¹⁴C]oleoyl-CoA into cholesterol esters in microsomes isolated from rat liver. The results were reported as the micromolar concentration of drug required to inhibit the enzymatic activity by 50% (IC₅₀).²³ There was no report of accurate estimation of error. For our study, to compress the range covered by the data, log(IC₅₀)

was used as the dependent variable. The experimental log(IC₅₀) values ranged from –1.00 to 2.21.

The data set of 167 compounds containing C, N, S, O, H, and halogens had a molecular formula range from 22 to 44 non-hydrogen atoms or 47–113 atoms total. The molecular weight varied from 314 to 635 Da. The data set included five classes of compounds with variations in substituents at positions R, R1, R2, R3, and R4, and in some classes at positions X and Ar. The following molecular frameworks illustrate the structural complexity of the data set.



Ten of the compounds which were ions or ionic salts were not used in the study due to software limitations. Thus, the data set consisted of 157 compounds. Twenty-seven of these compounds had IC₅₀ values reported as inequalities: 1 with IC₅₀ > 5, 1 with IC₅₀ > 10, 3 with IC₅₀ > 25, 18 with IC₅₀ > 50, and 4 with IC₅₀ > 100. These 27 compounds were held aside as an exclusion set leaving 130 compounds for development and testing of the QSAR models. All of the compounds, their experimental IC₅₀ values, and their observed and calculated log(IC₅₀) values are reported in Table 1.

* Corresponding author. E-mail: pcj@psu.edu.

Table 1. List of Compounds, Their IC₅₀ Values, and Observed and Calculated log IC₅₀ Values

Aminosulfonyl Ureas								
Com. No.	R	R 1	R 2	R 3	R 4	Obs. IC ₅₀	Obs. log IC ₅₀	Calc. log IC ₅₀
1	H		H			3.9	0.591	0.909
2	H		H	H		18	1.26	0.874
3 ^b	H		H			2.5	0.398	0.808
4	H		H	Ring at R3 & R4		15	1.18	1.11
5	H		H			2.3	0.362	0.502
6	H		H			2.1	0.322	0.383
7	H		H			2.7	0.431	0.161
8 ^c	H		H			1.1	0.0414	0.0537
9 ^b	H		H			1.2	0.0792	0.137
10	H		H	H		8.1	0.908	0.534
11	H		H	Ring at R3 & R4		17	1.23	1.16
12 ^c	H		H			0.24	-0.620	-0.298
13	H		H			0.14	-0.854	-0.676

Table 1 (Continued)

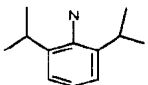



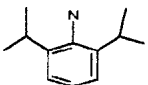


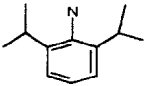
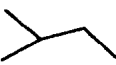
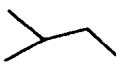
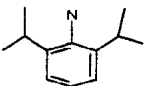
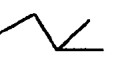

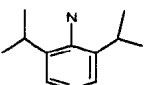


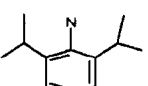


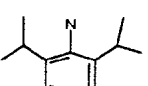


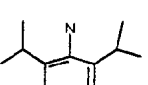

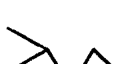
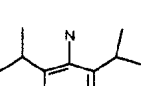


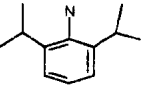


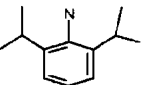
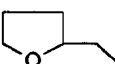
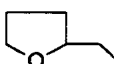
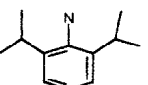
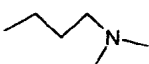
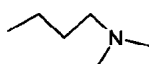
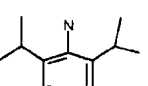

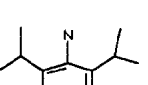


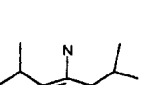

Com. No.	R	R 1	R 2	R 3	R 4	Obs. IC ₅₀	Obs. log IC ₅₀	Calc. log IC ₅₀
14	H		H			0.36	-0.444	-0.230
15 ^d			H			> 5	-	1.06
16	H		H			0.42	-0.377	-0.246
17	H		H			0.1	-1.00	-0.709
18	H		H			3.8	0.580	0.330
19	H		H			0.36	-0.444	-0.421
20	H		H			0.42	-0.377	-0.572
21 ^c	H		H			0.42	-0.377	-0.257
22	H		H			0.51	-0.292	-0.643
23	H		H			0.23	-0.638	-0.375
24	H		H			1.7	0.230	0.0271
25 ^b	H		H			55	1.74	2.00
26	H		H		CH ₃ - (CH ₂) ₁₃ -	0.33	-0.481	-0.0670
27 ^c	H		H			1.9	0.279	0.0565
28	H		H		CH ₃ - (CH ₂) ₁₇ -	1.9	0.279	0.382

Table 1 (Continued)

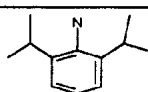
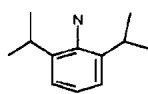
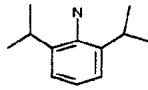
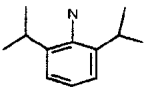
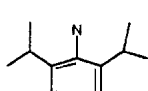
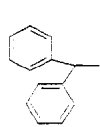
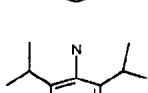
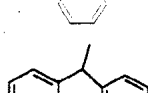
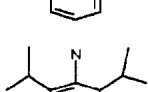
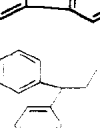
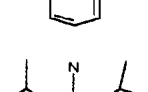
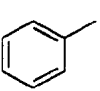
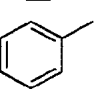
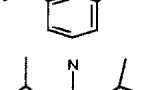
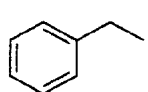
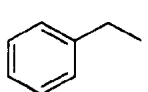
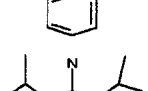

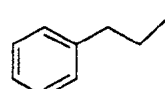
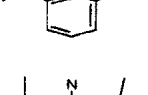

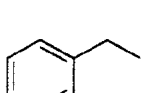
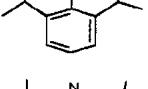
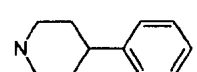
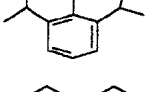
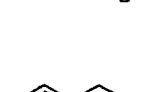
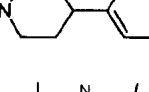

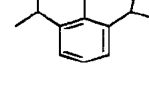
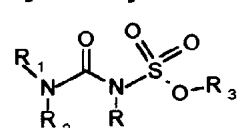
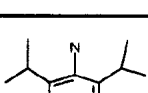

Com. No.	R	R 1	R 2	R 3	R 4	Obs. IC ₅₀	Obs. log IC ₅₀	Calc. log IC ₅₀
29	H		H	CH ₃ - (CH ₂) ₉ -	CH ₃ - (CH ₂) ₉ -	0.72	-0.143	-0.0135
30	H		H	CH ₃ - (CH ₂) ₁₁ -	CH ₃ - (CH ₂) ₁₁ -	69	1.84	1.79
31	H		H	H		11	1.04	0.913
32	H		H	H		4.1	0.613	1.04
33	H		H	H		25	1.40	1.43
34	H		H	H		5.9	0.771	0.883
35	H		H			16	1.20	1.36
36	H		H			2.2	0.342	0.440
37	H		H			1.7	0.230	0.655
38	H		H			0.4	-0.398	-0.322
39	H		H	Ring at R3 & R4		10.4	1.02	0.681
40	H			H		16	1.20	1.49
41	H	H		H		30	1.48	1.04
II. Oxsulfonyl Ureas								
								
