Use of Automatic Relevance Determination in QSAR Studies Using Bayesian Neural Networks

Frank R. Burden*

School of Chemistry, Monash University, Victoria 3800, Australia

Martyn G. Ford and David C. Whitley

Centre for Molecular Design, Institute of Biomedical and Biomolecular Sciences, Portsmouth University, Portsmouth PO1 2DY, U.K.

David A. Winkler

CSIRO Division of Molecular Science, Private Bag 10, Clayton South MDC, Clayton, Victoria 3169, Australia

Received March 21, 2000

We describe the use of Bayesian regularized artificial neural networks (BRANNs) coupled with automatic relevance determination (ARD) in the development of quantitative structure−activity relationship (QSAR) models. These BRANN-ARD networks have the potential to solve a number of problems which arise in QSAR modeling such as the following: choice of model; robustness of model; choice of validation set; size of validation effort; and optimization of network architecture. The ARD method ensures that irrelevant or highly correlated indices used in the modeling are neglected as well as showing which are the most important variables in modeling the activity data. The application of the methods to QSAR of compounds active at the benzodiazepine and muscarinic receptors as well as some toxicological data of the effect of substituted benzenes on *Tetetrahymena pyriformis* is illustrated.

1. INTRODUCTION

Quantitative structure−activity relationships (QSARs) have been applied successfully to many drug and agrochemical design and optimization problems.[1] Finding structure−activity relationships is essentially a pattern recognition process, and, historically, QSAR models have been developed using linear methods such as multiple linear regression (MLR) and partial least squares (PLS). Recently, regression methods based on neural networks have shown that they can account for nonlinear SARs and can deal with the linear dependencies that sometimes appear in real SAR problems.[2] Traditional feed-forward, back-propagation neural networks still present some problems, however, including overtraining, overfitting, optimization of the network architecture, and selection of the best QSAR model. Overtraining arises from running the neural network training for too long and results in a loss of ability of the trained network to generalize. Overtraining can be avoided by using an "early-stopping" procedure in conjunction with a validation set. However where the training data set is large and diverse, as may occur in combinatorial discovery, this can result in prohibitively large training times. Overfitting results from the use of too many adjustable parameters to fit the training data and can be avoided by using principal component analysis (PCA) to reduce the number of input variables. To date, there has been no objective method for architecture optimization, and trial-and-error, or rule-of thumb, is generally used to find the network with best performance. A validation procedure such as "leave-N-out" produce a family of models, and it is not clear which should be used. In addition, the validation effort scales as the square of the number of compounds, making validation very tedious for very large data sets.

Our research focuses on the development of novel molecular descriptors and the improvement of the variable selection, structure−activity mapping, and validation processes in QSAR.[3−5] The aim is to develop tools applicable to the development of robust QSAR models from large or fuzzy data sets. In a previous paper[6] we showed how the application of a regularizer based on Bayesian statistics can be added to a standard back-propagation neural network to overcome the problems of overtraining and overfitting in QSAR modeling. These Bayesian regularized neural networks (BRANNs) do not require a validation set during network training, allowing all of the data to be used to develop a single, robust QSAR model. BRANNs have found applicability in numerous other areas of data modeling but, with the exception of our recent work, have not been used for QSAR analysis.

This paper reports how the application of a method known as automatic relevance determination (ARD) can be used in conjunction with a Bayesian neural network to determine objective variable relevance and develop optimal QSAR models. Only a subset of the many molecular descriptors that could be used contains meaningful and statistically significant SAR information. A number of methods for variable reduction have been reported in the literature. The simplest include stepwise regression methods, and the use of principal

* To whom correspondence should be addressed. Telephone: +613 9905 4559. Fax: +613 9905 4597. E-mail: frank.burden@sci.monash.edu.au.

10.1021/ci000450a CCC: $19.00
Published on Web 10/28/2000

components analysis to find a new, smaller set of composite descriptors which are more relevant to the QSAR model. These methods, being essentially linear, suffer from the disadvantage that they may not be effective where the relationship between descriptors and activity is nonlinear.

Recently a number of novel methods for variable selection have been reported. For example, Hasegawa and Funatsu[7] used a combined genetic algorithm/partial least squares method (GAPLS) to select the best combination of descriptors to form a model. However, Jouan-Rimbaud et al.[8] showed that even if this is carefully done, the procedure could still select irrelevant variables. Waller and Bradley[9] refined this evolutionary approach in their FRED (fast random elimination of descriptors) algorithm. Nath and co-workers[10] used a saliency measure for variable selection based on partitioning of neural network weights. They showed that it performed quite well in comparison with a stepwise variable selection rule using synthetic data. Kovalishyn et al.[11] also used a method based on the weights of a cascade correlation neural network to prune input variables. Todeschini and coworkers[12] used a Kohonen neural network to select variables for PLS calibration models for analytical studies. Zheng and Tropsha[13] developed a useful variable selection method based on the use on *K*-nearest neighbor principles. Tetko et al.[14] have compared the efficacy of five pruning algorithms for variable selection. These methods have several disadvantages: many are nonoptimal, and they may cause chance correlations when many variables, or combination of variables, are screened for inclusion in the model.

We have employed ARD to ensure that irrelevant variables are excluded from the model and that the most important variables are highlighted. This feature helps to overcome the criticism of neural networks that they do not provide easily interpretable models. In ARD the network weights are divided into several classes and input weights are assigned individual decay rates in the Bayesian regularizer. Inputs with large decay rates have small weights and low relevances. Depending on the type of molecular representations (descriptors) used, we propose that BRANN-ARD will be useful for deriving robust QSAR models for specific biological activities to obtain a detailed understanding of drug or toxicity mechanisms, and screening large virtual databases for leads (virtual screening).

2. METHODS

2.1. Molecular Indices. We employed a combination of five easily computed molecular indices in this work. These include the well-studied Randic[15] index *R*; the valence modification to the Randic index by Kier and Hall (*K*);[16] and an atomistic (*A*)[17] index developed by Burden, which has now been enhanced by the recognition of aromatic atoms and hydrogen atom donors and acceptors (*B*). The *R* and *K* indices are produced from the path length and valence electron counts in the molecule. The enhanced atomistic indices (*A*, *B*) count the numbers of each type of these atoms present in the molecule. Two further indices have been added: the first counts the number of rings of various sizes (*G*), and the second counts some common functional groups (*F*). The indices and their associations are enumerated in Table 1.

The six types of index, *R*, *K*, *A*, *B*, *F*, and *G* are complementary, and we have shown in previous studies[18] that the combination RKA yields better QSAR models than the individual indices. In previous studies, principal component analysis (PCA) was used to reduce redundant information and minimize overfitting. Since PCA is a linear transformation, the number of principal components that give the lowest standard error of prediction was used as the measure for selection of the model with the best predictivity.

