# Genetic Neural Networks for Quantitative Structure−Activity Relationships: Improvements and Application of Benzodiazepine Affinity for Benzodiazepine/ $GABA_A$ Receptors

Sung-Sau So*,† and Martin Karplus*,†,‡

*Department of Chemistry and Chemical Biology, Harvard University, 12 Oxford Street, Cambridge, Massachusetts 02138, and Laboratoire de Chimie Biophysique, Institut le Bel, Université Louis Pasteur, 4 rue Blaise Pascal, 67000 Strasbourg, France*

*Received July 25, 1996*[⊗]

A novel tool, called a genetic neural network (GNN), has been developed for obtaining quantitative structure−activity relationships (QSAR) for high-dimensional data sets (*J. Med. Chem.* **1996**, *39*, 1521−1530). The GNN method uses a neural network to correlate activity with descriptors that are preselected by a genetic algorithm. To provide an extended test of the GNN method, the data on 57 benzodiazepines given by Maddalena and Johnston (MJ; *J. Med. Chem.* **1995**, *38*, 715−724) have been examined with an enhanced version of GNN, and the results are compared with the excellent QSAR of MJ. The problematic steepest descent training has been replaced by the scaled conjugate gradient algorithm. This leads to a substantial gain in performance in both robustness of prediction and speed of computation. The cross-validation GNN simulation and the subsequent run based on an unbiased and more efficient protocol led to the discovery of other 10-descriptor QSARs that are superior to the best model of MJ based on backward elimination selection and neural network training. Results from a series of GNNs with a different number of inputs showed that a neural network with fewer inputs can produce QSARs as good as or even better than those with higher dimensions. The top-ranking models from a GNN simulation using only six input descriptors are presented, and the chemical significance of the chosen descriptors is discussed. The statistical significance of these GNN QSARs is validated. The best QSARs are used to provide a graphical tool that aids the design of new drug analogues. By replacing functional groups at the 7- and 2′-positions with ones that have optimal substituent parameters, a number of new benzodiazepines with high potency are predicted.

## Introduction

A new method for quantitative structure−activity relationships (QSAR), called genetic neural networks (GNN), has been described recently.[1] In GNN, selections of descriptors are made using a genetic algorithm (GA). This facilitates rapid discovery of optimal solutions through its evolutionary and parallel features. Correlations of biological activities with these descriptors are performed by a neural network, whose nonlinear attribute has an advantage in modeling complex relationships. In the published study, we showed that GNN is superior in one application (the Selwood data set[2]) to genetic multiple linear regressions[3,4] and other methods,[2,5] one of which makes use of neural networks[5] but not a GA for property selection. In this paper we apply the GNN method to a series of benzodiazepines (BZs) which have played an important role in medicinal chemistry over 3 decades[6−14] and have been used recently as a nucleus for solid-state combinatorial synthesis.[15,16] They are of special interest because of the putative role of BZs in the mammalian central nervous system, which derives from their interaction with a macromolecular supercomplex involving their receptor and the $GABA_A$ receptor. The action of GABA, a major inhibitory neurotransmitter, has been shown to be modulated upon the binding of various BZs. Because of this regulatory role and the exceptional tolerance to BZs even at high doses, they have become an important class of therapeutic agents with widespread applications in treatments related to anxiety and emotional disorders.[8,10]

A GNN study of the BZs is an excellent test application of the GNN method because a wealth of chemical and pharmacological information is available on this series of compounds, and they have been used in many QSAR and SAR studies.[17−20] Maddalena and Johnston (MJ) obtained excellent results in a recent QSAR study.[20] They considered a homogeneous set of 57 classical 1,4-benzodiazepin-2-ones (1,4-BZs) that had well-defined binding affinities (log $IC_{50}$) for the BZ/$GABA_A$ receptor in a competition assay against tritiated diazepam.[6,10] The substitution patterns of these compounds were described by several commonly used substituent constants at each of the six variable positions (Figure 1).[21] A backward elimination strategy was applied to discard descriptors that were nonessential, and a steepest descent (SD) back-propagation neural network was used to derive QSAR models. The final 10-descriptor QSAR proposed by MJ had high values of correlation coefficients for both training (0.938) and cross-validation (0.896). The availability of the BZ data and their analysis provide the opportunity to test the GNN methodology because this data set is considerably larger than the one used in the initial application.[1,2] Further, the QSAR results of MJ are of high quality, making possible a stringent comparison. One question is whether the selection of descriptors made by MJ is optimal.

† Harvard University.
‡ Université Louis Pasteur.
[⊗] Abstract published in *Advance ACS Abstracts,* November 15, 1996.

S0022-2623(96)00536-5 CCC: $12.00 

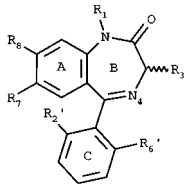

**Figure 1.** 1,4-Benzodiazepin-2-ones considered in this study. The six substituent positions are $R_7$ and $R_8$ in the A-ring, $R_1$ and $R_3$ in the B-ring, and $R_{2'}$ and $R_{6'}$ in the C-ring.

