Quantitative Structure−Activity Relationships in Carboquinones and Benzodiazepines Using Counter-Propagation Neural Networks

Keith L. Peterson

Division of Science and Mathematics, Wesleyan College, 4760 Forsyth Road, Macon, Georgia 31297

Received May 18, 1993[⊗]

Counter-propagation neural networks are used to model and predict activities of carboquinones and of benzodiazepines from physicochemical parameters. For carboquinones, networks with one hidden layer processing element (PE) for each compound achieved significantly better training set RMSE values than corresponding back-propagation and multiregression results and test set RMSE values as good or slightly worse than back-propagation. Test set results improved by 10−15% using networks with fewer hidden layer PEs than carboquinones; the smallest test set RMSE values are between 0 and 10% better than back-propagation values, about 1.3 times greater than corresponding training set values, and occur when there are about as many competitive layer PEs as there are compounds in the data set. Training set RMSE values increase with decreasing number of competitive layer PEs and approach those of test sets. Both counter-propagation and back-propagation networks, however, have worse predictive capability than multiregression. For benzodiazepines, networks with one hidden layer PE for each compound achieved significantly better training set RMSE values than back-propagation and multiregression results and test set RMSE values slightly worse than back-propagation. Test set results improved by 10−15% using fewer hidden layer PEs than benzodiazepines; the smallest test set RMSE values are 0−10% better than back-propagation values, about 1.3 times greater than training set values, and occur when there are about half as many competitive layer PEs as there are compounds in the data set. Training set RMSE values increase with decreasing number of competitive layer PEs and approach those of test sets. Counter-propagation, back-propagation, and multiregression all have similar predictive capabilities.

I. INTRODUCTION

In a previous paper[1] the utility of counter-propagation neural networks in quantitative structure-chromatographic retention relationships (QSRRs) was demonstrated by modeling and predicting the Kovats indices of substituted phenols. In this paper the same methodology is applied to quantitative structure−activity relationships (QSARs) by modeling and predicting the activities of carboquinones and benzodiazepines. The results will be compared to previous results obtained from back-propagation neural networks[2] and from multiple regression analyses.[3] This paper is one of the first (perhaps the first) computational studies where results from multiple regression and back-propagation and counter-propagation neural networks are directly compared. For the carboquinones and benzodiazepines used in this study the counter-propagation results are at least as good as and often much better than back-propagation results. However, for carboquinones, both network types give results slightly worse than those for multiple regression; for benzodiazepines all three methods give similar results.

The remainder of the paper is organized as follows. Section II contains a brief summary of counter-propagation neural networks, the data to which these networks were applied, and the methodology of this application. Section III contains the computational results and a comparison to back-propagation and multiple regression results. Section IV contains a summary and conclusions.

II. COUNTER-PROPAGATION NETWORKS, METHODOLOGY, AND DATA

A detailed discussion of neural networks in general, of counter-propagation networks in particular, and of the methodology of their application to structure−property relationship problems is given in ref 1. Only a brief summary of these matters will be given here.

A neural network consists of many processing elements (PEs) which are joined by input and output paths. Typically the PEs are organized into a series of layers. The first layer is the input layer with one PE for each variable or feature of the data. The last layer is the output layer consisting of one PE for each variable to be recognized. In between are a series of one or more hidden layers consisting of a number of PEs which are responsible for learning. PEs in any layer are fully or randomly connected to PEs of a succeeding layer. Each connection is represented by a number called a weight.

This paper will be concerned with networks undergoing supervised learning, wherein the network is provided with a set of inputs (in our case a set of pattern vectors each describing a given carboquinone or benzodiazepine by a set of physicochemical parameters—see below) and a set of desired outputs. Each output represents an activity for a given drug compound. Upon repeatedly presenting the input and output sets to the network, the weights between PEs are gradually adjusted so that finally the network will generate the correct output when given a certain input. After this training or learning phase the network can be presented with an input whose corresponding output is unknown. When this input is fed through the successive PE layers, it is

[⊗] Abstract published in *Advance ACS Abstracts,* August 15, 1995.

0095-2338/95/1635-0896$09.00/0

transformed into a predicted output (drug activity). This prediction process can also be used with inputs whose outputs are known in order to test the training level of the network. If the network is fully trained, the predicted output should match the known output.

A brief description of the operation of a typical PE in a counter-propagation neural network will now be given. During the operation of a network a PE will receive inputs from other PEs to which it is connected. These inputs are the outputs of the connected PEs and are combined via a summation function. Two types of summation functions are relevant to counter-propagation networks. The first is the weighted sum

$$I_i = \sum w_{ij} x_j \tag{1}$$

where I_i is the result of the summation function, W_{ij} is the weight of the connection from processing element j to processing element i, and x_j is the output of the PE from which the connection originates. The second summation function is a normalization function whereby the input vector to the network $(x_1, x_2, ..., x_n)$ is replaced by an augmented vector $(x_0, x_1, ..., x_n)$, where x_0 is chosen such that the augmented vector is normalized. The result of the summation function for a PE is transformed to an output of the PE with a transfer function. In a counter-propagation network this function is

$$T_i = I_i \tag{2}$$

This is a linear transfer function where I_i is, for example, obtained from eq 1. The result of the transfer function is then acted upon by an output function. One type of such function is a simple direct output function. Another type allows for competition between PEs in a layer. Each PE in a layer has a number associated with it as a result of the transfer function. The PE with the highest number is said to win. Use of a "one-highest" output function allows the value from the winning PE to be passed to PEs in the next layer but prevents values from all other PEs in the layer from being passed. Only weights associated with the winning PE can change. In a competitive layer of a counter-propagation network, the weights change according to the Kohonen learning rule

$$W_{ij}' = W_{ij} + C(x_{ij} - W_{ij}) \tag{3}$$

where W_{ij} is the weight before learning has occurred, x_{ij} is the jth input to the ith PE in the layer, C is a learning coefficient, and W_{ij}' is the weight after learning. The learning coefficient satisfies $0 < C \ll 1$. In all work reported here C was set to 0.1; changing the value of C had no effect on the performance of the networks. The output layer of a counter-propagation network uses the Widrow–Hoff learning rule

$$W_{ij}' = W_{ij} + Cx_{ij}(d_i - y_i) \tag{4}$$

where d_i is the desired output (in our case a drug activity) of the ith PE and y_i is the output of the ith PE; that is, $d_i - y_i$ is the error in the output of the ith PE.

Counter-propagation neural networks will now be briefly discussed. In order to optimally use this network for the prediction of drug activities two requirements should be met. First, there must be enough PEs in the competitive layer to ensure that they can adapt sufficiently well to the outputs. Second, the inputs should be selected so that they uniformly cover the value ranges of inputs and outputs expected to be encountered; this is more or less a requirement for all neural network types. The first requirement can be met by experimenting with networks containing different numbers of PEs in the competitive layer. Typically, there will be one PE in this layer for each input vector. Such a counter-propagation network will be referred to as a standard or default network. It will be shown later that the best predictions of drug activities are obtained from networks with fewer PEs in the competitive layer than there are drugs in a data set. The second requirement may or may not be met depending on the completeness of the data set.