Table 1. Molecular Indices Used in the QSAR Analyses

index	element	no. of connections	atom type		
atomistic				rings	ring size
A01	mol mass			G03	3
A02	H	1	H1	G04	4
A03	C	2	C2 (sp)	G05	5
A04	C	3	C3 (sp_2)	G06	6
A05	C	4	C4 (sp_3)	G07	7
A06	N	1	N1	G08	8
A07	N	2	N2		
A08	N	3	N3	fragments	
A09	N	4	N4	F01	H−O−C
A10	O	1	O1	F02	H−O−N
A11	O	2	O2	F03	C−O−C
A12	F	1	F1	F04	N−O−C
A13	Si	2	Si2	F05	N−O−N
A14	Si	3	Si3	F06	C═O
A15	Si	4	Si4	F07	O═C−N
A16	P	2	P2	F08	O═C−O
A17	P	3	P3	F09	N═O
A18	P	4	P4	F10	O═N═O
A19	P	5	P5		
A20	S	1	S1	Randic[15]	
A21	S	2	S2	R01	$^0\chi$
A22	S	3	S3	R02	$^1\chi$
A23	S	4	S4	R03	$^2\chi$
A24	Cl	1	Cl1	R04	$^3\chi$
A25	Br	1	Br1	R05	$^4\chi$
A26	I	1	I1		
extended				Kier and Hall[16]	
B01	C(Ar)		c	K01	$0\chi^v$
B02	N(Ar)		n	K02	$1\chi^v$
B03	O(Ar)		o	K03	$2\chi^v$
B04	S(Ar)		s	K04	$3\chi^v$
B09	H donor		(N)H	K05	$4\chi^v$
B10	H donor		(O)H		
B11	H acceptor		N═(O)		

2.2. Bayesian Regularized Artificial Neural Networks. The Bayesian framework for neural networks is based on a probabilistic interpretation of network training.[19−21] The networks considered here are multilayer perceptrons with a single hidden layer and a single output. Such a network determines a nonlinear transformation from a vector of inputs (molecular indices or descriptors) to the output (biological activity) parametrized by a set of network weights. In contrast to conventional network training where an optimal set of weights is chosen by minimizing a suitable error function, the Bayesian approach considers a probability distribution of network weights. Initially this is described by a prior distribution representing general beliefs about the network weights before any data are observed (that the weights are normally distributed, for instance). Having observed a set of data, the posterior weight distribution is found by Bayesian inference, and the most probable set of weights occurs at the maximum of the posterior distribution. If a Gaussian prior distribution that penalizes large network weights is used, and the data are assumed to be generated by a smooth function with additive Gaussian noise, then maximizing the posterior distribution is equivalent to minimizing the standard sum-of-squares error together with a weight decay regularizer,[19]

as is frequently done in conventional network training. (In QSAR neural network studies early-stopping procedures are often used in place of the regularizer to avoid overfitting. Bishop,[19] p 343, indicates why these two methods are likely to produce similar results.)

The Bayesian approach offers more than conventional network training, however, in that the trained network is represented by a posterior distribution of weights, rather than a single set of weights. An input vector combined with the posterior weight distribution generates a distribution of network outputs. The mean μ variance σ^2 of a Gaussian approximation to this predictive distribution can then be calculated to provide $\pm\sigma$ bars on the mean prediction μ. This is a feature that is difficult to achieve with other network training methods.

In this scenario the prior distribution depends on a hyperparameter α representing the weight decay regularization, while a second hyperparameter β governs the variance of the noise. These hyperparameters may be determined simultaneously with the maximum of the posterior distribution by Mackay's evidence procedure.[20]

2.3. Automatic Relevance Determination. The implementation of BRANNs outlined above assumes a single rate of weight decay α for all the network weights, but the scaling properties of networks suggest that weights in different network layers should employ different regularization coefficients. By separating the weights into different classes, MacKay and Neal[20,21] developed a method for soft network pruning called automatic relevance determination. In ARD the weights are divided into one class for each input (containing all the weights from that input to the hidden layer), one class for the hidden layer biases, and one class for each output (containing all the weights from the hidden layer to that output). Inputs with large decay rates have small weights, so, in problems with many input variables, some of which may be irrelevant to the prediction of the output, ARD allows the network to "estimate" the importance of each input, effectively turning off those which are not relevant. This allows all variables, including those that have little impact on the output, to be included in the analysis without ill-effect, as irrelevant variables will have their weights reduced automatically. On the other hand, in problems with very large numbers of inputs, it may be more efficient to remove variables with large decay rates and train a new network on a reduced set of inputs, especially if the trained network is to be used to screen a very large virtual database.

ARD has two main advantages over other pruning methods: it is firmly based on probability theory, and it is carried out automatically.

2.4. Characteristics of the Neural Network. The BRANN-ARD networks used in this study are three layer, fully connected, feed-forward, back-propagation neural networks to which has been added a Bayesian regularizer. As in previous work[6] on the BRANN method, four nodes using a sigmoidal transfer function in the hidden layer were used together with a linear transfer function for the output layer. It was shown that the number of nodes in the hidden layer needs only to be large enough to produce the effective number of parameters γ and that extra nodes are superfluous. In this work, the calculations were performed 10 or more times to overcome false starts caused by the random choice of initial weights from normal prior distributions and to allow the conjugate gradient routine to find the evidence maximum. In previous work the minimum standard error of prediction (SEP) of the test set was used. The present procedure is more likely to find the evidence maximum and thereby provide an objective criterion for the best model. This is more important with the BRANN-ARD method since the Bayesian regularization is being applied independently to the weights associated with each input variable and bias as well as those associated with the output layer. Because of the larger number of parameters, convergence of the calculations is less reliable and demands more iterative steps. We report the calculation which gives the maximum evidence for the hyperparameters α and β.

The work was encoded in the MATLAB[22] programming language; some routines from the NETLAB[23] toolbox were used for the BRANN-ARD modeling. These modules are incorporated in a comprehensive chemometrics package written in MATLAB by F.R.B.