In addition to the BZ application, we present a methodological improvement of the neural network simulator that overcomes the problem associated with multiple local minima. We also propose a more efficient GNN protocol that is used to obtain models with high predictivity. Some new and simpler QSAR models for the BZs are suggested, and the reliability of such models is validated with a number of statistical tools. New BZ analogues are designed by optimizing the values of chosen input descriptors, and the predicted activities of the new compounds are determined by the refined GNN QSARs.

## Methods

**Data Set.** The 57 BZs used in the study of MJ were examined in the present work (Table 1). These BZs contain six variable substituent positions (Figure 1), though the data are extensive for only three positions ($R_7$, $R_1$, and $R_{2'}$). Each substituent is parameterized by seven physicochemical descriptors. The values of dipole

**Table 1.** Data Set[a 20]

| ID | name | $R_7$ | $R_1$ | $R_{2'}$ | $R_{6'}$ | $R_3$ | $R_8$ | log $IC_{50}$ |
|---|---|---|---|---|---|---|---|---|
| **1** | Ro 05-3061 | F | H | H | H | H | H | 1.602 |
| **2** | Ro 05-4865 | F | $CH_3$ | H | H | H | H | 1.230 |
| **3** | Ro 05-6820 | F | H | F | H | H | H | 0.869 |
| **4** | Ro 05-6822 | F | $CH_3$ | F | H | H | H | 0.708 |
| **5** | nordazepam | Cl | H | H | H | H | H | 0.973 |
| **6** | diazepam | Cl | $CH_3$ | H | H | H | H | 0.908 |
| **7** | Ro 05-3367 | Cl | H | F | H | H | H | 0.301 |
| **8** | delorazepam | Cl | H | Cl | H | H | H | 0.255 |
| **9** | Ro 07-9957 | I | $CH_3$ | F | H | H | H | 0.462 |
| **10** | Ro 05-2904 | $CF_3$ | H | H | H | H | H | 1.114 |
| **11** | Ro 14-3074 | $N_3$ | H | F | H | H | H | 0.724 |
| **12** | nitrazepam | $NO_2$ | H | H | H | H | H | 1.000 |
| **13** | Ro 05-4435 | $NO_2$ | H | F | H | H | H | 0.176 |
| **14** | flunitrazepam | $NO_2$ | $CH_3$ | F | H | H | H | 0.580 |
| **15** | clonazepam | $NO_2$ | H | Cl | H | H | H | 0.255 |
| **16** | Ro 05-4082 | $NO_2$ | $CH_3$ | Cl | H | H | H | 0.342 |
| **17** | Ro 05-3590 | $NO_2$ | H | $CF_3$ | H | H | H | 0.544 |
| **18** | Ro 20-7736 | NHOH | $CH_3$ | F | H | H | H | 1.982 |
| **19** | Ro 05-3072 | $NH_2$ | H | H | H | H | H | 2.587 |
| **20** | Ro 05-4318 | $NH_2$ | $CH_3$ | H | H | H | H | 2.663 |
| **21** | Ro 20-1815 | $NH_2$ | $CH_3$ | F | H | H | H | 1.813 |
| **22** | Ro 05-4619 | $NH_2$ | H | Cl | H | H | H | 1.875 |
| **23** | Ro 05-4528 | CN | $CH_3$ | H | H | H | H | 2.580 |
| **24** | Ro 20-2541 | CN | $CH_3$ | F | H | H | H | 1.477 |
| **25** | Ro 20-2533 | $CH_2CH_3$ | H | H | H | H | H | 1.556 |
| **26** | Ro 20-5747 | $CH{=}CH_2$ | H | H | H | H | H | 1.380 |
| **27** | Ro 20-5397 | CHO | H | H | H | H | H | 1.633 |
| **28** | Ro 20-3053 | $COCH_3$ | H | F | H | H | H | 1.255 |
| **29** | Ro 05-2921 | H | H | H | H | H | H | 2.544 |
| **30** | Ro 05-4336 | H | H | F | H | H | H | 1.322 |
| **31** | Ro 05-4520 | H | $CH_3$ | F | H | H | H | 1.146 |
| **32** | Ro 05-4608 | H | $CH_3$ | Cl | H | H | H | 0.580 |
| **33** | halazepam | Cl | $CH_{2CF3}$ | H | H | H | H | 1.964 |
| **34** | Ro 06-9098 | $NO_2$ | $CH_{2OCH3}$ | H | H | H | H | 2.633 |
| **35** | Ro 20-1310 | Cl | $C(CH_{3)3}$ | H | H | H | H | 2.792 |
| **36** | Ro 07-2750 | Cl | $(CH_{2)2OH}$ | F | H | H | H | 1.389 |
| **37** | Ro 22-4683 | $NO_2$ | $C(CH_{3)3}$ | F | H | H | H | 2.477 |
| **38** | Ro 07-4419 | H | H | F | F | H | H | 1.279 |
| **39** | Ro 07-3953 | Cl | H | F | F | H | H | 0.204 |
| **40** | Ro 07-4065 | Cl | $CH_3$ | F | F | H | H | 0.613 |
| **41** | Ro 07-5193 | Cl | H | Cl | F | H | H | 0.477 |
| **42** | Ro 22-3294 | Cl | H | Cl | Cl | H | H | 0.845 |
| **43** | Ro 07-5220 | Cl | $CH_3$ | Cl | Cl | H | H | 0.740 |
| **44** | Ro 13-3780 | Br | $CH_3$ | F | F | H | H | 0.380 |
| **45** | Ro 11-4878 | Cl | H | F | H | (*S*)-$CH_3$ | H | 0.544 |
| **46** | meclonazepam | $NO_2$ | H | Cl | H | (*S*)-$CH_3$ | H | 0.079 |
| **47** | Ro 11-6896 | $NO_2$ | $CH_3$ | F | H | (*S*)-$CH_3$ | H | 0.845 |
| **48** | Ro 06-7263 | Cl | Cl | H | H | rac $CH_3$ | H | 1.690 |
| **49** | oxazepam | Cl | H | H | H | rac OH | H | 1.255 |
| **50** | temazepam | Cl | $CH_3$ | H | H | rac OH | H | 1.204 |
| **51** | lorazepam | Cl | H | Cl | H | rac OH | H | 0.544 |
| **52** | Ro 20-7078 | Cl | H | F | H | rac Cl | H | 0.724 |
| **53** | Ro 07-6198 | H | H | F | F | H | Cl | 1.447 |
| **54** | Ro 20-8895 | H | H | F | H | H | $CH_3$ | 1.279 |
| **55** | Ro 22-6762 | Cl | $CH_3$ | H | H | H | Cl | 1.602 |
| **56** | Ro 20-8065 | Cl | H | F | H | H | Cl | 0.556 |
| **57** | Ro 20-8552 | $CH_3$ | H | F | H | H | Cl | 1.146 |