Com. No.	R	R 1	R 2	R 3	Obs. IC ₅₀	Obs. log IC ₅₀	Calc. log IC ₅₀	
42	H		H		27	1.43	1.62	

Table 1 (Continued)

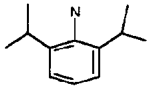

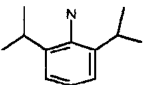
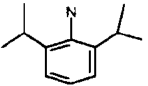

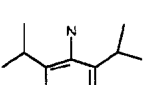
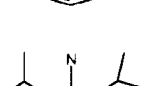
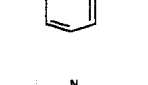

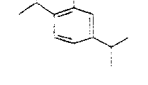
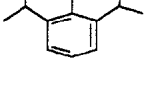
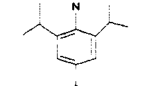
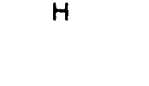
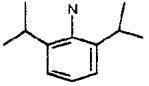
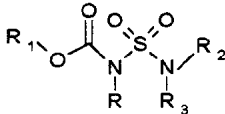
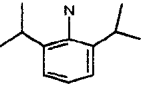

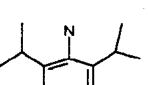

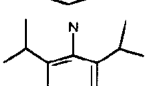

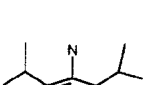
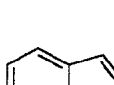
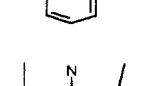


Com. No.	R	R 1	R 2	R 3	Obs. IC ₅₀	Obs. log IC ₅₀	Calc. log IC ₅₀
43	H		H		35	1.54	1.36
44 ^a	H		H	CH ₃ - (CH ₂) ₉ -	> 10	-	0.946
45	H		H		16	1.20	0.916
46	H		H	CH ₃ - (CH ₂) ₁₁ -	6.8	0.833	1.04
47	H		H	CH ₃ - (CH ₂) ₁₇ -	14	1.15	1.00
48 ^c	H		H		8.7	0.940	1.16
49	H		H		23	1.36	1.34
50	H				12	1.08	0.954
III. Aminosulfonyl Carbamates 							
Com. No.	R	R 1	R 2	R 3	Obs. IC ₅₀	Obs. log IC ₅₀	Calc. log IC ₅₀
51	H		Ring at R 2 R 3		53	1.72	1.31
52 ^d	H		Ring at R 2 R 3		> 100	-	1.12
53 ^d	H		Ring at R 2 R 3		> 100	-	1.23
54	H		Ring at R 2 R 3		43	1.63	1.30
55 ^b	H				15	1.18	0.858

Table 1 (Continued)

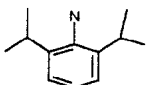

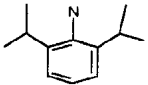


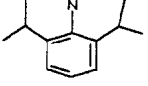


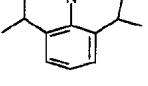


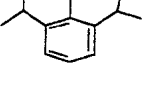


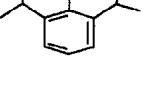


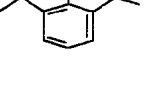


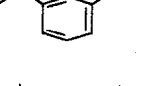

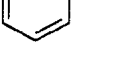
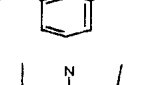
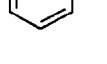
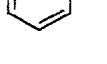
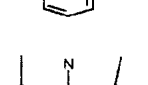
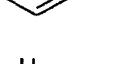
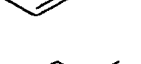
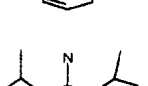
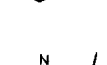
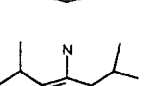
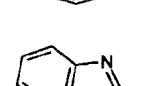
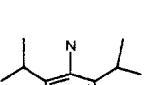
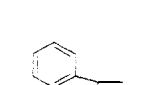
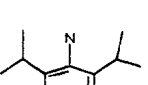



Com. No.	R	R 1	R 2	R 3	Obs. IC ₅₀	Obs. log IC ₅₀	Calc. log IC ₅₀
56 ^c	H		H		21	1.32	1.55
57	H				1.9	0.279	0.652
58	H				5.1	0.708	0.859
59	H				22	1.34	0.926
60	H				1.3	0.114	0.715
61	H				4.2	0.623	0.791
62	H				58	1.76	1.33
63 ^b	H				12	1.08	1.05
64 ^c	H				58	1.76	1.38
65	H				20	1.30	1.55
66	H		H		38.1	1.58	1.64
67	H		H		25.8	1.41	1.33
68	H		H		52	1.72	2.19
69	H		H		10.6	1.03	1.40
70	H		H		49	1.69	1.61

Table 1 (Continued)

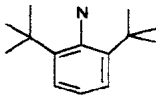
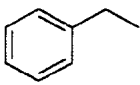
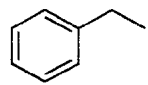
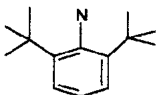
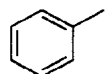
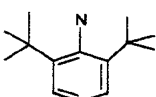
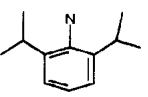
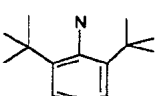
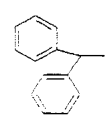
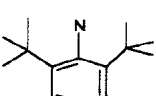
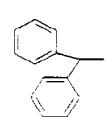
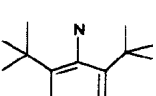
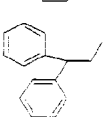
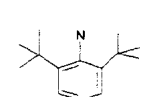


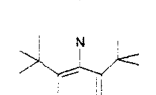

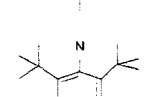


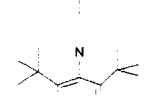
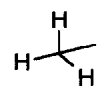

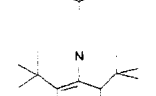


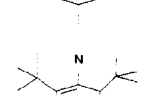


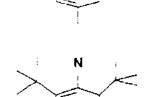


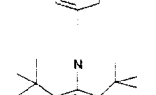
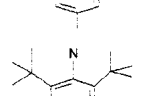
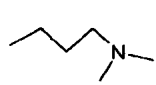
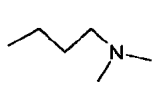
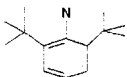
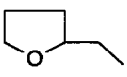
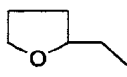
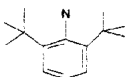

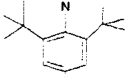
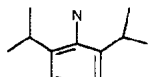
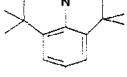

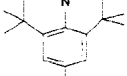
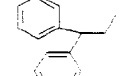
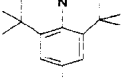
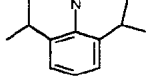
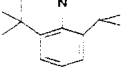
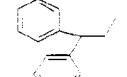
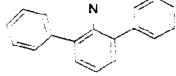
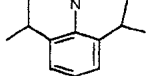
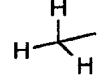
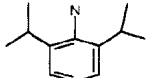
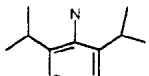
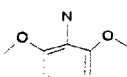
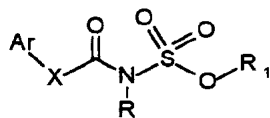
Com. No.	R	R 1	R 2	R 3	Obs. IC ₅₀	Obs. log IC ₅₀	Calc. log IC ₅₀
71	H				11	1.04	1.24
72	H		H		19	1.28	1.35
73	H		H		16.8	1.23	1.24
74	H		H		11.4	1.06	1.42
75	H		H		4.6	0.663	0.981
76	H		H		21.8	1.34	1.30
77 ^b	H				3.2	0.505	0.509
78	H		H		19.4	1.29	1.12
79 ^c	H				3.2	0.505	0.840
80	H				18	1.26	1.12
81	H				2.2	0.342	0.242
82	H				4.3	0.633	0.379
83	H				2.7	0.431	0.513
84	H		CH ₃ - (CH ₂) ₉ -	CH ₃ - (CH ₂) ₉ -	89	1.95	1.42
85	H				164	2.21	1.88