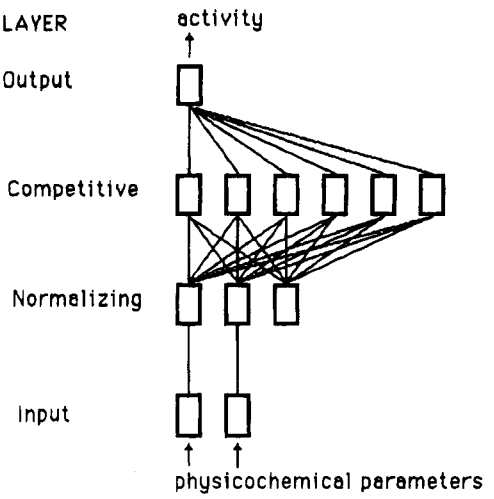

Figure 1. Sample counter-propagation neural network. In this example the input layer consists of two PEs; this would be appropriate if two physicochemical parameters were used as input features describing a data set. The normalizing layer contains one more PE than the input layer. The competitive layer containing six PEs would be appropriate for a data set consisting of six drug compounds (this assumes a "default" network—see text). The output layer contains one PE for the predicted activity.

The counter-propagation networks used in the present application consist of four layers of PEs. (See Figure 1 for a schematic of a sample counter-propagation network.) The first is the input layer with one PE for each feature. The input layer utilizes a weighted sum summation function, a linear transfer function, a direct output function, and no learning rule. The first hidden layer is a normalizing layer containing one more PE than the input layer. This layer ensures that every input vector has the same length, i.e., that the vectors are normalized. This layer utilizes the normalization summation function, a linear transfer function, a direct output function, and no learning rule. The second hidden layer is the competitive layer which acts as a nearest-neighbor classifier. The PEs in this layer learn with the Kohonen learning rule. During training each PE competes with the others in the layer and is equally likely to win for any randomly chosen input. For a given input vector, however, only one PE in this layer wins. The PE that wins is the PE which has the highest output. Equivalently, the winning PE is the PE whose weight vector (i.e., the set of numbers W_{ij}) is closest to the input vector. This is what is meant by the phrase "nearest-neighbor classifier". The competitive layer utilizes a weighted sum summation function, a linear transfer function, a one-highest output function, and the Kohonen learning rule. The output of the winning competitive layer PE is the input to the output layer which in our case contains

one PE. This PE will output the predicted drug activity for a given input vector (i.e., drug compound) by using the Widrow−Hoff learning rule to interpret the output from the competitive layer. For a given input to the network there will be only one output from the competitive layer, that output coming from the winning PE. The Widrow−Hoff rule enables the output PE to associate a drug activity (i.e., desired network output) with each competitive layer PE. The output layer utilizes a weighted sum summation function, a linear transfer function, a direct output function, and the Widrow−Hoff learning rule. The normalizing layer is fully interconnected to the competitive layer which is in turn fully connected to the output layer. The total number of learning weights in a counter-propagation network is equal to the number of competitive layer PEs plus the product of the number of competitive layer PEs and the number of normalizing layer PEs.

One problem which may arise in the operation of the network is that the competitive layer learns without supervision, meaning that it is possible for a PE in this layer to take responsibility for two or more input vactors. When this happens the ouput of the network will be ambiguous for any input vectors which activate this PE, i.e., for such an input it will be impossible for the network to predict or learn the correct drug activity. This does not mean, however, that the predicted numerical value of the drug activity has a large error associated with it. This point will be addressed in the next section.

The carboquinone data has appeared in several references,[2−4] and therefore only a brief description of it will be given here. Six physicochemical parameters (molecular refractivity constant MR1, hydrophobicity constant $\pi 1$, substituent constants F and R, and MR12 and $\pi 12$ to estimate the steric effects of functional groups at the 1- and 2-positions and the total hydrophobicity, respectively) were measured for each of 39 different carboquinones. Activities were expressed as minimum effective dose (MED) and as optimal dose (OD). Each of these doses was measured for a chronic treatment schedule and for a single injection. This gives four data sets: MED chronic, MED singular, OD chronic, and OD singular. There were 37, 35, 37, and 37 members (compounds), respectively, in each data set (not all activities were given for each of the 39 compounds).

The benzodiazepine data has also appeared in several references[2,3,5] and will only be briefly described here. Seven physicochemical parameters (these are similar to those for carboquinones and are denoted by MR3, MR7, $\pi 3$, $\sigma 3$, F4, R4, and I1) were measured for each of 60 different benzodiazepines. Activities were expressed as anti-pentylenetetrazole effect (anti-pent), anti-fighting effect (anti-fight), and clined screen test (clined scr). These three data sets contained 57, 53, and 53 members, respectively.

Within a given data set each physicochemical parameter (input feature) was rescaled to have values between 0.1 and 1.0 according to the following equation

$$X_i = (x_i - x_{\min} + 0.1)/(x_{\max} - x_{\min} + 0.1) \qquad (5)$$

where X_i is the rescaled value of a parameter, x_i is the unscaled value of the parameter, and $x_{\max}$ and $x_{\min}$ are the maximum and minimum unscaled values of the parameter within a given data set. Activities (desired outputs) within each data set were scaled to have values between 0.0 and 1.0. This scaling is the same as was done in a back-propagation study of the same data sets[2] and was done here to allow for direct comparison of results to the back-propagation study.

The counter-propagation networks used in the present study were constructed in accordance with paragraph six of this section. As an example, a standard or default network for the MED chronic data set consisted of an input layer containing six PEs (one for each of the physicochemical parameters), a normalizing layer containing seven PEs (one more PE than the input layer), a competitive layer containing 37 PEs (this is the default configuration with one PE for each member of the data set), and an output layer containing one PE (for the drug activity). Such a network has 296 learning weights; back-propagation networks used in ref 2 contained either 96 or 364 learning weights.

The authors of ref 2 also considered data sets which, in addition to six (for carboquinones) or seven (for benzodiazepines) physicochemical parameters, also contained the squares of each parameter. These data sets are also studied in this paper; a default network for such an MED chronic data set would be identical to that described in the above paragraph except that it would have 12 PEs in the input layer and 13 in the normalizing layer.

All counter-propagation networks were trained by presenting each input vector for each drug in a data set to the network 300 times. No improvement in the root mean square error (RMSE) of the output (drug activities) was observed for larger numbers of iterations. In fact, each output converged to a value which did not change after 300 iterations.

This paper will also report the results of back-propagation and multiple regression calculations for carboquinones and benzodiazepines, many of which have not appeared in refs 2−5. These calculations were performed by the author in order to compare them with the counter-propagation results. All back-propagation calculations were done using back-propagation neural networks which used the same parameters and were configured in exactly the same way as those used in ref 2. Multiple regression calculations were done using equations from refs 3 and 5. All equations contained a constant. Equations for carboquinone chronic MED and OD data sets used MR1, $\pi 12$, F, and R as variables, while those for singular MED and OD data sets used $\pi 12$, F, and R as variables. Equations for all benzodiazepine data sets used MR7, $\sigma 3$, F4, and I1 as variables. The anti-pent data set equations also used MR3 as a variable, the anti-fight data set equations also used $\pi 3$ and R4 as variables, and the clined scr data set equations also used $\pi 3$ and MR4 as variables.