[a] See Figure 1 for the substitution positions

**Table 2.** Substituent Constants[a]

| substituent | $\mu$ | $\pi$ | MR | F | R | $\sigma_m$ | $\sigma_p$ |
|---|---|---|---|---|---|---|---|
| Br | −1.57 | 0.86 | 8.88 | 0.44 | −0.17 | 0.39 | 0.23 |
| $C(CH_3)_3$ | 0.52 | 1.98 | 19.62 | −0.07 | −0.13 | −0.10 | −0.20 |
| $CF_3$ | −2.61 | 0.88 | 5.02 | 0.38 | 0.19 | 0.43 | 0.54 |
| $CH{=}CH_2$ | 0.20 | 0.82 | 10.99 | 0.07 | −0.08 | 0.05 | −0.02 |
| $CH_2CF_3$ | −2.07 | 1.34 | 9.64 | 0.34 | 0.09 | 0.40 | 0.50 |
| $CH_2CH_2OH$ | −0.60 | −0.31 | 12.10 | 0.01 | −0.29 | −0.05 | −0.23 |
| $CH_2OCH_3$ | −1.01 | −0.78 | 12.07 | 0.01 | 0.02 | 0.02 | 0.03 |
| CHO | −3.02 | −0.65 | 6.88 | 0.31 | 0.13 | 0.35 | 0.42 |
| Cl | −1.59 | 0.71 | 6.03 | 0.41 | −0.15 | 0.37 | 0.23 |
| CN | −4.08 | −0.57 | 6.33 | 0.51 | 0.19 | 0.56 | 0.66 |
| $COCH_3$ | −2.90 | −0.55 | 11.18 | 0.32 | 0.20 | 0.38 | 0.50 |
| $CH_2CH_3$ | 0.39 | 1.02 | 10.30 | −0.05 | −0.10 | −0.07 | −0.15 |
| F | −1.43 | 0.14 | 0.92 | 0.43 | −0.34 | 0.34 | 0.06 |
| H | 0.00 | 0.00 | 1.03 | 0.00 | 0.00 | 0.00 | 0.00 |
| I | −1.36 | 1.12 | 13.94 | 0.40 | −0.19 | 0.35 | 0.18 |
| $CH_3$ | 0.36 | 0.56 | 5.65 | −0.04 | −0.13 | −0.07 | −0.17 |
| $N_3$ | −1.56 | 0.46 | 10.20 | 0.30 | −0.13 | 0.27 | 0.15 |
| $NH_2$ | 1.53 | −1.23 | 5.42 | 0.02 | −0.68 | −0.16 | −0.66 |
| $NHCOCH_3$ | −3.65 | −0.97 | 14.93 | 0.28 | −0.26 | 0.21 | 0.00 |
| NHOH | −0.14 | −1.34 | 7.22 | 0.06 | −0.40 | −0.04 | −0.34 |
| $NO_2$ | −4.13 | −0.28 | 7.36 | 0.67 | 0.16 | 0.71 | 0.78 |
| OH | −1.59 | −0.67 | 2.85 | 0.29 | −0.64 | 0.12 | −0.37 |
| $OCF_3$ | −2.36 | 1.04 | 7.86 | 0.38 | 0.00 | 0.38 | 0.35 |
| $SO_2F$ | −4.59 | 0.05 | 8.65 | 0.75 | 0.22 | 0.80 | 0.91 |