Table 1 (Continued)

Com. No.	R	R 1	R 2	R 3	Obs. IC ₅₀	Obs. log IC ₅₀	Calc. log IC ₅₀
86	H				21	1.32	0.958
87	H		H		23.4	1.37	1.44
88	H		H		13.5	1.13	1.26
89	H		H		11.5	1.06	1.14
90	H		H		24.4	1.39	1.19
91	H		H		15	1.18	1.41
92	H		H		82	1.91	1.58
93 ^d	H		H		> 25	-	2.04
94 ^d	H		H		> 100	-	1.70
95	H	CH ₃ - (CH ₂) ₁₁ -	H		33	1.52	1.08
96	H	CH ₃ - (CH ₂) ₁₁ -	H		30	1.48	1.49

IV. Oxysulfonyl O-Aryl Carbamates & S-Aryl Thiocarbamates



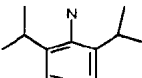

Com. No.	X	R	Ar	R1	Obs. IC ₅₀	Obs. log IC ₅₀	Calc. log IC ₅₀
97	O	H			40	1.60	2.08

Table 1 (Continued)

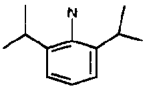
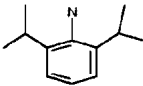
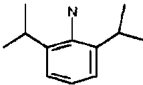

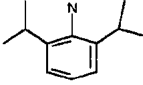
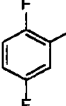
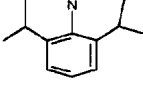

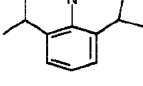

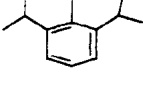

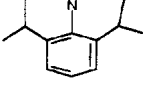

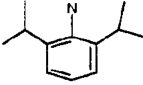
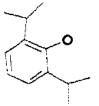

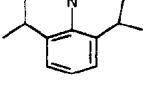
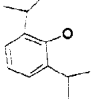
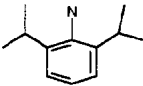
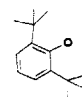
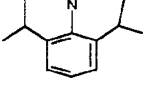

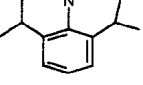
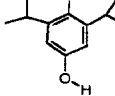
Com. No.	R	R 1	R 2	R 3	Obs. IC ₅₀	Obs. log IC ₅₀	Calc. log IC ₅₀
98 ^d	O	H		CH ₃ - (CH ₂) ₁₁ -	13	1.11	1.19
99	O	H		CH ₃ - (CH ₂) ₁₅ -	10.3	1.01	1.19
100 ^d	O	H			> 50	-	1.94
101 ^d	O	H			> 50	-	2.17
102	O	H			38	1.58	1.77
103 ^d	O	H			> 50	-	1.27
104 ^d	O	H			> 50	-	1.21
105 ^d	O	H			> 50	-	1.23
106	O	H			9.4	0.973	1.19
107		H			50	1.70	1.67
108	O	H			50	1.70	1.19
109	O	H			29	1.46	1.20
110 ^d	O	H			> 50	-	1.17

Table 1 (Continued)

Com. No.	R	R 1	R 2	R 3	Obs. IC ₅₀	Obs. log IC ₅₀	Calc. log IC ₅₀
111	O	H			95	1.98	2.25
112	O	H		CH ₃ - (CH ₂) ₁₁ -	52	1.72	1.43
113	O	H			90	1.95	2.26
114 ^c	O	H			7.7	0.886	1.19
115 ^b	O	H			14	1.15	1.20
116 ^c	O	H			22	1.34	1.22
117		H			12	1.08	1.19
118	O	H			30	1.48	1.21
119	S	H		CH ₃ - (CH ₂) ₁₁ -	48	1.68	1.32
120	S	H			8.7	0.940	1.19
Com. No.	R	R 1			Obs. IC ₅₀	Obs. log IC ₅₀	Calc. log IC ₅₀
121 ^d	H				> 100	-	1.91
122 ^d	H				> 50	-	1.91
123	H				26	1.41	1.60

Table 1 (Continued)


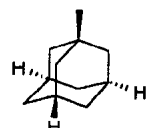
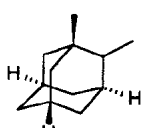
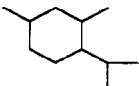


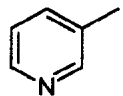
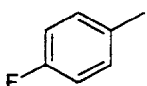
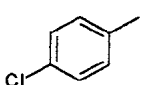
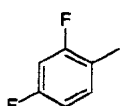
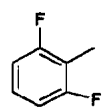
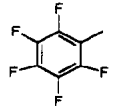
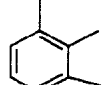
Com. No.	R	R 1	Obs. IC ₅₀	Obs. log IC ₅₀	Calc. log IC ₅₀
124 ^b	H	CH ₃ - (CH ₂) ₁₁ -	13	1.11	1.26
125	H		38	1.58	1.63
126	H		13	0.756	1.19
127	H		23	1.36	1.23
128	H		21	1.32	1.18
129	H		5.7	1.11	1.22
130	H		3.6	0.556	1.18
131 ^d	H		> 50	-	1.42
132 ^d	H		> 50	-	2.23
133	H		97	1.99	1.74
134 ^d	H		> 50	-	1.94
135	H		65	1.81	1.98
136	H		66	1.82	1.73
137 ^d	H		> 50	-	1.33

Table 1 (Continued)

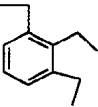
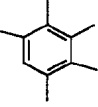
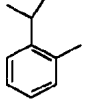
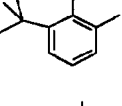
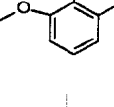
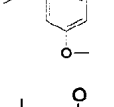
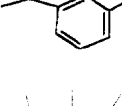
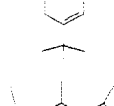
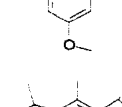
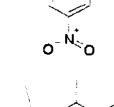
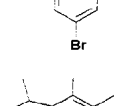
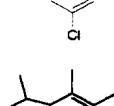
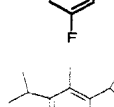
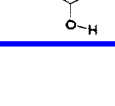
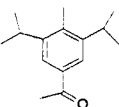
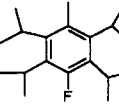
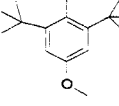
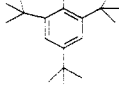
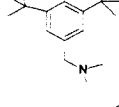
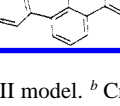
Com. No.	R	R 1	Obs. IC ₅₀	Obs. log IC ₅₀	Calc. log IC ₅₀
138	H		20	1.30	1.40
139 ^d	H		> 50	-	1.33
140 ^d	H		> 50	-	1.19
141	H		49	1.69	1.21
142 ^d	H		> 50	-	1.23
143 ^d	H		> 25	-	1.22
144 ^d	H		> 50	-	1.19
145	H		8.8	0.944	1.22
146 ^d	H		> 25	-	1.20
147 ^d	H		> 50	-	1.79
148 ^c	H		31	1.49	1.21
149 ^b	H		42	1.62	1.21
150 ^d	H		> 50	-	1.21
151 ^d	H		> 50	-	1.18

Table 1 (Continued)

Com. No.	R	R 1	Obs. IC ₅₀	Obs. log IC ₅₀	Calc. log IC ₅₀
152 ^c	H		41.8	1.62	1.22
153	H		13	1.11	1.25
154 ^b	H		4.8	1.62	1.22
155	H		26	1.41	1.20
156 ^c	H		22	0.681	1.20
157	H		12	1.08	1.41

^a Calculated log IC₅₀ values are from type III model. ^b Cross-validation set compounds. ^c Prediction set compounds. ^d Exclusion set compounds.