All networks were run on IBM AT compatible PCs with 80386SX chip operating at 16 MHz, 1-MB memory, 40-MB hard disk, and VGA monochrome monitor using the Neuralworks Professional II software package (NeuralWare Inc., Penn Center West, Building IV, Suite 227, Pittsburgh, PA 15276). Multiple regression calculations were performed on a Macintosh SE with 1-MB memory and 20-MB hard disk using the MYSTAT software package.

III. RESULTS AND DISCUSSION

a. Carboquinones. Table 1 shows a comparison of neural network and multiregression training set results for the carboquinone MED chronic data set. The counter-propagation networks are default networks, i.e., networks with one competitive layer PE for each member of the data

Table 1. Comparison of Default Counter-Propagation, Back-Propagation, and Multiple Regression Training Set Results for the Carboquinone MED Chronic Dtaa Set

no.[a]	obsd[b]	CP6[c]	BP6[d]	CP12[e]	BP12[f]	MR[g]
1						
2	4.33	4.33	4.26	4.33	4.36	4.05
3	4.47	4.47	4.63	4.47	4.55	4.61
4	4.63	4.63	4.33	4.63	4.30	4.31
5	4.77	4.84	5.10	4.77	5.14	5.26
6	4.85	5.00	5.11	5.00	5.09	5.18
7	4.92	4.84	4.98	4.92	5.01	5.15
8	5.15	5.00	5.13	5.00	5.21	5.21
9	5.16	5.16	5.19	5.16	5.05	5.21
10	5.46	5.46	5.46	5.46	5.52	5.57
11	5.57	5.57	5.98	5.57	6.02	5.98
12	5.59	5.59	5.71	5.59	5.73	5.74
13	5.60	5.60	5.54	5.60	5.58	5.58
14	5.63	5.63	5.80	5.63	5.81	6.03
15						
16	5.66	5.66	6.07	5.76	6.09	5.99
17	5.68	5.77	5.66	5.68	5.73	5.56
18	5.68	5.68	5.50	5.68	5.62	5.54
19	5.68	5.69	5.72	5.68	5.74	5.96
20	5.69	5.69	5.56	5.69	5.61	5.59
21	5.76	5.76	5.76	5.76	5.83	5.93
22	5.78	5.78	5.82	5.78	5.87	5.87
23	5.82	5.82	5.43	5.82	5.47	5.47
24	5.86	5.77	5.73	5.76	5.78	5.73
25	6.03	6.03	6.33	6.03	6.27	6.33
26	6.14	6.28	6.10	6.28	6.12	6.12
27	6.16	6.16	6.10	6.16	6.12	6.19
28	6.18	6.18	5.77	6.18	5.80	5.86
29	6.18	6.18	6.06	6.30	6.11	6.09
30	6.18	6.18	5.98	6.18	6.01	6.02
31	6.21	6.21	6.33	6.21	6.32	6.28
32	6.25	6.25	6.02	6.25	6.04	6.12
33	6.39	6.39	6.47	6.39	6.43	6.34
34	6.41	6.28	6.10	6.28	6.12	6.12
35	6.41	6.68	6.42	6.30	6.40	6.35
36	6.45	6.45	6.62	6.45	6.58	6.54
37	6.54	6.54	6.38	6.54	6.33	6.12
38	6.77	6.77	6.47	6.77	6.43	6.56
39	6.90	6.68	6.56	6.90	6.52	6.40

[a] Refers to compound number as given in ref 2. A blank entry means no value was given in this reference. [b] Drug activities from refs 2–4. [c] Default counter-propagation results with six physicochemical input parameters. [d] Back-propagation results with six physicochemical input parameters from ref 2. [e] Default counter-propagation results with six physicochemical parameters and their squares as input (12 total input parameters). [f] Back-propagation results from ref 2; input parameters same as *e*. [g] Multiple regression results from ref 3.

set (37 in this case). The counter-propagation results are much better than either back-propagation or multiregression (MR) results. (This is also reflected in Table 2 which gives a comparison of mean deviations (MD) and root mean square errors (RMSE) for all of the carboquinone training data sets.) The counter-propagation network with six physicochemical parameters as input (CP6) learns the correct activities for all but 11 of the carboquinones; the largest absolute error is 0.22 (compounds 35 and 39). In contrast, the analogous back-propagation network (BP6) learns two activities correctly with the largest absolute error being 0.41 (compounds 11, 16, and 28). The CP6 network gives better training results than the BP6 network for all compounds except 7, 8, 17, 26, and 35, and better results than MR for all compounds except 8, 26, and 35. The counter-propagation network with six physicochemical parameters and their squares as input (CP12) fails to learn only eight activities. The largest error is 0.15 (compounds 6 and 8). The analogous back-propagation network (BP12) does not learn any activities correctly and exhibits a largest absolute error of 0.43

Table 2. Comparison of Carboquinone Training Set Mean Deviation (MD) and Root Mean Square Error (RMSE) for Default Counter-Propagation, Back-Propagation, and Multiple Regression[a]

	chronic injection									
	MED					OD				
	CP6	BP6	CP12	BP12	MR	CP6	BP6	CP12	BP12	MR
MD	0.04	0.17	0.03	0.16	0.20	0.01	0.15	0.01	0.14	0.19
RMSE	0.08	0.21	0.06	0.21	0.24	0.02	0.19	0.05	0.18	0.23
	single injection									
	MED					OD				
	CP6	BP6	CP12	BP12	MR	CP6	BP6	CP12	BP12	MR
MD	0.02	0.20	0.08	0.19	0.23	0.02	0.16	0.01	0.14	0.20
RMSE	0.04	0.25	0.20	0.24	0.27	0.08	0.20	0.05	0.17	0.24

[a] See footnotes in Table 1 for explanation.

(compound 11). The CP12 network gives better training results than the BP12 network for all compounds except 8, 24, 26, 29, and 35 and better results than MR for all compounds except 8, 26, 29, and 36. The MR also fails to learn any activities correctly and has a largest absolute error of 0.5 (compound 39).

Table 2 gives a comparison of mean deviations (MD) and root mean square errors (RMSE) for all of the carboquinone training data sets. The CP6 results are better than both the BP6 and MR results and the CP12 results are better than those of both BP12 and MR. The differences between using six physicochemical parameters as input and six parameters and their squares as input are very small; the BP results are slightly better (by 0.01 to 0.03) with the added input parameters, while the CP results are sometimes better and sometimes worse (although again by 0.01 to 0.03). The only exception to this occurs for the single injection MED data set where the MD value goes from 0.02 for CP6 to 0.08 for CP12, and the RMSE value goes from 0.04 for CP6 to 0.2 for CP12.

All of the counter-propagation training errors reflected in Tables 1 and 2 are the result of some PEs in the competitive layer taking responsibility for learning about two or more drug activities instead of just one. This gives rise to ambiguous output. This is an example of the problem discussed in section II between eqs 4 and 5; the same phenomenon was observed in ref 1.

A set of computer experiments[6] designed to test the usefulness of counter-propagation neural networks in predicting drug activities was performed next. This method is the leave-one-out method and is described in detail in ref 1. For the carboquinone MED chronic data set 37 disjoint sets, each composed of one carboquinone, were formed. The 36 carboquinones remaining after each test set selection served as the corresponding training set. A given network was trained and tested on each training/testing set pair. The networks were configured in the same manner as described earlier; a default network contained 36 competitive layer PEs. All other details of these calculations are the same as described earlier. The default network results are given in the third column of Table 3. Each value in this column is the predicted drug activity of a compound (labeled in column 1) from a network which was trained on all other compounds. Generally, the results are quite good; notable exceptions are compounds 32, 38, 36, 12, 11, and 3, for which the absolute value of differences between predicted and observed values are 1.69, 1.2, 0.86, 0.86, 0.66, and 0.53, respectively.