[a] The first 22 substituents are taken from Table 2 in MJ.[20] The substituent constants of the final two substituents that are used in the new compounds are taken from refs **22** and **23**.

**Table 3.** Correlation ($R^2$) Matrix for the Seven Physicochemical Parameters Listed in Table 2

| | $\mu$ | $\pi$ | MR | F | R | $\sigma_m$ | $\sigma_p$ |
|---|---|---|---|---|---|---|---|
| m | 1 | | | | | | |
| p | 0.03 | 1 | | | | | |
| MR | 0.00 | 0.09 | 1 | | | | |
| F | 0.74 | 0.00 | 0.04 | 1 | | | |
| R | 0.35 | 0.07 | 0.02 | 0.17 | 1 | | |
| $\sigma_m$ | 0.81 | 0.01 | 0.02 | 0.94 | 0.37 | 1 | |
| $\sigma_p$ | 0.73 | 0.03 | 0.00 | 0.66 | 0.75 | 0.86 | 1 |

moment ($\mu$), lipophilicity ($\pi$), molar refractivity (MR), polar constant ($F$), resonance constant ($R$), and Hammett meta constant ($\sigma_m$) and para constant ($\sigma_p$) for the functional groups present in the data set are listed in Table 2. The values are taken from MJ and refs 22 and 23. Table 3 is a correlation matrix which shows the pairwise correlation coefficients between the descriptors based on the variation within the substituents that are listed in Table 2. It indicates that the hydrophobic $\pi$ and the steric MR are independent but the remaining five electronic descriptors are somewhat correlated. A complete cross-correlation matrix for the QSAR parameters by position of substitution is also available as Supporting Information. The additional random input variables introduced in the study of MJ were not used here, i.e., only the 42 physically meaningful descriptors were considered. The input and output vectors of the data set were scaled to take values between 0.1 and 0.9.

**Neural Network.** The most popular type of neural networks used in QSAR studies have been based on SD back-propagation training[24] because of ease of implementation. However, this kind of algorithm generally has a poor convergence behavior. It is also known that its success often depends on the setting of a number of user-defined parameters such as learning rates and momentum terms.[25] Furthermore, simple back-propagation has a weakness in dealing with multiple local minima, and in some cases satisfactory training cannot be achieved, e.g., neural networks with weights that are initialized by different random seeds can lead to solutions of disparate qualities. One solution to this kind of dependency is to perform multiple simulations with a number of different seeds on an otherwise identical system. Only the most optimized solution, judged in terms either of its training or cross-validated error, is accepted as the final output.

To avoid these problems we have employed an alternative training paradigm that is naturally more deterministic. Recent advances in neural network research indicate that optimization algorithms that include pseudo-second-derivative information are more capable of dealing with local minima. One such algorithm, the scaled conjugate gradient (SCG) method,[26] has been incorporated into the neural network simulator. A recent example of the utility of the SCG method has been given in a neural network prediction study of secondary structure in proteins.[27] Unlike SD algorithms, the two scalar parameters ($\sigma$ and $\lambda_1$) which specify step size in the SCG implementation are not crucial for its success.[26] In the present analysis they were set at $\sigma = 1 \times 10^{-5}$ and $\lambda_1 = 1 \times 10^{-7}$, which are in accord with the ranges recommended by Møller.[26] In all simulations the weights of networks were initialized in the range of ±1.0. Multiple random number seeds were employed for the first SCG benchmark study. Since no dependence on the random seed was found, a single seed was used for all the other applications.

**Genetic Algorithm.** In this study the backward elimination selection strategy used by MJ was replaced by a GA.[28−30] Although two different GAs were tried in the earlier work,[1] only the evolutionary programming (EP) algorithm was employed here to choose the important physicochemical descriptors because previous studies showed that it gave an optimal set of solutions.[1,4] The details of EP implementation are given elsewhere.[1,4] A population of 200 individuals was generated for each of the GNN runs. Single-point mutations were used for a total of 50 reproduction cycles.