A subset of 13 compounds was chosen randomly as a prediction set (pset), and this set of compounds was used for external validation of all the models that were developed in the study. It was checked for representation of all classes of compounds in the data set. For the development of type I linear regression models, the training set (tset) included all the remaining 117 compounds. For the generation of type II and type III nonlinear CNN models, the training set was further subdivided into a 106-compound training set and an 11-compound cross-validation set (cvset). The cross-validation set was used to avoid overtraining during the development of neural network models. The cvset, pset, and exclusion set compounds are identified with footnotes in Table 1.

The compounds were sketched as two-dimensional representations using HyperChem.²⁴ Initial modeling was performed to generate starting three-dimensional conformations. The three-dimensional conformations were further refined to their lowest energy states using MOPAC,²⁵ a semiempirical molecular modeling routine, using the PM3 Hamiltonian.²⁶ These optimum three-dimensional conformations were used for generation of descriptors dependent on geometry. All computations were done on a DEC 3000AXP Model 500 workstation at Penn State University.

The Automated Data Analysis and Pattern Recognition Toolkit (ADAPT) software package^{27,28} was used to calculate molecular structure descriptors to encode the geometric, topological, and electronic features of the compounds. Geometric descriptors included solvent-accessible surface area,²⁹ molecular volume,²⁹ molecular polarizability,³⁰ and moments of inertia.³¹ Topological descriptors^{32–35} included counts of atom types, bond types, numbers of basis rings,

and functional groups, as well as molecular connectivity indices to represent size and degree of branching. Electronic descriptors³⁶ stored the partial atomic charge descriptors, such as the most positive or most negative atoms, HOMO and LUMO energies, and dipole moments. Descriptors that combined geometric as well as electronic information such as hydrogen bonding and charged partial surface area were also included.^{37,38} As the pharmacophore had the dissociable acidic hydrogens and heteroatoms, intramolecular as well as intermolecular hydrogen bonding was expected. Therefore, hydrogen bonding descriptors were generated for pure and for solvated states.

In addition a set of six descriptors was calculated using information from MOPAC runs using the MNDO Hamiltonian. These descriptors were developed by Lowery et al.³⁹ from semiempirical molecular orbital calculations, and they include a term represented by van der Waals volume, a volume-independent molecular polarizability term, covalent hydrogen bonding acidity and basicity, and electrostatic hydrogen bonding acidity and basicity. In this work five descriptors were calculated and the steric term was represented by the volume descriptor generated by other ADAPT descriptor development routines.

A set of eight electrotopological state (E-state) descriptors were calculated as they measure the reactivity of each atom. They encoded the information regarding intermolecular attractions. These descriptors were developed by Kier and Hall.⁴⁰ For each atom *i* there is an intrinsic state *I_i*

$$I_i = (Dv_i + 1)/D_i \quad (1)$$

where *Dv_i* is the number of electrons on an atom *i* involved

in all types of bonds and lone pair of electrons and D_i is the number of σ electrons. Then

$$S_i = I_i + DI_i \quad (2)$$

where DI_i is the sum over all atoms j , where i is not equal to j and

$$DI_i = (I_i - I_j) \quad (3)$$

and let the denominator be r_{ij}^2 where r_{ij} is the shortest path length from atom i to j .

A new set of topological descriptors developed by Cao⁴¹ was calculated and was found to be useful in this study. The molecular distance edge (MDE) vector consisted of the descriptor values for 10 distance edge terms between four different types of carbon atoms. For example, D_{12} was the distance edge term between C_1 and C_2 carbons. The four types of carbons were classified as C_1 primary carbon, C_2 secondary carbon, C_3 tertiary carbon, and C_4 quaternary carbon. The distance-edge d_{ij} was defined as

$$d_{ij} = \prod_{k < l}^{n_{ij}} (d_{ik,jl})^{1/2n_{ij}} \quad (i = 1, 2, 3, 4; j \geq i) \quad (4)$$

where $d_{ik,jl}$ was the number of bonds between carbons C_i^k and C_j^l with k and l denoting the carbon identity and i and j denoting the classification types described above. The carbon atoms are arbitrarily assigned. In the above equation n_{ij} was the number of times a particular interaction occurred in the compound of interest. The members of the MDE vector were then defined as

$$D_{ij} = (n_{ij}/d_{ij}^2) \quad (i = 1, 2, 3, 4; j \geq i) \quad (5)$$

and represented 10 new descriptor values that were added to the descriptor pool. The new MDE descriptors did not correlate highly with other topological descriptors.

Initially, a set of 233 descriptors was calculated for each compound. All the descriptors were subjected to objective feature selection (selection not using the dependent variables) to remove those that did not contribute useful information to the pool. Pairwise correlations between descriptors were examined, and only one descriptor was retained from a pair contributing similar information ($r \geq 0.93$). Descriptors with greater than 90% identical values were removed since they were not encoding the structural differences between compounds that accounts for their IC₅₀ values.

The reduced pool of information-rich descriptors was then examined to find subsets of descriptors that accurately represented the relationship between molecular structure and the IC₅₀ values. Simulated annealing^{42,43} and genetic algorithm routines,⁴⁴ along with an interactive regression routine, were used to generate statistically valid linear models, termed type I models. Genetic algorithm and simulated annealing routines each investigated a large number of descriptor subsets. The quality of each model was based on root-mean-square (rms) error between calculated and observed log(IC₅₀) as well as statistical integrity. A variety of subset sizes were investigated to determine the optimum number of descriptors to be used. Subset size was considered to be optimum when addition of one more descriptor did not improve the statistics of the model.

The set of descriptors chosen for the type I linear model was subsequently submitted to a CNN to develop a nonlinear type II model. A fully connected, feed-forward, three-layer CNN was used. The input layer had the same number of descriptors as the type I model. The number of neurons in the hidden layer was considered to be optimized when addition of another neuron did not decrease the training set rms error significantly. A single neuron in the output layer provided the predicted log(IC₅₀) values. A linear transformation of the descriptor values restricted them to the interval between [0, 1], and they were then used as the input for the network. A quasi-Newton BFGS (Broyden–Fletcher–Goldfrab–Shanno)^{45–49} algorithm²¹ was used to train the networks. A cross-validation set was used to prevent overtraining. The cross-validation set was a small subset (about 10%) of compounds, drawn randomly from the training set. It was not used during training but was used for periodic testing during training. Training was stopped when the cross-validation set error was minimized, as beyond this point the CNN was fitting information specific to individual compounds rather than general information of the entire data set relating to the property of interest.

The CNN results were significantly dependent on starting weights and biases, which were chosen randomly. A variety of starting points were sampled by two methods. A generalized simulated annealing algorithm was used to find optimum starting weights and biases. An automated CNN program selected a user-determined number of starting weights and biases. The results from these large number of trials were reviewed to select a good starting point for weights and biases.