2.5. Structure−Activity Data Sets. We used three well-behaved literature data sets, one of which (BZD) has become a de facto standard for QSAR methods development, and which we have used previously for this purpose. This gives us an objective basis for comparing the performance of new QSAR methods with more traditional methods. The benzodiazepine data set (BZD) is a set of 245 compounds that act on the benzodiazepine receptor and was culled from the literature.[24−31] They do not have a common substructure so that the nature of the molecular indices used in forming the model become more important. The muscarinic data set (MUS) is a set of 162 compounds that act on the M_1 muscarinic receptor and was culled from the literature.[32−37] Muscarinic compounds are used in the treatment of memory related problems such as Alzheimer's disease. The compounds in this data set do not have a common substructure but do fall into small subsets with common structures. IC_{50} values were measured as the concentration necessary to displace 50% of [^{3}H]quinuclidinyl benzilate (QNB) from the M_1 muscarinic receptor. The toxicological data set (TOX) has been modeled recently[38] by the BRANN method. It consists of 277 substituted benzenes and their toxicity to the ciliate *Tetrahymena pyriformis* as measured by Schultz and modeled in parts by Cronin et al.[39−41] using MLR techniques. A test set of 15% of compounds was used in the analysis of all three data sets.

2.6. Procedure Used in Forming the Model. As described in a previous paper[6] on the use of BRANNs in data modeling, the following steps were taken. The data set was divided into training and test sets (15%) by a *K*-means clustering algorithm clustering on descriptors (*X*) and biological activity (*Y*) values taken together. Clustering on *X* and *Y* data together, rather than just on *X*, is our preferred method in that it clusters compounds according to all of the given information in a manner akin to PLS. This may lead to different test sets for different groups of indices but is appropriate when searching for the best model to represent a data set. The training set data was standardized and the test set data was transformed with these means and standard deviations.

We also compared the models produced by BRANN-ARDs with BRANN calculations without ARD, multiple linear regression (MLR), multiple linear regression with

Table 2. Statistics of QSAR Models Derived from the Three Data Sets Using the RKABGF Indices

	data set	method	variables	SEF[b]	R^2	SEP[b]	Q^2	comments
1	BZD[c]	MLR	39	0.18	0.47	0.20	0.32	
2	BZD	MLRBE	33	0.18	0.47	0.20	0.32	$p > 0.9$, A03, A05, A11, A24, B04, G07 excluded
3	BZD	MLRFI	10	0.18	0.44	0.18	0.38	$p < 0.1$, R04, K03, K04, A02, A05, A25, A26, F07, F08, F10 included
4	BZD	ANN	22[d]	0.13	0.73	0.14	0.66	
5	BZD	BRANN	39	0.12	0.75	0.12	0.71	
6	BZD	ARD	39	0.12	0.76	0.13	0.70	
7	BZD	ARD	34	0.12	0.75	0.14	0.67	A26, B10, F01, F04, F06 removed
8	BZD	ARD	29	0.12	0.74	0.15	0.58	A03, A25, A26, B10, G 04, G07, F01, F04, F06, F07 removed
9	MUS[c]	MLR	29	0.12	0.56	0.19	0.34	
10	MUS	MLRBE	25	0.13	0.53	0.17	0.28	$p > 0.7$, K04, K05, A05, G06 excluded
11	MUS	ANN	20[d]	0.10	0.72	0.12	0.47	
12	MUS	BRANN	29	0.13	0.57	0.13	0.39	
13	MUS	ARD	29	0.12	0.56	0.13	0.39	
14	MUS	ARD	24	0.12	0.60	0.14	0.42	R02, K03, A08, A11, G06 removed
15	MUS	ARD	19	0.15	0.54	0.19	0.455	R02, K03, A04, A08, A11, A24, A25, B03, G06, F04 removed
16	TOX[c]	MLR	32	0.09	0.69	0.11	0.71	
17	TOX	MLRBE	28	0.09	0.69	0.11	0.71	$p > 0.7$, R05, K03, A08, A11 excluded
18	TOX	ANN	14[d]	0.07	0.85	0.10	0.67	
19	TOX	BRANN	32	0.06	0.86	0.08	0.81	
20	TOX	ARD	32	0.07	0.80	0.09	0.80	
21	TOX	ARD	27	0.07	0.84	0.09	0.80	A04, A06, A12, A26, B01 removed
22	TOX	ARD	22	0.07	0.83	0.09	0.78	K04, A04, A05, A25, A06, A12, A26, B01, F02, G06 removed

[a] BZD variables: R01-K05, A01, A02, A04-A08 A10-A12, A21, A24, A25, A26, B01, B02, B04, B10, G04-G07, F01, F03, F04, F06, F07, F08, F10. MUS variables: R01-K05, A01, A02, A04, A05, A08, A10, A11, A21, A24, A25, B01, B02, B03, B04, G05, G06, F03, F04, F08. TOX variables: R01-K05, A01, A02, A04, A05 A06, A08, A10-, A12, A24-A26, B01, B02, B10, G06, F01-F03, F07, F08, F10. [b] SEF = standard error of fit; SEP = standard error of prediction [c] BZD set, 245 samples; MUS set, 162 samples; TOX set, 277 samples. [d] Number of principal components used to train the networking giving the lowest SEP value.

backward elimination (MLRBE), and multiple linear regression with forward inclusion (MLRFI). The MLRBE eliminations and MLRFI inclusions were carried out using hypothesis testing where various p values were used to eliminate or include numbers of variables. The MLRBE and MLRFI calculations were done for the best characterized data set, BZD, to assess their performance relative to the neural net methods. Calculations using a regular back-propagation artificial neural network (ANN) were also carried out using four hidden nodes and an early-stopping validation set of 15% of the samples. This reduced the training set to 70% of the total samples given that a test set of 15% was also used. The ANN giving the best test set predictions derived from 10 repeat training sessions are shown in Table 2. Since the test set predictions were used to select the model, the test set is no longer independent, and the results for the ANN in Table 2 overestimate the model's predictive ability. Strictly speaking a further, independent test set should be used for comparison purposes. This illustrates one of the difficulties of deriving QSAR models using conventional ANNs.

3. RESULTS AND DISCUSSION

Table 2 shows statistics for QSAR models obtained by analysis of the three data sets using MLR, MLRBE, and MLRFI, and ANN as well as the BRANN and BRANN-ARD methods. The quality of the results depends on the reliability of the data, the diversity of the molecular structure of the samples, the relevance of the indices, and the power of the model in capturing the underlying structure−activity relationships. The three data sets show different characteristics: the benzodiazepine set (BZD) has a moderate diversity of structures and reliable data; the smaller muscarinic set (MUS) has a wider diversity of structures and less reliable data; and the toxicity set (TOX) has very low diversity of structures and highly reliable data. The benzodiazepine set shows the main advantages of the BRANN-ARD method with the two other sets, showing similar results with less emphasis.

Table 2 shows that the ANN, BRANN, and BRANN-ARD methods perform similarly and are all superior to the MLR calculations for the three data sets. In the BZD calculation with 39 input variables and 4 hidden nodes the number of weights is 165 for a non−Bayesian back-propagation neural network, whereas the effective number of parameters was found to be much smaller at 76 for the BRANN calculation and 55 for the BRANN-ARD calculation. The ratio of data set size to weights or number of effective parameters (ρ) is 1.5 (neural net), 3.2 (BRANN), and 4.5 (BRANN-ARD), showing that overfitting is very unlikely in the BRANN and BRANN-ARD models.