**Statistical Measures.** Correlation coefficients are used to give a measure of statistical fit in the QSAR models. $R_t$ is the correlation coefficient between the

**Table 4.** QSAR Models Used in This Study[a]

| | | | $R_7$ | | | | $R_1$ | | | | | $R_{2'}$ | | | | | | $R_{6'}$ | | | | | $R_3$ | | | $R_8$ | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| model | $R_t$ | $R_{cv}$ | p | F | R | $\sigma_m$ | m | p | MR | $\sigma_m$ | R | p | MR | F | R | $\sigma_m$ | $\sigma_p$ | m | p | MR | $\sigma_m$ | $\sigma_p$ | MR | F | $\sigma_m$ | p | MR | F | $\sigma_m$ | $\sigma_p$ |
| MJ 6-1 | 0.826 | 0.776 | ● | | | | | | ● | | | | | ● | | | | | | ● | | | | | ● | | | | | ● |
| 10-1 | 0.938 | 0.896 | ● | ● | | | ● | | ● | | ● | | ● | ● | | | | | | ● | | | | | ● | | | | | ● |
| 10-3 | 0.935 | 0.899 | ● | ● | | | ● | | ● | | ● | | ● | ● | | | | | | ● | | | | | ● | | | | | ● |
| SCG 10-3 | 0.948 | 0.897 | ● | ● | | | ● | | ● | | ● | | ● | ● | | | | | | ● | | | | | ● | | | | | ● |
| C10-3 #1 | 0.970 | 0.939 | ● | ● | ● | | | ● | ● | ● | | | ● | | | ● | | | ● | | | | | ● | | | | ● | |
| T10-3 #1 | 0.969 | 0.938 | ● | | ● | ● | | | ● | | | ● | | | ● | ● | | | ● | | | | ● | | | ● | | | |
| T6-2 #1 | 0.954 | 0.932 | ● | ● | | | | | ● | | | | | | | ● | | | ● | | | | | | | ● | | | |
| #2 | 0.953 | 0.929 | ● | | | ● | | | ● | | | | | | | ● | | | | ● | | | | | | | | | ● | |
| #3 | 0.954 | 0.928 | ● | | | ● | | | ● | | | | | | | ● | ● | | | | | ● | | | | | | | | |
| #4 | 0.953 | 0.928 | ● | | | ● | | | ● | | | | | | | ● | | | ● | | | | ● | | | | | | | |
| #5 | 0.952 | 0.927 | ● | ● | | | | | ● | | | | | | | ● | | | ● | | | ● | | | | | | | | |
| #6 | 0.951 | 0.927 | ● | ● | | | | | ● | | | | | | | ● | | | | | ● | | | | | | ● | | | |
| #7 | 0.956 | 0.926 | ● | | | ● | | | ● | | | | | | | ● | ● | | | ● | | | | | | | | | | |
| #8 | 0.952 | 0.926 | ● | ● | | | | | ● | | | | | | | ● | | ● | | ● | | | | | | | | | | |
| #9 | 0.951 | 0.925 | ● | ● | | | | | ● | | | | | | | ● | ● | | | | | | | ● | | | | | | |
| #10 | 0.954 | 0.925 | ● | ● | | | | | ● | | | | | | | ● | | | ● | | | | | | | ● | | | | |
| C6-2 #1 | 0.954 | 0.932 | ● | ● | | | | | ● | | | | | | | ● | | | ● | | | | | | | | ● | | | |
| #2 | 0.953 | 0.929 | ● | | | ● | | | ● | | | | | | | ● | | | | ● | | | | | | | | | ● | |
| #3 | 0.953 | 0.928 | ● | | | ● | | | ● | | | | | | | ● | | | | ● | | | | | | | ● | | | |
| #4 | 0.954 | 0.928 | ● | | | ● | | | ● | | | | | | | ● | ● | | | | | ● | | | | | | | | |

[a] See Figure 1 for the positions of substitution.

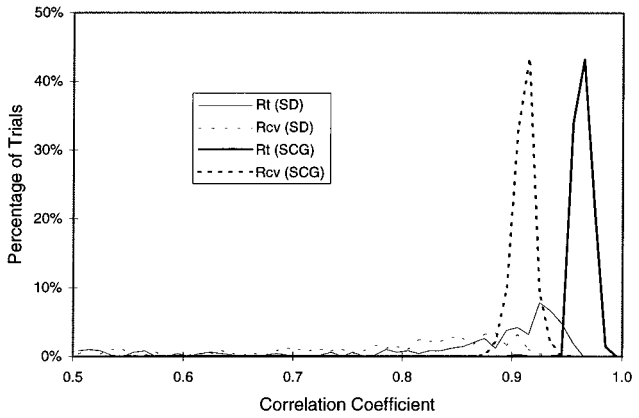


**Figure 2.** Distribution of the values of correlation coefficients for 500 trials of SD and SCG neural network simulation using different random seeds. The sharp peaks corresponding to the SCG set indicate that this type of neural network is more deterministic than its SD-based counterpart (see text).

*calculated* activities and the observed biological activity based on the result of neural network training. $R_{cv}$ is the correlation coefficients between the *predicted* activities and the observed activities based on a leave-one-out (LOO) cross-validation procedure.