Finally, fully nonlinear CNN type III models were developed using a genetic algorithm descriptor selection routine with a CNN for evaluating the fitness of each subset of descriptors selected. The quality of the models was based on the following fitness function:

$$\text{quality} = \text{TSET} + 0.4|\text{TSET} - \text{CVSET}| \quad (6)$$

where TSET and CVSET refer to the tset and cvset rms errors.⁴³ Models chosen with this quality factor performed better than models chosen with just training set rms error as the quality factor. Inclusion of the rms error of the cross-validation set in the quality factor gave models with good external predictive ability. The coefficient of 0.4 in the quality factor was determined empirically. Unlike type I model development, investigation of a large number of type III models was extremely computationally intensive. The architecture from type II models was initially retained in neural networks; however, a 8–3–1 architecture yielded significant improvements. Type III models were found that had lower rms errors than the corresponding type II models. This proved that the descriptors chosen on the basis of linear criteria were not the best subset that could be found for use with a fully nonlinear CNN model.

All the models that were developed in this study were validated with the same external prediction set. The models were able to accurately predict the log(IC₅₀) values of these compounds that were set aside from the start. The prediction set rms errors were of the same magnitude as that of the training set and cross-validation set errors. The statistical validity of type I models was also determined by using

Table 2. Nine Descriptors Used for Multiple Linear Regression Type I and Nonlinear CNN Type II Model for Prediction of $\log(\text{IC}_{50})$ Values of ACAT Inhibitors

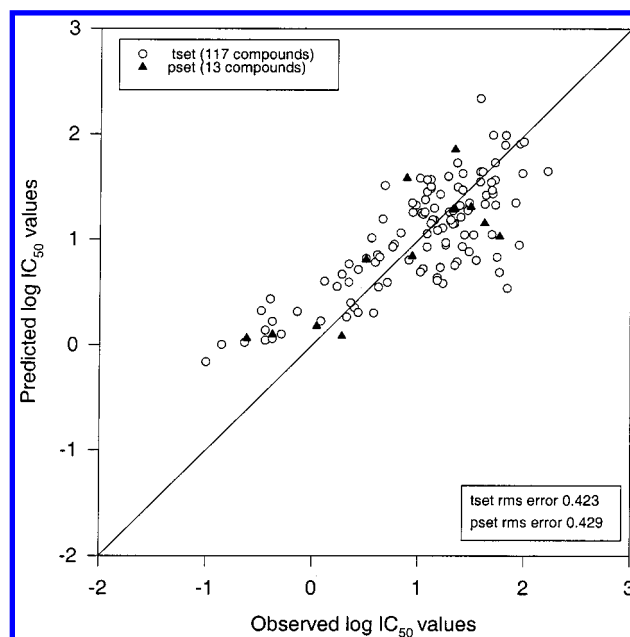
descriptor	coeff ^a	error est ^a	range	explanation ^b
RNCG1	0.751	0.149	0.097–0.16	rel chg of most electroneg atom
KAPA6	1.17	0.150	4.93–19.3	κ^3 index
NBR17	0.664	0.102	1–4	no. of basis rings
WTPT5	0.691	0.138	2.64–14.7	Σ path wts starting from N atoms
MCHG0	–0.723	0.109	–9.87–1.10	max chg diff between donor & acceptor atoms
ACHG0	0.714	0.125	0–1.09	avg chg diff between donor & acceptor atoms
MDE14	0.483	0.096	5.43–41.4	distance edge between 1° and 4° carbons
EMIN1	–0.816	0.138	–2.48–1.29	min E-state index value
ESUM1	–0.620	0.135	56.6–106	Σ E-state values
Intercept	–0.0586			

^a Values representing model I. ^b RNCG1, relative negative charge calculated from charge of most negative atom/sum total negative charge;³⁷ KAPA6, corrected κ^3 index based on paths of length three;³² NBR17, number of basis rings; WTPT5, sum all path weights starting from nitrogen atoms;³⁵ MCHG0, maximum charge difference between donor and acceptor atom;³⁸ ACHG0 average charge difference between donor and acceptor atoms;³⁸ MDE14, molecular distance edge between all primary and quaternary carbons;⁴¹ EMIN1, minimum atomic E-state value;⁴⁰ ESUM1, sum of E-state values over all heavy atoms.⁴⁰

standard statistical parameters such as the correlation coefficient, the overall *F*-value of the model, and the *T*-values of the individual descriptors. The presence of outliers was detected by looking at regression diagnostics that measured the effect of individual data points on the model, such as standardized and studentized residuals, leverage, and Cook's distance.⁵⁰ Compounds were often flagged as outliers if they were not well represented in the data set. An additional method of validation for all three model types was the visual inspection of calculated versus observed plots of $\log(\text{IC}_{50})$.

RESULTS AND DISCUSSION

Multiple linear regression^{42,44} was used to find the best type I linear model for predicting $\log(\text{IC}_{50})$, and it consisted of the nine structural descriptors shown in Table 2. The model contained three topological descriptors (KAPA6, NBR17, and WTPT5), two electrotopological (E-state) state descriptors (EMIN1, ESUM1), one charged partial surface area descriptor (RNCG1), one molecular distance edge descriptor (MDE14), and two hydrogen bonding descriptors (MCHG0, ACHG0). Connectivity was encoded with the κ^3 index (KAPA6), the number of basis rings in the molecules (NBR17), the sum of all path weights starting from nitrogen (WTPT5), and molecular distance edge between all primary and quaternary carbons (MDE14). The relative negative charge calculated from the ratio of the charge of the most negative atom to the sum of the total negative charge (RNCG1) represented information about the charge and area. The maximum charge difference between donor hydrogen and acceptor atom (MCHG0) and average charge difference between donor hydrogen and bonded heteroatom pairs (ACHG0) encoded information about hydrogen bonding ability. The minimum atomic E-state value (EMIN1) and the sum of all E-state values over all heavy atoms (ESUM1) encoded information about atom-based values that are an extension of the free valance concept extended to heteroatoms. Among the nine descriptors, pairwise correlations ranged from 0.354 to 0.807 with an average value of 0.682. The combination of these nine descriptors was able to accurately encode the structural differences and pharmacophore similarities that influence the IC_{50} values. The descriptors were not overly concentrated only in one area, which could have led to a poor quality model. It was not surprising to see the number of basis rings among them as

**Figure 1.** Calculated/predicted vs observed $\log(\text{IC}_{50})$ for type I linear model.

many of the compounds included varied rings as part of their structure. The hydrogen bonding descriptors were expected to be among this subset as all compounds had several heteroatoms, thus having large number of donor atoms leading to the hydrogen bonding interaction.

The type I model for $\log(\text{IC}_{50})$ had a tset root-mean-square (rms) error of 0.423 log units with a multiple correlation coefficient of 0.779. After diagnostic testing, two compounds (numbers 25 and 30) were flagged as outliers, but their removal did not significantly affect the errors so they were left in for the final error calculations. These were the only two compounds in their subclass with high experimental IC_{50} values. In validation of this model, the external prediction set had a rms error of 0.429 log units with a multiple correlation coefficient of 0.828. Figure 1 is a plot of calculated/predicted vs observed $\log(\text{IC}_{50})$ values for all 130 compounds.