Table 3. Comparison of Six Physicochemical Parameter Counter-Propagation Leave-One-Out Test Set Results for the Carboquinone MED Chronic Data Set

no.[a]	obsd[b]	CP[c]	CP29[d]	CP16[e]	BP[f]	MR[g]
1						
2	4.33	4.63	4.63	4.55	4.48	3.82
3	4.47	5.00	5.00	4.92	4.66	4.66
4	4.63	4.33	4.33	4.40	4.07	4.19
5	4.77	4.92	4.92	4.84	5.11	5.29
6	4.85	5.15	5.42	4.94	5.13	5.23
7	4.92	5.12	5.39	4.92	4.94	5.17
8	5.15	4.85	5.68	4.85	5.16	5.22
9	5.16	5.69	5.63	5.70	5.24	5.23
10	5.46	5.60	5.71	5.74	5.44	5.58
11	5.57	6.23	5.78	6.24	6.01	6.00
12	5.59	6.45	6.45	6.45	6.97	6.25
13	5.60	5.82	5.72	5.70	5.51	5.58
14	5.63	5.76	5.76	5.76	5.80	6.08
15						
16	5.66	5.86	5.66	6.39	6.28	6.01
17	5.68	5.86	5.71	5.68	5.87	5.54
18	5.68	5.67	5.67	5.69	5.59	5.53
19	5.68	5.69	5.69	5.69	5.65	5.99
20	5.69	5.68	5.68	5.69	5.57	5.58
21	5.76	6.16	5.89	5.90	5.77	5.94
22	5.78	5.57	6.18	5.93	5.84	5.88
23	5.82	5.53	5.64	5.72	5.37	5.45
24	5.86	5.62	5.63	5.70	5.77	5.73
25	6.03	6.18	6.29	6.18	6.57	6.37
26	6.14	6.27	6.29	6.26	6.18	6.12
27	6.16	5.70	5.69	5.69	6.09	6.19
28	6.18	6.18	6.18	6.25	5.76	5.85
29	6.18	6.20	6.41	6.03	6.20	6.08
30	6.18	6.27	6.25	6.25	6.04	6.01
31	6.21	6.27	6.24	6.27	6.32	6.29
32	6.25	4.56	6.23	6.25	6.02	6.12
33	6.39	6.21	6.72	6.25	6.69	6.33
34	6.41	6.14	6.18	6.03	6.14	6.11
35	6.41	6.12	6.11	6.03	6.60	6.34
36	6.45	5.59	5.59	5.59	5.96	6.58
37	6.54	6.21	6.21	6.58	6.52	6.04
38	6.77	5.57	6.21	6.58	5.86	6.39
39	6.90	6.41	6.21	6.30	6.69	6.34

[a] Refers to compound number as given in ref 2. A blank entry means no value was given in this reference. [b] Drug activities from refs 2–4. [c] Default (36 competitive layer PEs) counter-propagation results. [d] Twenty-nine competitive layer PE computer-propagation results. [e] Sixteen competitive layer PE counter-propagation results. [f] Back-propagation results. [g] Multiple regression results.

Analogous results for back-propagation networks and multiregression are not given in the literature.[2-5] Therefore, for purposes of comparison, back-propagation and multiple regression calculations were performed as described at the end of section II. The results are given in columns six and seven of Table 3. The back-propagation and multiple regression results are generally comparable to the counter-propagation results; this is consistent with the fact that test set RMSE values for both network types and multiple regression are very similar (see below and Table 4). Back-propagation performs poorly for compounds 12, 38, 16, and 4, for which the absolute value of differences between predicted and observed values are 1.38, 0.91, 0.62, and 0.56, respectively. Multiple regression performs poorly for compounds 12, 39, 5, 2, and 37, for which the absolute value of differences between predicted and observed values are 0.66, 0.56, 0.52, 0.51, and 0.5, respectively.

For default counter-propagation networks an RMSE value of 0.33 for the test sets was obtained by calculating the RMSE for each test set and then averaging over the 37 test sets. The standard deviation of RMSE values over the 37 test sets is 0.35. For back-propagation networks the averaged test set RMSE value is 0.26 with a standard deviation of 0.28. For multiple regression the averaged test set RMSE value is 0.26 with a standard deviation of 0.30. These values indicate that the predictive power of multiple regression and back-propagation is essentially the same, while that of the default counter-propagation network is not quite so good.

Table 4. Comparison of Carboquinone Leave-One-Out Root Mean Square Error (RMSE) and Standard Deviations (STD) for Counter-Propagation, Back-Propagation, and Multiple Regression[a]

	chronic injection								
	MED					OD			
	CP6	BP6	CP29[b]	CP16[b]	MR	CP6	BP6	CP24[b]	MR
RMSE	0.33	0.26	0.25	0.24	0.25	0.27	0.26	0.23	0.23
STD	0.35	0.28	0.22	0.25	0.18	0.29	0.27	0.23	0.16
	single injection								
	MED				OD				
	CP6	BP6	CP29[b]	MR	CP6	BP6	CP12[b]	MR	
RMSE	0.44	0.48	0.38	0.26	0.47	0.30	0.32	0.23	
STD	0.58	0.77	0.45	0.18	0.58	0.31	0.45	0.18	

[a] See footnotes in Table 1 for explanation. [b] A counter-propagation network with either 29, 16, 24, or 12 hidden layer PEs.

There are two concerns which arise when comparing the above counter-propagation results to those of back-propagation. The first is that the counter-propagation test set RMSE is a factor of 4.1 times greater than the training set RMSE. This may raise the question of the robustness of counter-propagation networks. In fact, this is not an issue for such networks; a detailed discussion of this point is given on page 383 of ref 1 and will not be repeated here. Additionally, in the next paragraph computational results which address this point directly will be described. The second concern is that the default counter-propagation networks do not perform as well as back-propagation networks.