## Results

**Comparison of the SD and the SCG Neural Networks.** The most predictive 10-descriptor QSAR model reported by MJ used a set of descriptors derived from a backward elimination scheme and a neural network that contained 10 input, 3 hidden, and 1 output nodes, i.e., a topology of 10-3-1. This model (MJ 10-3, Table 4) was used as a benchmark for the new SCG neural network simulator. Both the SD-optimized neural network used in the original GNN study[1] and the new SCG-based neural network (also with a 10-3-1 topology) were used to formulate QSARs with this selection of descriptors. The SD and SCG calculations were repeated 500 times using different initial random number seeds. The distributions of the resulting $R_t$ and $R_{cv}$ for the two sets of runs are shown in Figure 2. It is evident that the SCG simulator is much superior to the SD-based method. Less than one-half of the SD runs (46%) achieved satisfactory training ($R_t > 0.75$). Furthermore the correlation coefficients from these successful SD-based trials spanned a very wide range which introduces significant uncertainty in the results. This contrasted sharply with the SCG result. Every SCG run led to a low training and cross-validation error. Their $R_t$ values had a high average of 0.954 and, more importantly, a very low standard deviation of 0.007. This was mirrored by a narrow distribution for $R_{cv}$, which had an average of 0.901 and a small standard deviation of 0.009.

The superiority of the SCG-based method is clear. Further, the results are such that the standard deviations for $R_t$ and $R_{cv}$ become a useful measure for determining whether any change in predictivity of a new model is statistically significant. Also, the consistency of the runs with different seeds shows that a solution provided by a single SCG simulation could be used directly without the need for multiple neural network runs. Figure 3a shows the averages and the standard deviations of *calculated* activities for the 57 BZs from the above 500 SCG training runs, and Figure 3b shows the *predicted* activities from cross-validations. The standard deviations of these predictions were small (on average 0.06 for training and 0.08 for cross-validation).

Another advantage of the SCG algorithm was its speed. The relative convergence behavior of the two algorithms is shown in Figure 4. It suggests that the speedup of SCG relative to SD is approximately 2-fold in this system.

As a control, the original $R_{cv}$ values of MJ were reproduced with the SCG algorithm, and the same results were obtained (SCG 10-3, Table 4). Thus any observed difference in the predictive performance between their and our subsequent QSAR models was not due to neural network implementations or initialization conditions. The comparison also shows that they did obtain nearly converged results with a SD implementation of their neural network. The same seed that was used in this control run was kept for all the subsequent neural network simulations.

**Cross-Validated Genetic Neural Network.** A GNN (codename C10-3) simulation that maximized the

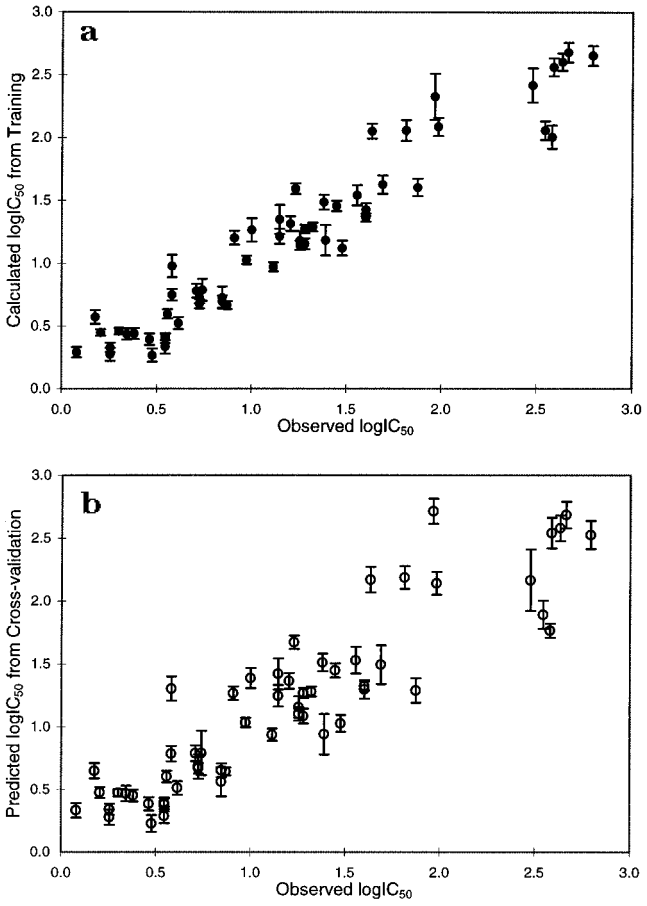


**Figure 3.** (a) Calculated activity from training and (b) predicted activity from cross-validation versus observed activity from the best 10-descriptor MJ model. The figures are made by averaging all 500 SCG trials (see text), and the error bar corresponds to one standard deviation from the average value.