The descriptors chosen by the linear regression feature selection routine were then submitted to a CNN to develop a nonlinear model. The type II model resulted in tset, cvset, and pset rms errors of 0.362, 0.334, and 0.336 log units,

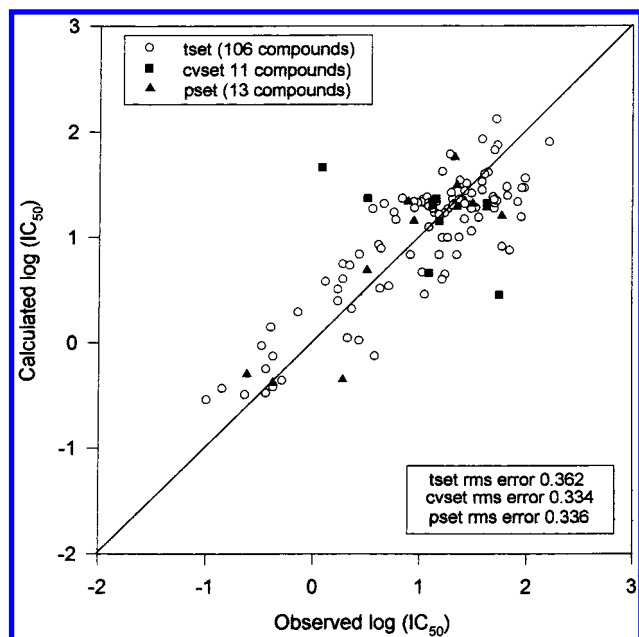


Figure 2. Calculated/predicted vs observed log(IC₅₀) for type II nonlinear model with 9–2–1 architecture.

respectively. These values represented a 14% improvement over the type I tset error, and a 22% improvement over the type I pset error. The CNN had a 9–2–1 architecture, consisting of nine input neurons corresponding to the descriptors chosen by the linear feature selection, two hidden neurons, and one output neuron corresponding to calculated log(IC₅₀). The number of hidden neurons was optimized by adding neurons until there was no further improvement in the cost function of neural networks. The 9–2–1 architecture of the CNN had 23 adjustable parameters for 117 training set compounds, well above the cutoff of 2.0 for the ratio of observations to adjustable parameters. Figure 2 shows a plot of calculated/predicted vs observed log(IC₅₀) values for this type II nonlinear CNN model.

The compounds were then submitted to the combined genetic algorithm/CNN feature selection routine⁴⁴ to search directly for a nonlinear relationship between the descriptors and log(IC₅₀). It was believed that such a routine would find a subset of descriptors that were more suited to take advantage of the nonlinear nature of CNN modeling. Several different computations were done in search of the best architecture. A high-quality CNN model with 8–3–1 architecture was found. The number of adjustable parameters of 31 was still well within the cutoff. The tset rms error for the type III model was 0.269 log units, an improvement of 26% over the type II CNN mode; the cvset rms error was 0.271 log units, an improvement of 19%; and the pset rms error was 0.265 log units, an improvement of 21%. Figure 3 shows a plot of calculated/predicted vs observed log(IC₅₀) values for the type III model. The improvement in going from type I to type II to type III models can easily be seen by comparing Figures 1, 2, and 3. The corresponding mean absolute errors for the type III model are tset 0.226 log units, cvset 0.208 log units, and pset 0.242 log units. This type III neural network model was the best found in the course of this work.

The descriptors chosen for the type III CNN model are listed in Table 3. The nonlinear feature selection routine

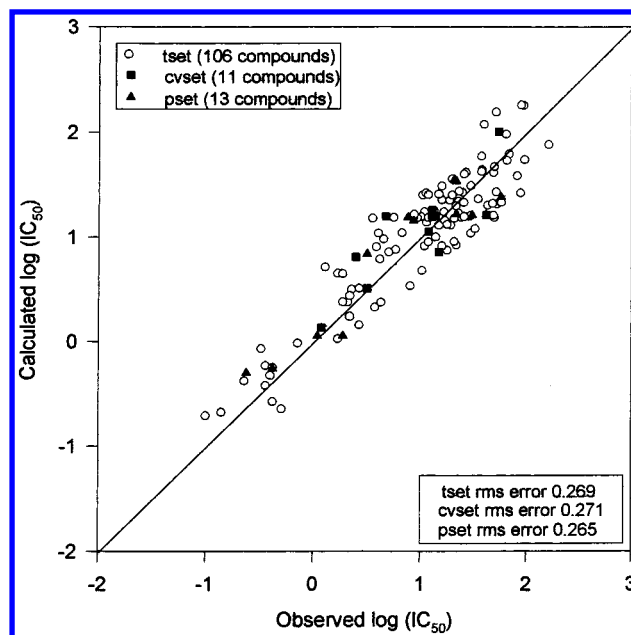


Figure 3. Calculated/predicted vs observed log(IC₅₀) for type III fully nonlinear model with 8–3–1 architecture.

Table 3. Eight Descriptors Used for a Nonlinear CNN Type III Model for Prediction of log(IC₅₀) Values of ACAT Inhibitors

descriptor	range	explanation ^a
SA	542–5186	surface area
WTPT 2	1.90–2.02	molecular ID/no. of atoms
WTPT 5	2.64–14.7	Σ path weights starting from N atoms
3SP3	0–6	no. of sp ³ carbons bonding to 3 other carbons
MCHG0	0.580–1.10	max chg diff between donor & acceptor atoms
ACHG0	0.580–1.09	avg chg diff between donor & acceptor atoms
SAAA2	1.03–9.96	Σ surface area of donatable H/no. of acceptor atoms
ESUM1	56.6–106	Σ E-state over all heavy atoms

^a SA, solvent accessible surface area;²⁹ WTPT2, molecular ID/number of atoms;³⁵ WTPT5, sum of all path weights starting from nitrogen atoms;³⁵ 3SP3, number of sp³ carbons bonded to three other carbon atoms; MCHG0, maximum charge difference between donor and acceptor atoms;³⁸ ACHG0, average charge difference between donor and acceptor atoms;³⁸ SAAA2, sum of surface area of donatable hydrogen atoms/number of acceptor atoms;³⁸ ESUM1, sum of E-state values over all heavy atoms.⁴⁰

chose eight descriptors, four of which were identical with descriptors chosen in the linear model. The eight descriptors consisted of three topological (WTPT2, WTPT5, 3SP3), one geometric (SA), three hydrogen bonding (MCHG0, ACHG0, SAAA2), and one electrotopological state descriptor (ESUM1). The solvent accessible surface area (SA) encoded the information about the shapes and sizes of the molecules. The sum of all path weights starting from nitrogen atoms (WTPT5), the structurally significant molecular identification numbers/number of atoms in the molecule (WTPT2), and number of tertiary sp³ carbons in a given molecule (3SP3) related the connectivity of different groups. The maximum charge difference between donor and acceptor atoms (MCHG0), the average charge difference between the donor and acceptor atoms (ACHG0), and the ratio of sum of surface area of donatable hydrogen atoms to number of acceptor atoms (SAAA2) encoded the information about hydrogen

bonding abilities of different structures. The sum of E-state indices for a molecule (ESUM1) encoded information about atom-based E-state values that are an extension of free valance concept extended to heteroatoms. As in the type I model, the descriptors had a uniform distribution over all categories. The nonlinear relationship among these descriptors was able to more accurately predict $\log(\text{IC}_{50})$ than the linear model had done.

Even though all the experimental IC_{50} values were reported on the basis of in vitro studies, the actual assay method involved microsomal assays using rabbits or rats as tissue donors.^{51,52} As there were no reported errors in the literature for this data set, a true comparison of errors between experiment and model could not be made.

One of the problems encountered while developing QSAR models is the possibility of models being found due to chance. It has been shown that performing feature selection on a pool of independent random variables can lead to linear correlation with the dependent variable if the number of independent variables in the pool is much larger than the total number of observations.⁵³ Obviously, if the independent variables are random, there can be no real meaning attached to such a correlation. To ensure that chance effects did not influence the current study, a randomized test was performed. The dependent variables of all of the compounds in the test and cvset were scrambled randomly, and the genetic algorithm feature selection routine was run again. The cost function of 0.462 was obtained from this run, compared to a value of 0.265 for the model based on real data. Thus, the cost function for the scrambled data is about 1.7 times the value for the model built with real data. This and other randomized studies not reported here clearly indicate that the probability of chance correlations being in the results of the current study is extremely low.