A set of computations was performed in order to achieve the best possible test set RMSE values. They are completely analogous to the leave-one-out calculations described above except that instead of using default networks there are differing numbers of PEs in the competitive layer. That is, a set of leave-one-out calculations with 35 competitive layer PEs was performed, as was a set with 34 competitive layer PEs, a set with 33 competitive layer PEs, etc., to a minimum of six competitive layer PEs. Using fewer competitive layer PEs than input vectors forces the networks to give ambiguous output; i.e., it forces some of the competitive layer PEs to take responsibility for two or more input vectors. The idea is to force the network to recognize statistical trends of the entire data set rather than focusing on each input vector of the data set individually. The results are shown in Figure 2. (With fewer than six competitive layer PEs the RMSE values begin to get very large.) The training set RMSE values show a generally increasing trend as the number of competitive layer PEs decreases. The test set RMSE values are generally best in the range of 22 to 14 competitive layer PEs. Over this range the ratio of test set RMSE to training set RMSE varies from 1.26 to 2.17. The best test set RMSE value of 0.24 occurs for 16 competitive layer PEs and is slightly better than the back-propagation result of 0.26; the test set RMSE to training set RMSE ratio for this network is 1.33, roughly comparable to the corresponding back-propagation ratio of 1.24 (see above). Thus, the robustness

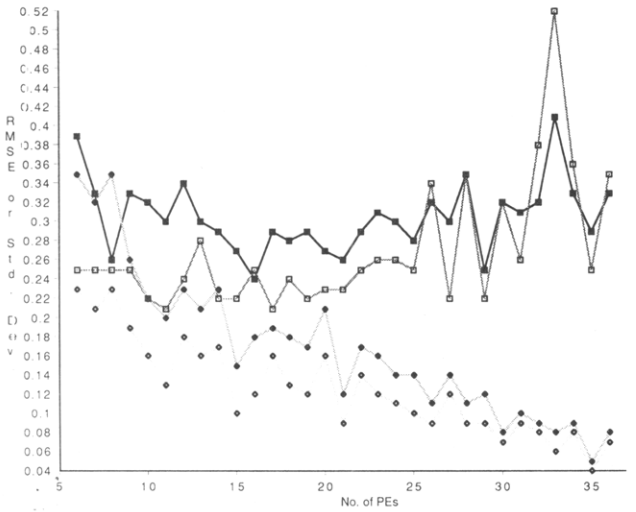

Figure 2. Variation of RMSE values and corresponding standard deviations for leave-one-out counter-propagation networks with number of competitive layer PEs for carboquinone MED chronic data. Solid squares represent RMSE values of test sets and open squares the corresponding standard deviations. Solid diamonds represent RMSE values of training sets and open diamonds the corresponding standard deviations. RMSE values are obtained by calculating the RMSE for each test/training set and then averaging over the 37 test/training sets. Standard deviations are those of RMSE values over the 37 test/training sets.

of the two network types is equivalent and the counter-propagation network can perform slightly better than the back-propagation network. Counter-propagation networks with 29 and 21 competitive layer PEs have test set RMSE values of 0.25 and 0.26, respectively, both of which are comparable to the back-propagation result.

The counter-propagation network test set results for 29 and 16 competitive layer PEs are given in the fourth and fifth columns of Table 3. Generally, the results are quite good; notable exceptions for the 29 PE case are compounds 36, 12, 6, and 8, for which the absolute value of differences between predicted and observed values are 0.86, 0.86, 0.57, and 0.53, respectively. Notable exceptions for the 16 PE case are compounds 36, 12, 16, 11, and 9, for which the absolute value of differences between predicted and observed values are 0.86, 0.86, 0.73, 0.67, and 0.54, respectively. Relative to the default network these errors are generally smaller. Training set results for the 29 and 16 PE networks are qualitatively very similar to those given in Table 1 for the default network and are therefore not shown here. The increase in RMSE values for these two nondefault networks over that for the default network arises primarily from their inability to correctly learn as many activities. This is to be expected since some of the fewer competitive layer PEs are forced to take responsibility for more than one drug activity.

Figures 3−5 show RMSE values and corresponding standard deviations of leave-one-out calculations for the other three carboquinone data sets using counter-propagation networks with differing numbers of competitive layer PEs. Six physicochemical parameters were used as input. Since the trends in RMSE values with number of PEs is the same for these data sets as described in the second-to-last paragraph above for chronic MED data, results are shown only for every sixth PE. Again, the training set RMSE values increase with decreasing number of competitive layer PEs, gradually approaching corresponding test set RMSE values. Test set RMSE values are best when there are roughly half as many competitive layer PEs as data set members. Leave-one-out

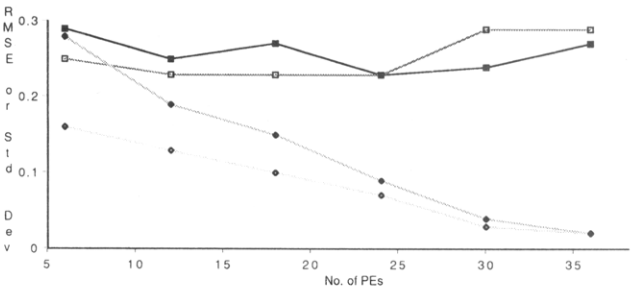

Figure 3. Variation of RMSE values and corresponding standard deviations for leave-one-out counter-propagation networks with number of competitive layer PEs for carboquinone chronic OD data. See Figure 2 for explanation. Number of test/training sets is 37.

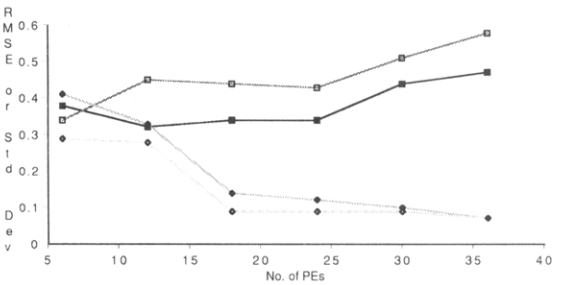

Figure 4. Variation of RMSE values and corresponding standard deviations for leave-one-out counter-propagation networks with number of competitive layer PEs for carboquinone singular OD data. See Figure 2 for explanation. Number of test/training sets is 37.

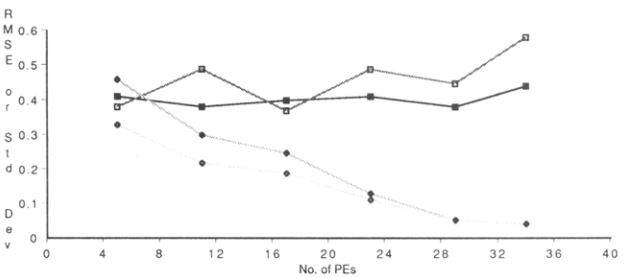

Figure 5. Variation of RMSE values and corresponding standard deviations for leave-one-out counter-propagation networks with number of competitive layer PEs for carboquinone singular MED Data. See Figure 2 for explanation. Number of test/training sets is 35.

RMSE values for these three data sets have not been reported in the literature for back-propagation or multiregression calculations. Therefore, for purposes of comparison, these calculations were performed. The results are summarized in Table 4. It is apparent that counter-propagation networks (perhaps with other than the default number of hidden layer PEs) and back-propagation networks have very similar predictive abilities but that multiple regression is better.