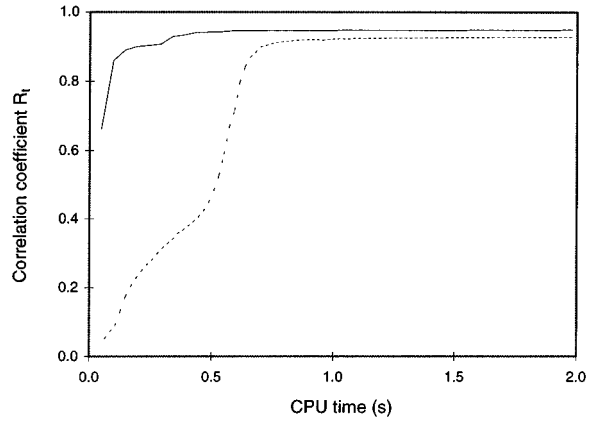


**Figure 4.** Relative efficiency of the SD (− −) and SCG (—) training algorithms for 10-3-1 neural networks. The plot shows the training profiles of a successful SD attempt and a typical SCG run.

value of $R_{cv}$ was performed with the SCG 10-3-1 neural network and the EP algorithm. A population of 200 models was generated. After 50 generations, highly predictive QSARs were found. The best model (C10-3 #1, Table 4) only had four descriptors in common with the MJ selection ($\pi$7, *F*7, MR1, and MR2′) and did not contain a substituent constant for position 3. The increase in $R_{cv}$ of this model (0.939) relative to the best MJ model (0.897) was significant because of the small uncertainty in the neural network outputs. From the comparison of SD and SCG results for the MJ model (see above), this increase arises from a better selection of descriptors and not from the more efficient neural network. The best MJ model was not discovered by our analysis. This is in accord with the result that even the last of the 200 GNN models had a $R_{cv}$ value of 0.924, still considerably better than that of the MJ model. The additional random inputs that were introduced in the MJ study were not used because they did not appear in the final MJ model, and furthermore we did not believe such nonphysical descriptors would play a useful role in the GNN simulation.

**Alternative Protocol To Obtain Highly Predictive Models.** Despite the use of a much improved neural network optimizer, a full cross-validated GNN simulation was computationally expensive. The C10-3 runs required nearly $4^1/_2$ days of CPU time on a fast workstation.[31] To put GNN on a more practical basis, it was clearly desirable to devise a new protocol that would substantially decrease the computational cost while producing predictive QSARs of equal quality. An alternative GNN procedure was proposed that led to a significant reduction of computational cost. In the previous GNN paper, we suggested the construction of a separate test set from existing training data and used the accuracy of test set prediction as a criterion for determining the fitness of the individuals in a GA.[1] In this way, the expensive cross-validation that was required in the current protocol could be avoided. However, a potential drawback of this approach was the difficulty in forming an objective test set, i.e., it was conceivable that the test set compositions might bias the final outcome.

Here we describe an unbiased procedure to obtain models with high predictivity. The previous C10-3 simulation evolved in such a way that the $R_{cv}$ values of the QSAR models were being maximized so that the most predictive models were discovered at the end of the GNN run. Figure 5a is a plot of the values of $R_{cv}$ and $R_t$ from these 200 models. Although there was no strong correlation ($R = 0.36$) between the two correlation coefficients, it was clear that sufficient neural network training is a prerequisite of high predictivity. All predictive models had high $R_t$ values, and over 96% of them were above 0.95. Thus, it seemed reasonable to expect that if one would carry out a GNN simulation that optimized on the value of $R_t$, a number of these highly *correlated* models (high $R_t$) found at the end would also be highly *predictive* (high $R_{cv}$). This is the basis of the new GNN protocol. A GNN simulation (T10-3) that optimized the values of $R_t$ was performed first. Full cross-validations were done to the QSAR models only at the final generation to obtain $R_{cv}$ values. These QSARs were then ranked according to their predictivity ($R_{cv}$ values).

The new protocol was applied to the present data set with the same (10-3-1) neural network topology (codename T10-3). This simulation took 4.1 CPU h,[31] which was only a small fraction of the time ($4^1/_2$ days) required for a full cross-validation procedure. The best QSAR values ($R_{cv} = 0.938$) found by the T10-3 simulation are also shown in Table 4. The fact that this model was different from the best one of C10-3 meant that both of the genetic searches were limited in scope. This is not surprising because of the very small sampling size[32] ($1.0 \times 10^5$) relative to the number of distinct combinations[33]