Once the models were finalized, the $\log(\text{IC}_{50})$ values for the 27 exclusion set compounds were predicted on the basis of the best model available, the type III nonlinear CNN model. The predicted $\log(\text{IC}_{50})$ values are reported in Table 1. As the compounds in this exclusion set had observed IC_{50} values that were reported as inequalities, it was not possible to compare these predictions to observed values. It should be pointed out that 22 of the 27 exclusion set compounds had observed $\text{IC}_{50} > 50$ (82%), whereas the remainder of the data set of compounds had only 15 compounds out of 130 (12%) with observed IC_{50} values greater than 50. This mismatch between the main data set characteristics and the exclusion set characteristics casts substantial doubt on these predictions. Nonetheless, they are the best estimated of the $\log(\text{IC}_{50})$ values based on the models developed from the present set of compounds.

CONCLUSIONS

Models have been developed using QSAR methodology that successfully relate molecular structure to inhibitory concentrations of ACAT inhibitors. The models reported here are the first with the ability to predict IC_{50} values from structural information alone for ACAT inhibitors derived from *N*-chlorosulfonyl isocyanate. The structural information of cholesterol- and triglyceride-lowering compounds is numerically encoded as molecular descriptors including hydrogen bonding descriptors and electrotopological state

index descriptors. Linear feature selection led to the development of linear type I models and improved nonlinear type II CNN models. Nonlinear feature selection led to development of fully nonlinear type III CNN models. The type III model had the best rms errors of all the models that were generated.

This study has shown that IC_{50} values can be predicted on the basis of molecular structure alone, without the inclusion of experimentally derived data such as partition coefficients. The models that have been developed with linear regression and computational neural network techniques can be applied to prediction of inhibitory activities of compounds that are not present in the data set used in this study as long as their structures are similar to those used in the study. The predictive power of these models could be useful in cases where biological assays are necessary to calculate the activities, which tend to be costly, time-consuming, or impossible to measure experimentally due to biological variation.

REFERENCES AND NOTES

- (1) Sliskovic, D. R.; Trivedi, B. K. ACAT Inhibitors: Potential Anti-atherosclerotic Agents. *Curr. Med. Chem.* **1994**, *1*, 204–225.
- (2) Picard, J. A. Patent Update: ACAT inhibitors. *Curr. Opin. Ther. Patents* **1993**, *3*, 151–160.
- (3) Matsuda, K. ACAT inhibitors as Anti-atherosclerotic agents: Compounds and Mechanisms. *Med. Res. Rev.* **1994**, *14*, 271–305.
- (4) Maduskuie, T. P.; Wilde, R. G.; Billheimer, J. T.; Cromley, D. A.; Design, Synthesis, and Structure–Activity Relationship Studies for a new Imidazole Series of J774 Macrophage Specific Acyl CoA-Cholesterol Acyl Transferase (ACAT) Inhibitors. *J. Med. Chem.* **1995**, *38*, 1067–1083.
- (5) Kumazawa, T.; Harakawa, H.; Fukui, H.; Shirakura, S. *Bioorg. Med. Chem. Lett.* **1995**, *5*, 1829–1832.
- (6) Fancelli, D.; Chiari, A.; Cozzi, P.; Lovisolio, P. Imidazolylbenzopyrane Derivatives A New Class of AcylCoA cholesterol Acyl Transferase (ACAT) Inhibitors. *J. Enzyme Inhibition* **1994**, *8*, 159–172.
- (7) Hainer, J. W.; Terry, J. G.; Connel, J. M.; Zyruk, H.; Jenkins, R. M.; Shad, D. L.; Gillies, P. J.; Livak, K. J.; Hunt, T. L.; Crouse, J. R., III. Effect of Acyl-CoA:Cholesterol Acyl transferase Inhibitor DuP 128 on Cholesterol absorption and Serum Cholesterol in Humans. *Clin. Pharmacol. Ther.* **1994**, *56*, 65–74.
- (8) Harris, W. S.; Dujovne, C. A.; von Bergman, K.; Neil, J.; Akester, J.; Windsor, S. L.; Greene, D.; Look, Z. Effects of the ACAT Inhibitor CL-277082 on Cholesterol Metabolism in Humans. *Clin. Pharmacol. Ther.* **1990**, *48*, 189–194.
- (9) Peck, R. W.; Wiggs, R.; Posner, J. The Tolerability, Pharmacokinetics and Effect on Plasma Cholesterol of 447C88, an ACAT Inhibitor with Low Bioavailability, in Healthy Volunteers. *Arteriosclerosis* **1994**, *109*, 155–156.
- (10) Austel, V.; Kutter, E. Absorption, Distribution, and Metabolism of Drugs. In *Quantitative Structure–Activity Relationships of Drugs*; Topliss, J. G., Ed.; Academic Press: New York, 1983; pp 437–496.
- (11) O'Brien, P. M.; Sliskovic, D. R.; Blankley, J.; Roth, B. D.; Wilson, M. W.; Hamelhele, K. L.; Krause, B. R.; Stanfield, R. Inhibitors of Acyl-CoA:Cholesterol O-Acyl Transferase (ACAT) as Hypocholesterolemic Agents. 8. Incorporation of Amide or Amine Functionalities into a Series of Disubstituted Ureas and Carbamates. Effects on ACAT Inhibition in Vitro and Efficacy in Vivo. *J. Med. Chem.* **1994**, *37*, 1810–1822.
- (12) Augelli-Szafran, C. E.; Blankley, C. J.; Roth, B. D.; Trivedi, B. K.; Bousley, R. F.; Essenberg, A. D.; Hamelhele, K. L.; Krause, B. R.; Stanfield, R. L. Inhibitors of Acyl-CoA:Cholesterol Acyl-transferase. 5. Identification and Structure–Activity Relationships of Novel β -Ketoamides as Hypocholesterolemic Agents. *J. Med. Chem.* **1993**, *36*, 2943–2949.
- (13) Trivedi, B. K.; Holmes, A.; Stober, T. L.; Blankley, J.; Roark, W. H.; Picard, J. A.; Shaw, M. K.; Essenburg, A. D.; Stanfield, R. L.; Krause, B. R. Inhibitors of Acyl-CoA:Cholesterol Acyltransferase. 4. A Novel Series of Urea ACAT Inhibitors as Potential Hypocholesterolemic Agents. *J. Med. Chem.* **1993**, *36*, 3300–3307.
- (14) Sutter, J. M.; Jurs, P. C. Prediction of Aqueous Solubility for a Diverse Set of Heteroatom-Containing Organic Compounds Using a Quantitative Structure–Property Relationship. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 100–107.