Leave-one-out calculations completely analogous to those just described were performed using counter- and back-propagation for all of the carboquinone data sets using six physicochemical parameters and their squares as input. The results are not shown here since they are extremely similar qualitatively and quantitatively to those just described.

b. Benzodiazepines. Table 5 shows a comparison of neural network and multiregression training set results for the benzodiazepine anti-pentylenetetrazole data set. The counter-propagation networks are default networks, i.e., networks with one competitive layer PE for each member of the data set (57 in this case) and use seven physicochemical parameters and their squares as input. (Calculations with

Table 5. Comparison of Default Counter-Propagation, Back-Propagation, and Multiple Regression Training Set Results for the Benzodiazepine anti-Pentylenetetrazole Data Set[f]

no.[a]	obsd[b]	CP[c]	BP[d]	MR[e]	no.[a]	obsd[b]	CP[c]	BP[d]	MR[e]
1	4.99	4.99	4.88	4.76	31	5.85	5.97	6.22	5.54
2	3.33	3.33	3.45	6.40	32	4.90	5.35	5.04	4.99
3	3.57	3.57	3.87	4.00	33	5.30	5.30	5.46	5.04
4	5.83	5.35	5.01	4.95	34	5.70	5.70	5.84	5.76
5	3.79	3.79	3.86	3.50	35	4.65	4.93	5.12	4.84
6	4.80	4.80	5.03	5.02	36	5.63	5.63	5.82	5.75
7	3.84	3.84	3.99	3.85	37	4.90	5.11	5.23	5.28
8	4.60	4.60	4.64	5.30	38	3.60	4.14	4.27	3.64
9	3.38	3.38	4.71	4.56	39	5.28	5.28	5.12	5.10
10	4.34	4.34	5.08	5.05	40	5.77	5.97	6.30	5.59
11	5.29	5.29	5.27	5.21	41	6.46	6.46	6.11	5.58
12	5.06	5.06	4.54	4.39	42	5.71	5.71	5.54	5.58
13	5.35	5.35	4.88	4.77	43	4.76	4.76	4.94	4.85
14	4.53	4.53	4.36	6.06	44	5.35	5.35	5.13	5.13
15	4.64	4.64	4.70	4.54	45	6.03	6.03	5.93	6.13
16	4.73	4.73	4.86	4.74	46	5.31	5.11	5.31	5.46
17	4.76	4.76	5.19	4.27	47	6.92	6.71	6.73	6.39
18					48	4.97	4.97	5.24	5.61
19	5.06	5.06	5.09	5.06	49	5.20	4.93	4.82	4.74
20	4.15	4.15	3.95	5.24	50	5.44	5.44	5.52	5.63
21					51	5.60	5.60	5.65	5.13
22	5.07	5.07	5.13	5.13	52	5.62	5.62	5.99	5.72
23					53	5.48	5.48	5.38	4.99
24	4.08	4.08	4.34	4.61	54	5.88	5.88	5.81	5.46
25	4.57	4.57	5.12	4.81	55	5.03	5.03	5.21	5.24
26	4.69	4.14	4.56	4.43	56	6.30	6.30	6.36	5.80
27	5.38	5.38	5.19	5.09	57	5.77	5.77	6.00	5.73
28	5.82	5.82	6.10	5.10	58	5.97	5.97	6.01	6.23
29	4.28	4.28	4.57	4.82	59	6.00	6.00	6.01	5.84
30	4.70	4.70	5.00	5.00	60	6.50	6.71	6.55	6.42

[a] Refers to compound number as given in ref 2. A blank entry means no value was given in this reference. [b] Drug activities from refs 2–5. [c] Default counter-propagation results. [d] Back-propagation results from ref 2. [e] Multiple regression results from ref 2, 3, and 5. [f] Seven physicochemical parameters and their squares used as input.

only seven physicochemical parameters as input were done even though such calculations were not reported in ref 2. The results of these calculations are not given here since all of the results are extremely similar both qualitatively and quantitatively to those using 14 input parameters.) The counter-propagation results are again better than either back-propagation or multiregression (MR) results, although not as good as in the carboquinone case. (This is also reflected in Table 6 which gives a comparison of MD and RMSE values for all of the benzodiazepine training data sets.) The counter-propagation network with seven physicochemical parameters and their squares as input (CP) fails to learn 12 activities. The largest error is 0.55 (compound 26). The analogous back-propagation network (BP) learns one activity correctly and exhibits a largest absolute error of 1.33 (compound 9). The CP network gives better training results than the BP network for all compounds except 26, 32, 46, 47, and 60 and better results than MR for all compounds except 26, 32, 35, 38, 40, 46, and 60. The MR also learns one activity correctly and has a largest absolute error of 3.07 (compound 2); compounds 14, 9, and 20 have absolute errors of 1.53, 1.18, and 1.09, respectively.

Table 6 gives a comparison of MD and RMSE values for all of the benzodiazepine training data sets. The CP14 results are better than both the BP14 and MR results. All of the counter-propagation training errors reflected in Tables 5 and 6 are the result of some PEs in the competitive layer taking responsibility for learning about two or more drug activities instead of just one. This gives rise to ambiguous output. This is an example of the problem discussed in section II

Table 6. Comparison of Benzodiazepine Training Set Mean Deviation (MD) and Root Mean Square Error (RMSE) for Default Counter-Propagation, Back-Propagation, and Multiple Regression

	anti-pent[a]			anti-fight[b]			clined scr[c]		
	CP14[d]	BP14[e]	MR	CP14	BP14	MR	CP14	BP14	MR
MD	0.07	0.24	0.39	0.07	0.16	0.30	0.07	0.25	0.33
RMSE	0.16	0.33	0.61	0.15	0.21	0.39	0.15	0.31	0.41

[a] Anti-pentylenetetrazole effect. [b] Anti-fighting effect. [c] Clined screen test effect. [d] Default counter-propagation results with seven physicochemical parameters and their squares as input (14 total input parameters). [e] Default back-propagation results from ref 2; input parameters same as *d*.

between eqs 4 and 5; the same phenomenon was observed for the carboquinones.

Again, a set of leave-one-out calculations was performed to test the usefulness of counter-propagation neural networks in predicting benzodiazepine activities. For the benzodiazepine anti-pent data set 57 disjoint sets, each composed of one benzodiazepine were formed. The 56 benzodiazepines remaining after each test set selection served as the corresponding training set. A given network was trained and tested on each training/testing set pair. The networks were configured in the same manner as described earlier; a default network contained 56 competitive layer PEs. All other details of these calculations are the same as described earlier. The default network results are given in the third column of Table 7. This table is completely analogous to Table 3. Generally, the results are not as good as those for carboquinones; for example, compounds 2, 4, 8, 9, 10, 12, 14, 15, 24, 26, 29, 38, 42, and 45 all exhibit absolute value of differences between predicted and observed values of 0.9 or more; the largest difference is 2.7 (compound 2). Analogous results for back-propagation networks and multiregression are not given in the literature.[2-5] Therefore, for purposes of comparison, back-propagation and multiple regression calculations were performed as described at the end of section II. The results are given in columns six and seven of Table 7. Compared to the carboquinone MED chronic results there appears to be a greater variation in the anti-pent results among the various methods. This is consistent with the fact that the RMSE values and standard deviations for the three methods are not as similar as those for the MED chronic results (see above and Table 8). Back-propagation performs relatively poorly for compounds 2, 54, 14, 60, 26, 9, 38, 8, 29, 5, 41, 12, and 4, for which the absolute value of differences between predicted and observed values are 2.64, 2.56, 2.08, 1.72, 1.65, 1.46, 1.40, 1.28, 1.28, 1.17, 1.08, 1.02, and 1.01, respectively. Multiple regression performs relatively poorly for compounds 2, 28, 9, 41, 4, 24, and 47, for which the absolute value of differences between predicted and observed values are 2.74, 1.58, 1.40, 1.20, 1.01, 0.96, and 0.92. The fact that back-propagation performs poorly is reflected in its large RMSE value and especially the large corresponding standard deviation.