Figure 1shows the relative magnitudes of the coefficients of all 39 (normalized) descriptors from the MLR calculation on the BZD data set in entry 1 of Table 2. Figure 1 shows that this linear fit ascribes the four largest coefficients to the variables F03 (ether linkage), A11 (oxygen atoms with two connections), the topological index K01, and F04 (the N−O−C moiety). Entries 2 and 3 of Table 2 show the results of MLR calculations using both MLRFI and MLRBE. In these cases the number of relevant descriptors has been reduced by including only the most important (MLRFI), or eliminating the least important (MLRBE) successively using hypothesis testing criteria. Clearly, linear dependence between descriptors is a major determinant of those to be included, or excluded, from the MLR model. For example, the MLRBE calculation removes the A11 index which is highly correlated with the F03 and F04 indices. The results show that an MLR model can be obtained with far fewer variables than the 39 originally selected.

The results of modeling the BZD data set using a standard back-propagation neural network are summarized in entry 4

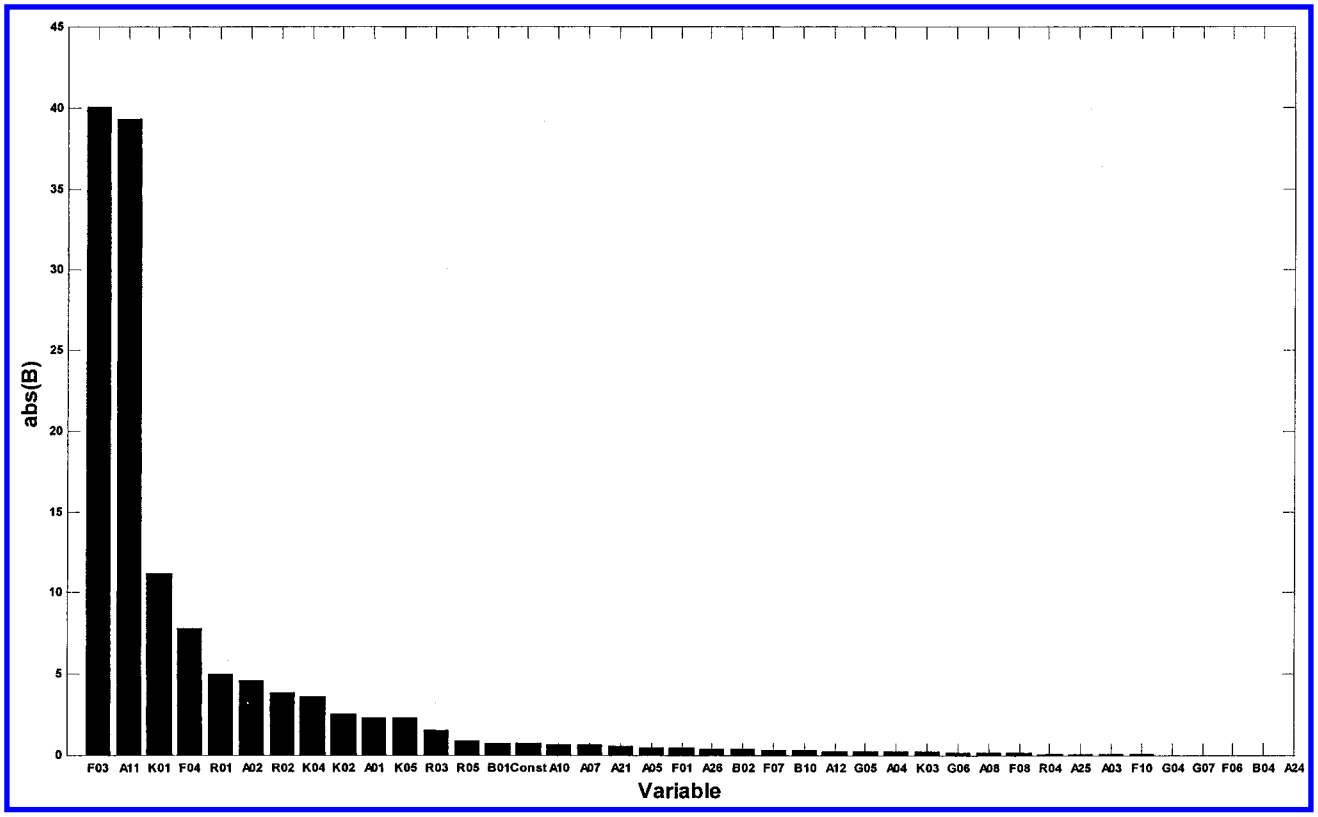

Figure 1. MLR coefficients for the BZD data.

of Table 2. The model which uses 22 principal components derived from the 39 descriptors and giving the best SEP was selected. Clearly this gives superior fit and prediction statistics compared with the MLR models, suggesting that nonlinearity in the response surface, and variable dependencies are important in this SAR.

The BRANN calculation (entry 5) gives slightly better results using all 39 descriptors and choosing the final model as that giving the maximum evidence. The BRANN-ARD calculation (entry 6) has essentially identical results using only those descriptors which contain relevant information. Figure 2 illustrates the relevances ascribed by the ARD procedure to the descriptors used to model the BZD data set. Note that the relevances are plotted on a log scale, with long bars denoting the most relevant descriptors, and short bars the least relevant. Alternative calculations, leading to very similar fits, may slightly reorder the relevances. The relevances are inversely proportional to the weights, which can overemphasize the relevance scale. We have arbitrarily selected log relevances of >10 as the most relevant to focus the discussion. Figure 2 shows that the six least relevant indices are the number of hydrogen atoms (A02), the number of hydrogen bond donors (B10), the numbers of iodine atoms (A26), the number of O═C−N groups (F07), the number of OH groups (F01), and the number of N−O−C moieties (F04). The most relevant descriptors are the number of sp^2 carbon atoms (A04), the number of tertiary amine nitrogens (A08), and the number of esters/carboxylates (F08), with some topological and atomistic descriptors (R02, B01), while relevant, having lesser importance. Interpretation of the least relevant descriptors is not particularly useful, but the high relevance of the carboxylic moiety, and amine, is consistent with the putative site of action of these compounds at the $GABA_A$-sensitive ($HOOCCH_2CH_2NH_2$) binding site of the benzodiazepine receptor. Again these relevances are not the same as those suggested by MLR or MLRBE/MLRFI methods, suggesting that nonlinearity is an important feature in this QSAR model.