- (15) Johnson, S. R.; Jurs, P. C. Prediction of Acute Mammalian Toxicity from Molecular Structure for a Diverse Set of Substituted Anilines Using Regression Analysis and Computational Neural Networks. In *Computer-Assisted Lead Finding and Optimization*; van de Waterbeemd, H., Testa, B., Folkers, G., Eds.; Verlag Helvetica Chimica Acta: Basel, 1997.
- (16) Eldred, D. V.; Jurs, P. C. Prediction of Acute Mammalian Toxicity of Organo-phosphorous Pesticide Compounds from Molecular Structure. *SAR QSAR Environ. Res.* **1999**, *10*, 75–79.
- (17) Eldred, D. V.; Jurs, P. C. Prediction of Fathead Minnow Acute Toxicity of Organic Compounds from Molecular Structure. *Chem. Res. Toxicol.* **1999**, *12*, 670–678.
- (18) Wessel, M. D.; Jurs, P. C. Prediction of Normal Boiling Points for a Diverse Set of Industrially Important Organic Compounds *J. Chem. Inf. Comput. Sci.* **1995**, *5*, 841–850.
- (19) Mitchell, B. E.; Jurs, P. C. Development of QSPR Models to Predict the Infinite Dilution Activity Coefficients of Organic Compounds in Aqueous Solutions From Molecular Structure *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 489–496.
- (20) Wessel, M. D.; Jurs, P. C.; Tolan, J. W.; Muskal, S. M. Prediction of Human Intestinal Absorption of Drug Compounds from Molecular Structure. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 726–735.
- (21) Xu, L.; Ball, J. W.; Dixon, S. L.; Jurs, P. C. Quantitative Structure–activity Relationships for Toxicity of Phenols Using Regression Analysis and Computational Neural Networks. *Environ. Toxicol. Chem.* **1994**, *13*, 841–851.
- (22) Picard, J. A.; O'Brien, P. M.; Sliskovic, D. R.; Anderson, M. K.; Bousley, R. F.; Hamelehle, K. L.; Krause, B. R.; Stanfield, R. L. Inhibitors of Acyl-CoA:Cholesterol O-Acyltransferase. 17. Structure–Activity Relationships of Several Series of Compounds Derived from N-Chlorosulfonyl Isocyanate. *J. Med. Chem.* **1996**, *39*, 1243–1252.
- (23) Krause, B. R.; Black, A.; Bousley, R.; Essenburg, A.; Cornicelli, J.; Holmes, A.; Homan, R.; Kieft, K.; Sekerke, C.; Shaw-Hes, M. K.; Stanfield, R.; Trivedi, B.; Woolfe, T. Divergent Pharmacologic Activities of PD 132301–2 and CL 277,082 Urea Inhibitors of Acyl-CoA:Cholesterol Acyltransferase (ACAT). *J. Pharmacol. Exp. Ther.* **1993**, *267*, 734–743.
- (24) Hypercube Inc., Waterloo, ON.
- (25) Stewart, J. P. P. *MOPAC 6.0*; Quantum Chemistry Program Exchange, Indiana University, Bloomington, IN, Program 455.
- (26) Stewart, J. P. P. *MOPAC—A Semiempirical Molecular-Orbital Package*. *J. Comput.-Aided Mol. Des.* **1990**, *4*, 1–45.
- (27) Stuper, A. J.; Brugger, W. E.; Jurs, P. C. *Computer-Assisted Studies of Chemical Structure and Biological Function*; Wiley-Interscience: New York, 1979.
- (28) Jurs, P. C.; Chou, J. T.; Yuan, M. In *Computer-Assisted Drug Design*; Olsen, E. C., Christoffersen, R. E., Eds.; American Chemical Society: Washington, DC, 1979; pp 103–129.
- (29) Pearlman, R. S. In *Physical Chemical Properties of Drugs*; Yalkowsky, S. H., Sinkula, A. A., Valvani, S. C. Eds.; Marcel Dekker: New York, 1980.
- (30) Miller, K. J.; Savchik, J. A. A New Empirical Method to Calculate Average Molecular Polarizabilities. *J. Am. Chem. Soc.* **1979**, *101*, 7206.
- (31) Goldstein, H. *Classical Mechanics*; Addison-Wesley: Reading, MA, 1950; pp 144–156.
- (32) Kier, L. B. A Shape Index of Orders One and Three from Molecular Graphs. *Quant. Struct.-Act. Relat. Pharmacol., Chem. Biol.* **1986**, *5*, 1–7.
- (33) Kier, L. B.; Hall, L. H. *Molecular Connectivity in Structure-Activity Analysis*; John Wiley & Sons Inc.: New York, 1986.
- (34) Balaban, A. T.; Highly Discriminating Distance-Based Topological Index. *Chem. Phys. Lett.* **1982**, *89*, 399.
- (35) Randic, M. On Molecular Identification Numbers. *J. Chem. Inf. Comput. Sci.* **1984**, *24*, 164–175.
- (36) Dixon, S. L.; Jurs, P. C. Atomic Charge Calculations for Quantitative Structure–Property Relationships. *J. Comput. Chem.* **1992**, *13*, 492–504.
- (37) Stanton, D. T.; Jurs, P. C. Development and Use of Charged Partial Surface Area Structural Descriptors in Computer-Assisted Quantitative Structure Property Relation Studies. *Anal. Chem.* **1990**, *62*, 2323–2329.
- (38) Vinogradov, S. N.; Linnell, R. H. *Hydrogen Bonding*, Van Nostrand Reinhold: New York, 1971.
- (39) Lowrey, A. H.; Cramer, C. J.; Urban, J. J.; Famini, G. R.; Quantum Chemical Descriptors for Linear Solvation Energy Relationships. *Comput. Chem.* **1995**, *19*, 209–215.
- (40) Kier, L. B.; Hall, L. H.; The E-State as an Extended Free Valance. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 548–552.
- (41) Cao, C.; Distance-Edge Topological Index—Research on Structure–Property Relationships of Alkenes. *Huaxue Tongbao* **1996**, *22*, 1238–1244.
- (42) Sutter, J. M.; Jurs, P. C. Selection of Molecular Structure Descriptors for Quantitative Structure–Activity Relationships. In *Adaption of Simulated Annealing to Chemical Problems*; Kalivas, J. H., Ed.; Elsevier Science Publishers B. V.: Amsterdam, 1995.
- (43) Sutter, J. M.; Dixon, S. L.; Jurs, P. C. Automated Descriptor Selection for Quantitative Structure–Activity Relationship Using Generalized Simulated Annealing. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 77–84.
- (44) Wessel, M. D. Computer-Assisted Development of Quantitative Structure–Property Relationships and Design of Feature Selection Routines. Ph.D. Dissertation, Pennsylvania State University, University Park, PA, 1996.
- (45) Broyden, C. G. The Convergence of a Class of Double-rank Minimization Algorithms *J. Inst. Math. Appl.* **1970**, *6*, 76.
- (46) Fletcher, R. A New Approach to Variable Metric Algorithms. *Comput. J.* **1970**, *13*, 317.
- (47) Goldfarb, D. A Family of Variable-Metric Methods Derived by Variational Means. *Math. Comput.* **1970**, *24*, 23.
- (48) Shanno, D. F. Conditioning of Quasi-Newton Methods for Function Minimization. *Math. Comput.* **1970**, *24*, 647.
- (49) Fletcher, R. *Practical Methods of Optimization*; Vol. 1, Unconstrained Optimization; Wiley: New York, 1980.
- (50) Neter, J.; Wasserman, W.; Kutner, M. H. *Applied Linear Statistical Models*; Irwin: Homewood, IL, 1990.
- (51) Krause, B. R.; Black, A.; Bousley, R.; Essenburg, A.; Cornicelli, J.; Holmes, A.; Homan, R.; Kieft, K.; Sekerke, C.; Shaw-Hes, M.; Stanfield, R.; Trivedi, B.; Woolf, T. Divergent Pharmacologic Activities of PD 132301–2 and CL 277,082, Urea Inhibitors of Acyl-CoA:Cholesterol Acyltransferase (ACAT). *J. Pharmacol. Exp. Ther.* **1993**, *267*, 734–743.
- (52) Roth, B. D.; Blankley, C. J.; Hoefle, M.; Holmes, A.; Rorak, W. H.; Trivedi, B. K.; Essenburg, A. D.; Kieft, K. A.; Krause, B. R.; Stanfield, R. L. Inhibitors of acylCoA:cholesterol acyl-transferase. 1. Identification and structure activity relationships of a novel series of fatty acid anilide hypocholesterolemic agent. *J. Med. Chem.* **1992**, *35*, 1609–1617.
- (53) Hansch, C. Quantitative Structure–Activity Relationship and the Unnamed Science. *Acc. Chem. Res.* **1993**, *2*, 147–153.

CI990125R