For default counter-propagation networks an RMSE value of 0.59 for the test sets was obtained by calculating the RMSE for each test set and then averaging over the 57 test sets. The standard deviation of RMSE values over the 57 test sets is 0.51. For back-propagation networks the averaged test set RMSE value is 0.53 with a standard deviation of 0.63. For multiple regression the averaged test set RMSE value is 0.47 with a standard deviation of 0.44. These values indicate that the predictive power of multiple regression is

Table 7. Comparison of Counter-Propagation Leave-One-Out Test Set Results for the Benzodiazepine Anti-Pentylenetetrazole Data Set[h]

no.[a]	obsd[b]	CP[c]	CP20[d]	CP32[e]	BP[f]	MR[g]
1	4.99	5.04	4.94	5.04	4.72	4.75
2	3.33	6.03	4.53	4.53	5.97	6.07
3	3.57	3.99	4.64	3.99	3.94	4.33
4	5.83	4.90	5.01	5.01	4.82	4.82
5	3.79	3.85	4.46	3.86	2.62	4.16
6	4.80	5.00	5.21	5.00	4.97	5.36
7	3.84	4.15	4.61	3.86	3.74	3.76
8	4.60	6.46	5.43	6.13	5.88	5.20
9	3.38	4.85	4.66	4.85	4.84	4.78
10	4.34	5.28	5.26	5.18	5.04	4.92
11	5.29	4.80	4.75	4.75	5.13	5.40
12	5.06	3.99	4.01	4.62	4.04	4.51
13	5.35	4.86	4.58	4.83	4.72	4.74
14	4.53	3.33	5.07	3.33	2.45	5.30
15	4.64	3.38	4.23	4.21	4.62	4.66
16	4.73	5.17	5.17	5.02	4.76	4.75
17	4.76	5.38	5.38	5.38	5.43	5.24
18						
19	5.06	5.35	5.10	5.35	5.11	4.98
20	4.15	3.70	3.70	3.57	3.78	4.34
21						
22	5.07	5.35	5.18	5.19	5.08	4.94
23						
24	4.08	5.45	5.45	5.30	4.82	5.04
25	4.57	4.65	5.49	5.12	5.24	4.85
26	4.69	3.60	4.92	3.60	3.04	3.84
27	5.38	4.76	4.76	4.76	4.65	4.99
28	5.82	6.03	4.96	6.03	6.07	4.24
29	4.28	3.38	4.02	4.32	5.56	3.64
30	4.70	4.80	5.43	5.04	5.00	5.36
31	5.85	6.09	5.31	6.09	6.38	5.29
32	4.90	5.18	5.09	5.44	5.08	4.87
33	5.30	4.84	4.83	4.82	5.11	5.16
34	5.70	5.97	5.97	5.97	5.71	5.85
35	4.65	5.20	5.50	5.09	5.32	4.89
36	5.63	6.00	6.15	6.00	5.80	5.68
37	4.90	5.31	5.21	5.31	5.24	4.70
38	3.60	4.69	4.69	4.69	5.00	4.27
39	5.28	4.93	5.13	5.15	5.02	4.60
40	5.77	5.85	5.27	5.65	6.26	5.33
41	6.46	6.09	5.18	4.60	5.38	5.26
42	5.71	6.92	5.70	5.62	5.38	6.12
43	4.76	4.73	5.02	5.02	4.87	4.56
44	5.35	4.91	5.09	5.12	5.05	4.92
45	6.03	4.57	5.85	4.62	6.17	5.53
46	5.31	4.90	5.08	4.90	5.28	5.08
47	6.92	6.50	6.12	6.50	6.79	6.00
48	4.97	5.70	5.84	5.20	5.51	5.46
49	5.20	4.65	5.36	4.91	4.64	4.81
50	5.44	5.70	5.21	5.10	4.90	5.41
51	5.60	5.30	4.85	4.69	5.61	5.38
52	5.62	5.60	5.49	5.64	5.97	5.64
53	5.48	4.65	4.83	4.92	5.37	4.93
54	5.88	5.10	5.11	5.10	3.32	5.04
55	5.03	5.23	5.04	5.01	5.33	5.07
56	6.30	6.00	5.78	5.82	6.31	5.83
57	5.77	5.62	5.53	5.59	5.88	5.63
58	5.97	5.70	5.17	5.70	5.80	6.21
59	6.00	6.30	5.82	5.97	6.40	5.89
60	6.50	6.92	5.93	5.13	4.78	6.08

[a] Refers to compound number as given in ref 2. A blank entry means no value was given in this reference. [b] Drug activities from refs 2–5. [c] Default (56 competitive layer PEs) counter-propagation results. [d] Twenty competitive layer PE counter-propagation results. [e] Thirty-two competitive layer PE counter-propagation results. [f] Back-propagation results. [g] Multiple regression results. [h] Seven physicochemical parameters and their squares used as input.

better than that of back-propagation and the default counter-propagation network. As with the carboquinones there are two concerns which arise when comparing the above counter-propagation results to those of back-propagation. The first

Table 8. Comparison of Benzodiazepine Leave-One-Out Root Mean Square Error (RMSE) and Standard Deviations (STD) for Counter-Propagation, Back-Propagation, and Multiple Regression[a]

	anti-pent				
	CP14	BP14	CP20[b]	CP32[b]	MR
RMSE	0.59	0.53	0.56	0.53	0.47
STD	0.51	0.63	0.36	0.44	0.44
	anti-fight				
	CP14	BP14	CP41[b]	CP8[b]	MR
RMSE	0.47	0.38	0.44	0.45	0.40
STD	0.35	0.26	0.32	0.30	0.31
	clined scr				
	CP14	BP14	CP19[b]	CP30[b]	MR
RMSE	0.52	0.41	0.43	0.44	0.41
STD	0.37	0.37	0.34	0.31	0.37

[a] See footnotes in Table 6 for explanation. [b] A counter-propagation network with either 20, 32, 41, 8, 19, or 30 hidden layer PEs.

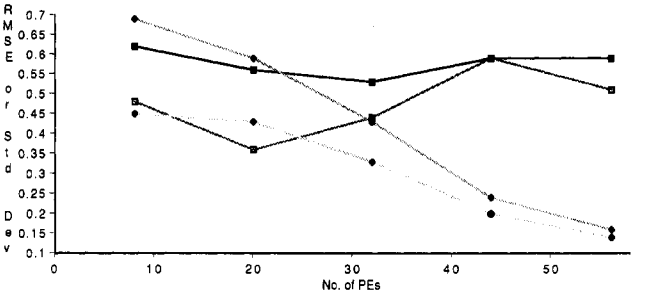

Figure 6. Variation of RMSE values and corresponding standard deviations for leave-one-out counter-propagation networks with number of competitive layer PEs for benzodiazepine anti-pentylenetetrazole data. See Figure 2 for explanation. Number of test/training sets is 57.