The validity of ARD in determining the relevance of the descriptors was tested for the BZD data set by deleting (hard pruning) some of the less relevant descriptors from the BRANN model. Entry 7 shows that when the 5 least relevant descriptors (A26, B10, F01, F04, F06) are pruned, the model has almost identical statistics to that incorporating ARD. When the descriptors are pruned more severely and 10 are removed (entry 8), there is noticeable deterioration in the model statistics. This is in stark contrast to the MLRFI procedure which suggested removal of 21 descriptors, with a resulting poor model. It also illustrates how the variable elimination method MLRBE, which selected the same number of relevant variables, also gives a poor model because of the inherent nonlinearity of the underlying SAR. These results illustrate the importance of not relying on a linear method to carry out variable selection. Table 2 and Figures3 and 4 show the results for similar calculations on the muscarinic (MUS) and toxicity (TOX) data sets. The MUS set is not as well-modeled as the BZD set, whereas the TOX set is well-modeled and gives good statistics even at the MLR level.

The most relevant descriptors in the MUS model (log relevance $>$ 10) include topological indices R01 and K01, which are relatively uninformative, plus the numbers of carboxylate or ester moieties (F08) and amine (A08) moieties. These descriptors are consistent with the pharmacophore for this receptor, which requires a heterocyclic nitrogen atom connected to an ester mimetic by (most commonly) a three-atom linker.[42] Commonly the ester-mimetic is a five-membered oxazole or oxadiazole.[43]

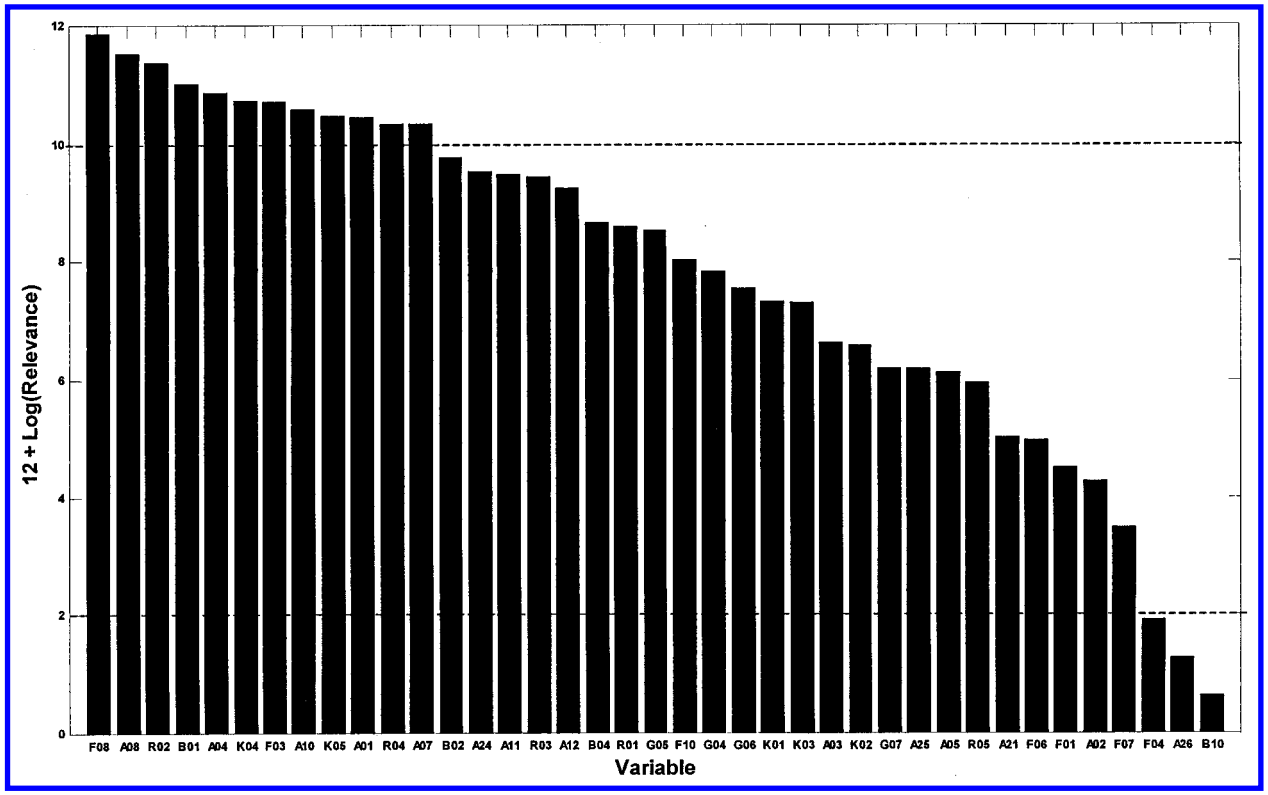

Figure 2. ARD relevances for the BZD data set.

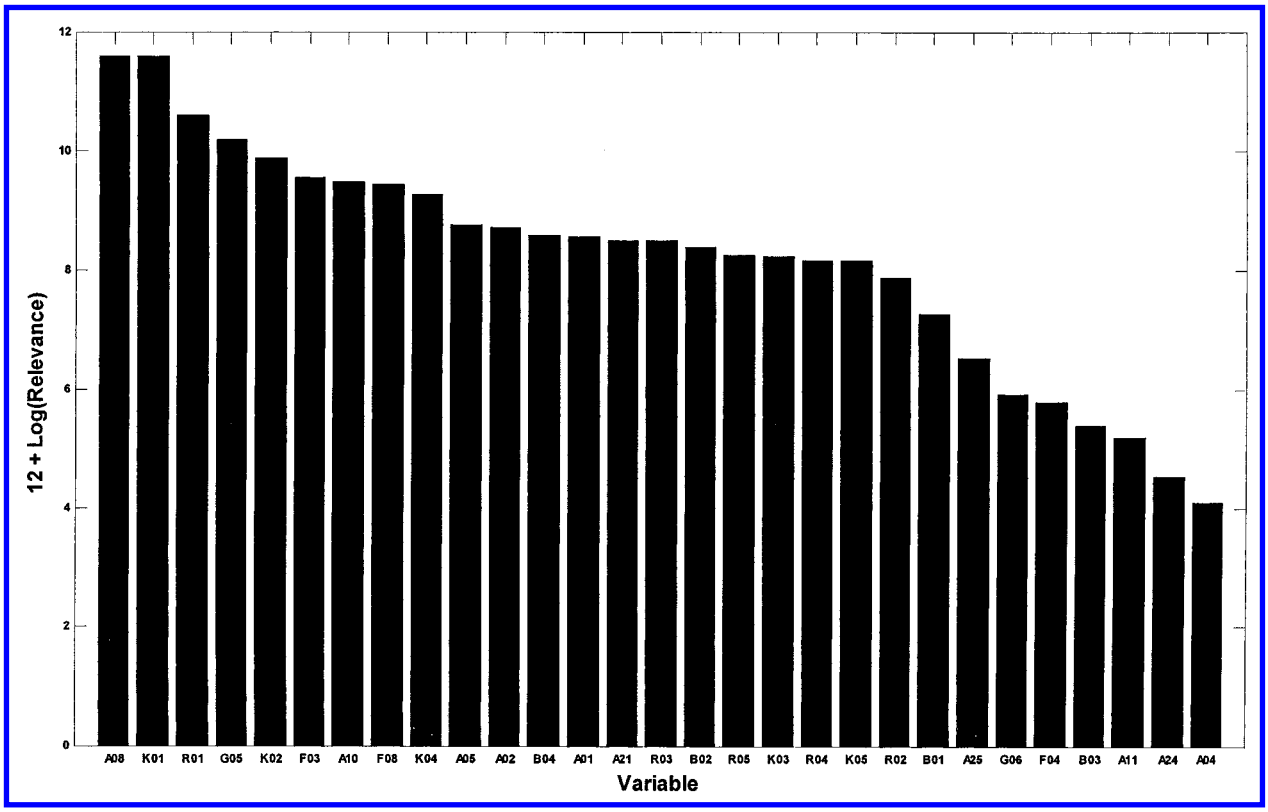

Figure 3. ARD relevances for the MUS data set.

The toxicology data are interesting as the compounds belong to four classes with potentially four different mechanisms of toxicity. The toxicity model produced by the regression methods is therefore a composite of four submodels, something general regression methods such as neural networks are very good at accounting for. The most relevant descriptors for this data set are the number of tertiary nitrogens (A08), the number of hydrogen bond donors (B10), the number of chlorine atoms (A24) and the number of ester/carboxylate groups (F08). Interpretation of these descriptors in terms of SAR is complex because of the multiple mechanisms operating and the complex nature of the topological indices used. However, a number of relevant descriptors are consistent with at least one of the toxicity

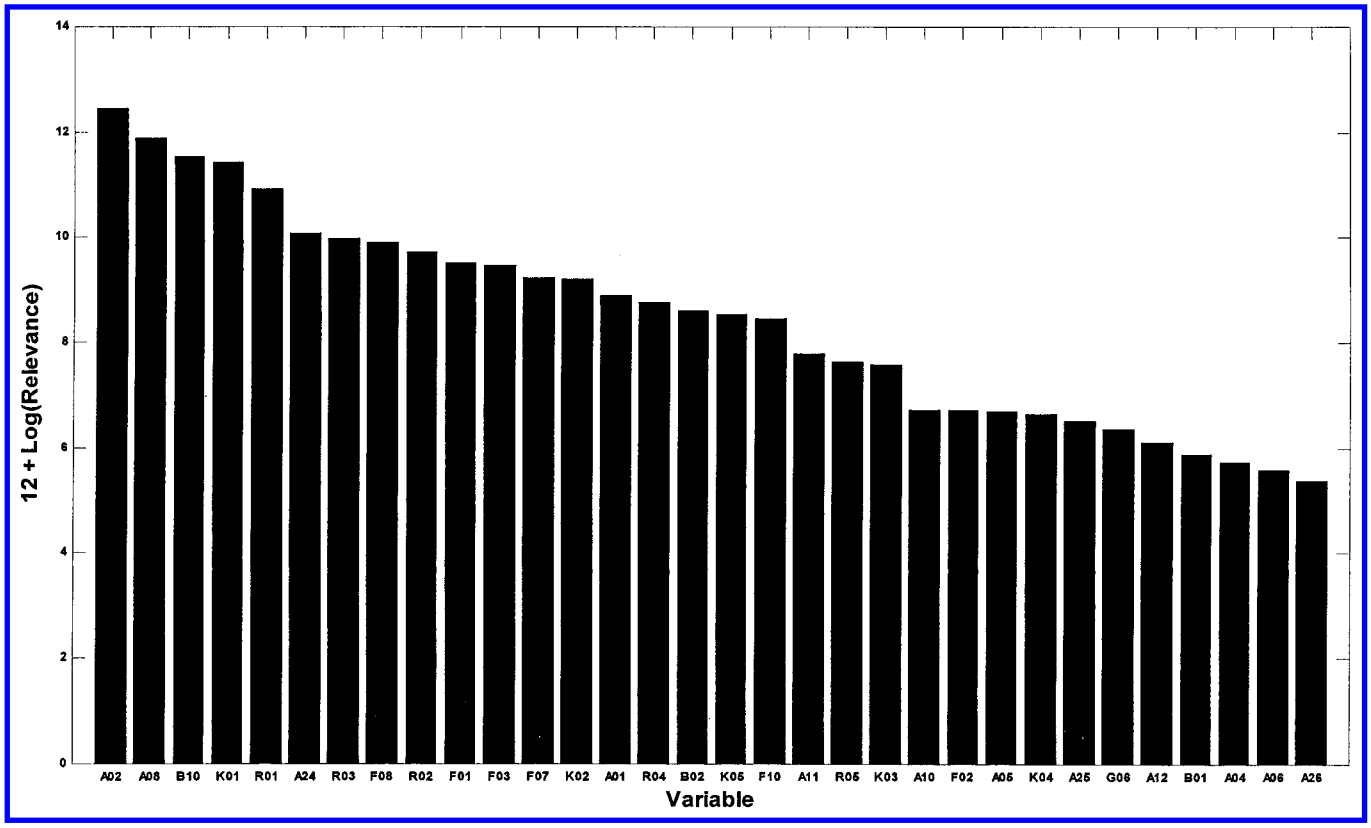

Figure 4. ARD relevances for the TOX data set.

mechanisms, uncoupling of oxidative phosphorylation, which requires lipophilic (halogens A24), weak acids (proton donors, eg B10, F08). Use of descriptors which are more chemically interpretable would considerably enhance the interpretation of the ARD relevances in terms of structure−activity or toxicity mechanisms.

4. CONCLUSIONS

The quality of the QSAR models produced by the BRANN and BRANN-ARD methods, compared with those from standard back-propagation neural net and MLR methods, illustrate their superiority. The BRANN method reduces the risk of overfitting and overtraining by means of the probabilistic interpretation of network training and produces the model using an objective criterion: the evidence for each hyperparameter. It also produces an estimate of the number of effective variables actually used in the production of the model that can be used to automatically optimize neural network architecture. The results show the importance of using a SAR mapping method able to deal with nonlinearities and variable dependences.