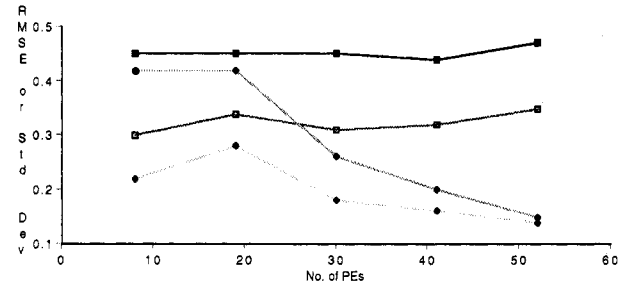

Figure 7. Variation of RMSE values and corresponding standard deviations for leave-one-out counter-propagation networks with number of competitive layer PEs for benzodiazepine anti-fighting data. See Figure 2 for explanation. Number of test/training sets is 53.

is that the counter-propagation test set RMSE is a factor of 3.7 times greater than the training set RMSE (it was a factor of 4.1 for the carboquinone MED chronic data). This may raise the question of the robustness of counter-propagation networks. In fact, this is not an issue for such networks; a detailed discussion of this point is given on page 383 of ref 1 and will not be repeated here. The second concern is that the default counter-propagation networks do not perform as well as back-propagation networks, although in this case the difference is very small (RMSE of 0.59 for counter-propagation test sets compared to 0.53 for back-propagation).

Figures 6–8 show RMSE values and corresponding standard deviations of leave-one-out calculations for the three benzodiazepine data sets using counter-propagation networks with differing numbers of competitive layer PEs. Seven

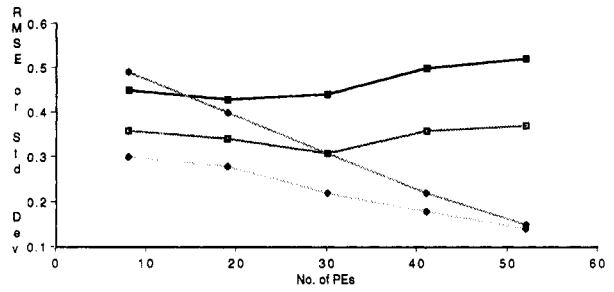

Figure 8. Variation of RMSE values and corresponding standard deviations for leave-one-out counter-propagation networks with number of competitive layer PEs for benzodiazepine clined screen data. See Figure 2 for explanation. Number of test/training sets is 53.

physicochemical parameters and their squares were used as input. Since the trends in RMSE values with number of PEs is the same for these data sets as described above for chronic MED data, results are shown only for every 11th PE (12th for the anti-pent data). Again, the training set RMSE values increase with a decreasing number of competitive layer PEs, gradually approaching corresponding test set RMSE values. Test set RMSE values are best when there are roughly half as many competitive layer PEs as data set members. Leave-one-out RMSE values for the anti-fighting and clined screen data sets have not been reported in the literature for back-propagation or multiregression calculations. For purposes of comparison these calculations were performed. The results are summarized in Table 8. It is apparent that, in contrast to the carboquinone case, neural networks and multiple regression all have similar predictive capabilities.

The counter-propagation network test set results for 20 and 32 competitive layer PEs are given in the fourth and fifth columns of Table VII. Generally, the results are better than those of the default network; for example, for the 20 PE case compounds 2, 3, 9, 10, 12, 24, 25, and 38 exhibit absolute value of differences between predicted and observed values of 0.9 or more; the largest difference is 1.37 (compound 24). For the 32 PE case compounds 2, 8, 9, 14, 24, 26, 38, 41, 45, and 51 exhibit differences of 0.9 or more; the largest difference is 1.86 (compound 41). Relative to the default network these errors are generally smaller. Training set results for the 20 and 32 PE networks are qualitatively very similar to those given in Table 5 for the default network and are therefore not shown here. The increase in RMSE values for these two nondefault networks over that for the default network again arises primarily from their inability to correctly learn as many activities. This is to be expected since some of the fewer competitive layer PEs are forced to take responsibility for more than one drug activity.

IV. SUMMARY AND CONCLUSIONS

This paper describes attempts to predict activities of a set of carboquinones and of a set of benzodiazepines from physicochemical parameters using counter-propagation neural networks. For carboquinones, networks containing one hidden layer PE for each compound were able to achieve very small RMSE values (ranging from 0.02 to 0.08) for training sets and relatively good RMSE values (ranging from 0.27 to 0.47) for test sets. These training results are significantly better than corresponding back-propagation and multiregression results from the literature; these test set results are as good as or slightly worse than back-propagation results. These one PE per carboquinone test set results were improved by 10−15% using networks containing fewer hidden layer PEs than there were carboquinones. This attempts to force the network to recognize statistical trends of the entire data set, rather than focusing on each input vector of the data set individually. In general the smallest test set RMSE values occur when there are about half as many competitive layer PEs as there are compounds in the data set; under this condition test set RMSE values are between 0 and 10% better than back-propagation RMSE values and are about 1.3 times greater than corresponding training set RMSE values (this is comparable to reported back-propagation results). Training set RMSE values increase with decreasing number of competitive layer PEs and approach those of test sets under this condition, perhaps indicating that these networks are more robust and that they are more successful in learning statistical trends of the data.

For benzodiazepines, all of the qualitative observations in the above paragraph hold. All RMSE values are larger than for carboquinones, however. Networks containing one hidden layer PE for each compound were able to achieve very small RMSE values (ranging from 0.15 to 0.16) for training sets and relatively good RMSE values (ranging from 0.47 to 0.59) for test sets. These training results are significantly better than corresponding back-propagation and multiregression results from the literature; these test set results are slightly worse than previously reported back-propagation results. These one PE per benzodiazepine results were improved by 10−15% using networks containing fewer hidden layer PEs than there were benzodiazepines.

Interestingly, multiple regression results for all carboquinone data sets have essentially the same test set RMSE values as for the neural network results; however the corresponding test set standard deviations for multiple regression are smaller than those for neural networks. In this sense multiple regression appears to be the preferred method for this data. For all benzodiazepine data sets multiple regression test set RMSE values and their corresponding standard deviations are similar to neural network RMSE values and standard deviations. It therefore appears that for data which is quite linear (carboquinones^2) multiple regression is preferred, while for data which is significantly nonlinear (benzodiazepines^2) neural networks are preferred. This observation is currently being investigated further.

REFERENCES AND NOTES

(1) Peterson, K. L. Counter-Propagation Neural Networks in the Modeling and Prediction of Kovats Indices for Substituted Phenols. *Anal. Chem.* **1992**, *64*, 379−386.

(2) Aoyama, T.; Suzuki, Y.; Ichikawa, H. Neural Networks Applied to Quantitative Structure-Activity Relationship Analysis. *J. Med. Chem.* **1990**, *33*, 2583−2590.

(3) Yoshimoto, M.; Miyazawa, H.; Nakao, H.; Shinkai, K.; Arakawa, M. Quantitative Structure-Activity Relationships in 2,5-Bis(1-aziridinyl)-*p*-benzoquinone Derivatives Against Leukemia L-1210. *J. Med. Chem.* **1979**, *22*, 491−496.

(4) Aoyama, T.; Ichikawa, H. Obtaining the Correlation Indices Between Drug Activity and Structural Parameters Using a Neural Network. *Chem. Pharm. Bull.* **1991**, *39*, 372−378.

(5) Kubota, T.; Yamakawa, M.; Terada, H.; Yoshimoto, M. In Structure−Activity Relationships−Quantitative Approaches; the QSAR Research Group, Ed.; *Kagaku no Ryoichi Suppl. Ed., No. 122*, 1979.

(6) Gorman, P. R.; Sejnowski, T. J. Analysis of Hidden Units in a Layered Network Trained to Classify Sonar Targets. *Neural Networks* **1988**, *1*, 75−89.

CI930043H