Our calculations also demonstrate the utility of adding ARD to BRANN to improve the robustness and ability to generalize the QSAR models. By separating the weights into different classes and effectively pruning them in the course of training, ARD objectively eliminates variables containing little relevant information. ARD automatically weights each of the input variables according to their importance to the model. These weightings can also be used to interpret the QSAR model by highlighting the features of the compounds that are important in producing the bioactivity.

We propose that BRANN coupled with ARD is useful in developing robust, interpretable QSAR models for drug discovery and optimization, or virtual screening. The type of indices used should be chosen to match the application, i.e., computationally efficient for screening, or chemically interpretable for drug design, optimization, or toxicity modeling. The BRANN-ARD model has addressed some of the perceived shortcomings of neural networks in QSAR modeling. We are extending this work to find more efficient methods (e.g., genetic algorithms) for optimizing the hyperparameters in ARD. This process, which currently uses conjugate gradient minimization methods to find the maximum likelihood of the parameters, may cause difficulties for large numbers of molecular descriptors. We are also employing BRANN-ARD to carry out virtual screening studies.

REFERENCES AND NOTES

(1) Hansch, C.; Fujita, T. $\rho-\sigma$-Analysis. A Method for the Correlation of Biological Activity and Chemical Structure. *J. Am. Chem. Soc.* **1964**, *86*, 1616−1626.
(2) Devillers, J., Ed. *Neural Networks in QSAR and Drug Design*: Academic Press: London, 1996.
(3) Winkler, D. A.; Burden, F. R. Robust QSAR Models from Novel Descriptors and Bayesian Regularized Neural Networks. *Mol. Simul.*, in press.
(4) Burden, F. R. Using Artificial Neural Networks to Predict Biological Activity from Simple Molecular Structural Considerations. *Quantum Struct.-Act. Relat*. **1996**, *15*, 7−11.
(5) Winkler, D. A.; Burden, F. R.; Watkins, A. J. R. Atomistic Topological Indices Applied to Benzodiazepines using Various Regression Methods. *Quantum Struct.-Act. Relat.* **1998**, *17*, 14−19.
(6) Burden, F. R.; Winkler, D. A. Robust QSAR Models Using Bayesian Regularized Artificial Neural Networks. *J. Med. Chem*. **1999**, *42*, 3183−3187.
(7) Hasegawa, K.; Funatsu, K. Ga Strategy for Variable Selection in QSAR Studies: GAPLS and D-Optimal Design for Predictive QSAR Studies. *J. Mol. Struct. (THEOCHEM)* **1998**, *425*, 255−262.
(8) Jouan-Rimbaud, D.; Massart, D. I.; de Noord, O. E. Random Correlation in Variable Selection for Multivariate Calibration with a genetic Algorithm. *Chemom. Intell. Lab. SysT*. **1996**, *35*, 213−220.

(9) Waller, C. L.; Bradley, M. P. Development and Validation of a Novel Variable Selection Technique with Application to Multidimensional Quantitative Structure−Activity Relationship Studies. *J. Chem. Inf. Comput. Sci*. **1999**, *39*, 345−355.
(10) Nath, R.; Rajagopalan, B.; Ryker, R. Determining the Saliency of Input Variables in Neural Network Classifiers. *Comput. Ops. Res*. **1997**, *24*, 767−773.
(11) Kovalishyn, V.; Tetko, V.I.; Luik, A. I.; Kholodovych, V. V.; Villa, A. E. P.; Livingstone, D. J. Neural Network Studies. 3. Variable Selection in the Cascade-Correlation Learning Architecture. *J. Chem. Inf. Comput. Sci*. **1998**, *38*, 651−659.
(12) Todeschini, R.; Galvagni, D.; Vilchez, J. L.; del Olmo, M.; Navas, N. Kohonen Artificial Neural Networks As A Tool For Wavelength Selection In Multicomponent Spectrofluorimetric PLS Modelling: Application to Phenol, o-Cresol, m-Cresol and p-Cresol Mixtures. *Trends Anal. Chem.* **1999**, *18*, 93−98.
(13) Zheng, W.; Tropsha, A. Novel Variable Selection Quantitative Structure−Property Relationship Approach Based on the *k*-Nearest-Neighbor Principle. *J. Chem. Inf. Comput. Sci*. **2000**, *40*, 185−194.
(14) Tetko, I. V.; Villa, A. E. P.; Livingstone, D. J. Neural Network Studies. 2. Variable Selection. *J. Chem. Inf. Comput. Sci*. **1996**, *36*, 794−803.
(15) Randic, M. On Characterization of Molecular Branching. *J. Am.Chem. Soc.* **1975**, *7*, 6609−6615.
(16) Kier, L. B.; Hall, L. H. The Molecular Connectivity Chi Indexes and Kappa Shape Indexes in Structure−Property Modeling. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; VCH Publishers: New York, 1995; Vol. 2, pp 367−422, and references cited therein.
(17) Burden, F. R. Using Artificial Neural Networks to Predict Biological Activity from Simple Molecular Structural Considerations. *Quantum Struct.-Act. Relat*. **1996**, *15*, 7−11.
(18) Winkler, D. A.; Burden, F. R.; Watkins, A. J. R. Atomistic Topological Indices Applied to Benzodiazepines using Various Regression Methods. *Quantum Struct.-Act. Relat*. **1998**, *17*, 14−19.
(19) Bishop, C. M. *Neural Networks for Pattern Recognition*; Oxford University Press: Oxford, U.K., 1995.
(20) MacKay, D. J. C. Bayesian methods for back-propagation networks. In *Models of Neural Networks III*; Domany, E., van Hemmen, J. L., Schulten, K., Eds.; Springer: New York, 1994; pp 211−254.
(21) Neal, R. M. *Bayesian Learning for Neural Networks*; Springer: New York, 1996.
(22) *MATLAB;* The MathWorks, Inc.: Natick, MA, 1999.
(23) *NETLAB Neural Network Software*; Neural Computing Research Group, Aston University: Birmingham, U.K. (http://www.ncrg.aston.ac.uk).
(24) Harrison, P. W.; Barlin, G. B.; Davies, L. P.; Ireland, S. J.; Matyus, P.; Wong, M. G. Syntheses, pharmacological evaluation and molecular modeling of substituted 6-alkoxyimidazo[1,2-b]pyridazines as new ligands for the benzodiazepine receptor. *Eur. J. Med. Chem*. **1996**, *31*, 651−662.
(25) Davies, **Λ. Π.;** Barlin G. B.; Ireland, S. J.; Ngu, M. M. L. Substituted imidazo[1,2-β]pyridazines. New compounds with activity at central and peripheral benzodiazepine receptors. *Biochem. Pharmacol*. **1992**, *44*, 1555−1561.
(26) Barlin, G. B.; Davies, L. P.; Davis, R. A.; Harrison, P. W. Imidazo-[1,2-β]pyridazines. XVII* Synthesis and central nervous system activity of some 6-(alkylthio and chloro)-3-(methoxy, unsubstituted and benzamidomethyl)-2-aryl-imidazo[1, 2-β]pyridazines containing methoxy, methylenedioxy and methyl substituents. *Aust. J. Chem.* **1994**, *47*, 2001−2012.
(27) Fryer, R. I.; Zhang, P.; Rios, R.; Gu, Z.-Q.; Basile, A. S.; Skolnick, P. Structure−activity relationship studies at the benzodiazepine receptor (BzR): A comparison of the substituent effects of pyrazoloquinoline analogues. *J. Med. Chem*. **1993**, *36*, 1669−1673.
(28) Wang, C.-G.; Langer, T.; Kamath, P. G.; Gu, Z.-Q.; Skolnick, P.; Fryer, R. I. Computer-aided molecular modeling, synthesis and biological evaluation of 8-(benzyloxy)-2-phenylpyrazolo[4,3-c]quinoline as a novel benzodiazepine receptor agonist ligand. *J. Med. Chem.* **1995**, *38*, 950−957.
(29) Hollinshead, S. P.; Trudell, M. L.; Skolnick, P.; Cook, J. M. Structural requirements for agonist actions at the benzodiazepine receptor: Studies with analogues of 6-(benzyloxy)-4-(methoxymethyl)-*b*-carboline-3-carboxylic acid ethyl ester. *J. Med. Chem.* **1990**, *33*, 1062−1069.
(30) Allen, M. S.; Hagen, T. J.; Trudell, M. L.; Codding, P. W.; Skolnick, P.; Cook, J. M. Synthesis of novel 3-substituted b-carbolines as benzodiazepine receptor ligands: Probing the benzodiazepine pharmacophore. *J. Med. Chem.* **1988**, *31*, 1854−1861.
(31) Yokoyama, N.; Ritter, B.; Neubert, A. D. 2-Arylpyrazolo[4,3-c]-quinolin-3-ones: Novel agonist, partial agonist and antagonist benzodiazepines. *J. Med. Chem.* **1982**, *25*, 337−339.
(32) Orlek, B. S.; Blaney, F. E.; Brown, F.; Clark, M. S. G.; Hadley, M. S.; Hatcher, J.; Riley, G. J.; Rosenberg, H. E.; Wadsworth, H. J.; Wyman, P., Comparison of Azabicyclic Esters and Oxadiazoles as Ligands for the Muscarinic Receptor. *J. Med. Chem.* **1991**, *34*, 2726−2735.
(33) Wadsworth, H. J., Jenkins, S. M.; Orlek, B. S.; Cassidy, F.; Clark, M. S. G.; Brown, F.; Riley, G. J.; Graves, D.; Hawkins, J.; Naylor, C. Synthesis and Muscarinic Activity of Quinuclidin-3-yltriazole and -tetrazole Derivatives. *J. Med. Chem*. **1992**, *35*, 1280−1290.
(34) Ward, J. S.; Merritt, L.; Klimkowski, V. J.; Lamb, M. L.; Mitch, C. H.; Bymaster, F. P.; Sawyer, B.; Shannon, H. E.; Olesen, P. H.; Honoré, T.; Sheardown, M. J.; Sauerberg, P. Novel functional M1 selective muscarinic agonists. 2. Synthesis and structure−activity relationships of 3-pyrazinyl-1,2,5,6-tetrahydro-1-methylpyridines. Construction of a molecular model for the M1 pharmacophore. *J. Med. Chem.* **1992**, *35*, 4011−4019.
(35) Sauerberg, P.; Olesen, P. H.; Nielsen, S.; Treppendahl, S.; Sheardown, M. J.; Honoré, T.,; Mitch, C. H.; Ward, J. S.; Pike, A. J.; Bymaster, F. P.; Sawyer, B. D.; Shannon, H. E. Novel functional M1 selective muscarinic agonists. Synthesis and structure−activity relationships of 3-(1,2,5-thiadiazolyl)-1,2,5,6-tetrahydro-1-methylpyridines. *J. Med. Chem.* **1992**, *35*, 2274−2283.
(36) Jenkins, S. M.; Wadsworth, H. J.; Bromidge, S.; Orlek, B. S.; Wyman, P. A.; Wiley: G. J.; Hawkins, J. Substituent Variation in Azabicyclic Triazole and Tetrazole-Based Muscarinic Receptor Ligands., *J. Med. Chem.* **1991**, *35*, 2392−2406.
(37) Sauerberg, P., Kindtlet, J. W.; Nielsen, L.; Sheardown, M. J.; Honoré, T. Muscarinic Cholinergic Agonists and Antagonists of the 3-(3-Alkyl-1,2,4-oxadiazol-5-yl)-1,2,5,6-tetrahydropyridine Type. Synthesis and Structure−Activity Relationships. *J. Med. Chem.* **1991**, *34*, 687−692.
(38) Burden, F. R.; Winkler, D. A. A Quantitative Structure−Activity Relationships Model for the Acute Toxicity of Substituted Benzenes to *Tetrahymena pyriformis* using Bayesian Regularized Neural Networks. *Chem. Res. Toxicol.* **2000**, *13*, 436−440.
(39) Dearden, J. C.; Cronin, M. T. D.; Schultz, T.W. Lin, D.T. QSAR Study of the Toxicity of Nitrobenzenes to Tetrahymena Pyriformis. *Quantum Struct.-Act. Relat.* **1995**, *14*, 427−432.
(40) Cronin, M. T. D.; Bryant, S. E.; Dearden, J. C.; Schultz, T. W. Quantitative Structure−Activity Study of the Toxicity of Benzonitriles to the Ciliate Tetrahymena Pyriformis. *SAR QSAR Environ. Res.* **1995**, *3*, 1−13.
(41) Cronin, M. T. D.; Gregory, B. W.; Schultz, T. W. Quantitative Structure−Activity Analyses of Nitrobenzene to Tetrahymena Pyriformis. *Chem. Res. Toxicol*. **1998**, *11*, 902−908.
(42) Shapiro, G.; Floersheim, P.; Boelsterli, J.; Amstustz, R.; Bollinger, G.; Gamenthaler, H.; Gmelin, G.; Supavilai, P.; Walkinshaw, M. *J. Med. Chem*. **1992**, *35*, 15−27.
(43) Sanders, J.; Cassidy, M.; Freedman, S. B.; Harley, E. A.; Iverson, L. L.; Kneen, C.; MacLeod, A. M.; Merchant, K. J.; Snow, R. J.; Baker, R. *J. Med. Chem.* **1990**, *33*, 1128−1138.

CI000450A