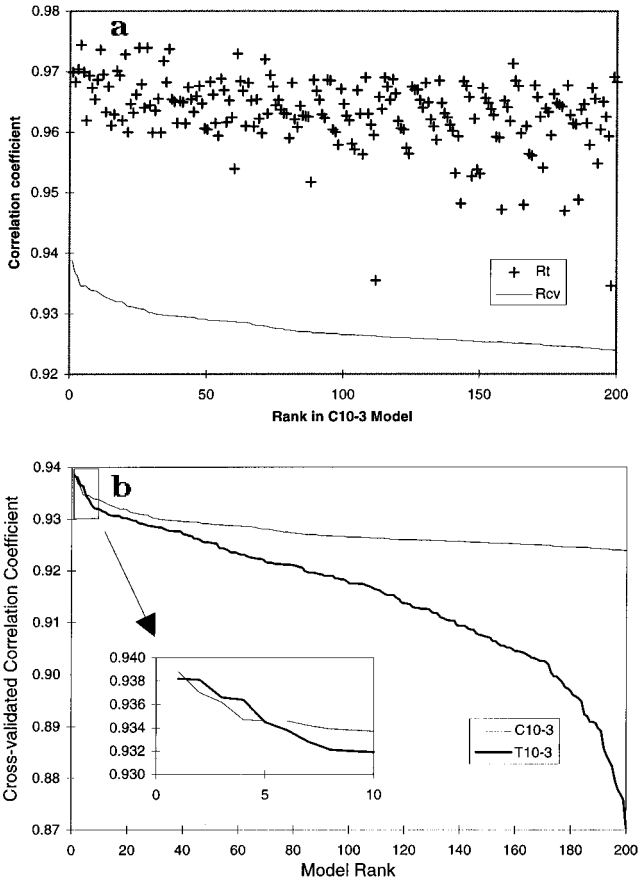


**Figure 5.** (a) Correlation coefficients in training ($R_t$) and cross-validation ($R_{cv}$) plotted against their ranks for the 200 models generated in the C10-3 GNN simulation. It shows that the corresponding $R_t$ values for the highly predictive models are generally high (see text). (b) Plot of cross-validated correlation coefficients ($R_{cv}$) for the C10-3 and T10-3 GNN simulations as a function of their rank. The quality of prediction given by both sets is essentially the same when the top models are considered. C10-3 becomes considerably better than T10-3 only when models beyond a rank of 20 are required.

($1.5 \times 10^9$). Nevertheless, it was encouraging that the two top models converged to a similar level of predictivity, i.e., both of them appear to be near the true optimum for the available data. Figure 5b shows the values of $R_{cv}$ as a function of model rank for the two sets of GNN simulations. Only when models beyond a rank of 20 are needed does C10-3 become considerably better than T10-3. Nevertheless, for most practical purposes, only the top five or so GNN models would normally be considered in further data analysis.

The most important finding in the present context is that there was no obvious advantage of C10-3 over the much faster T10-3 for obtaining highly predictive models. A simple analysis showed that the time advantage of the new protocol over that of full cross-validation was a factor of $N \times G/(N + G)$, where $N$ = number of compounds in the data set and $G$ = number of evolutionary generations. The current GNN system had $N = 57$ and $G = 50$, and we would therefore expect T10-3 to run about 27 times faster than C10-3. This was indeed the observed ratio in CPU usage between the two runs.

**Variation of Number of Inputs.** Another way to save CPU time was to minimize the size of the neural network without losing effectiveness. This has the additional advantage that the risk of overfitting due to a limited data set is reduced. The new protocol was efficient enough so that a large number of GNN simulations could be performed to investigate the effect of variation in network topology. With the number of hidden nodes reduced to two, a series of T*X*-2 simulations that varied the number of input nodes ($X$) was made.

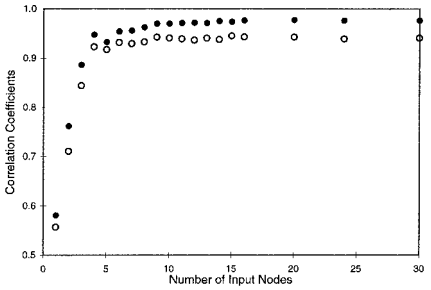


**Figure 6.** Variation of $R_t$ and $R_{cv}$ as a function of the number of input nodes used in the GNN simulation: (●) training and (○) cross-validation.

Figure 6 shows the values of $R_{cv}$ of the best models in each of these runs. This plot demonstrates that $R_{cv}$ does not appear to increase significantly with the number of input nodes after a sharp initial rise. It was noteworthy that the best 6-descriptor QSAR ($R_{cv}$ = 0.932) obtained by GNN was almost as predictive as its 10-descriptor counterpart ($R_{cv} = 0.938$). This disagrees with the result of MJ, who had found a substantial decrease in the predictivity of their 6-descriptor QSAR ($R_{cv} = 0.776$) relative to the 10-descriptor QSAR ($R_{cv}$ = 0.896). We attribute the difference in performance between the MJ and the GNN 6-descriptor QSARs to the following factors. The MJ selection procedure permitted a very limited scope for coupling between descriptors, and some of the potentially important descriptors might have been incorrectly eliminated due to redundancy in the early stages. Furthermore, MJ required the inclusion of at least one descriptor for each of the substituent positions. This could result in redundant or noisy signals when binding affinity was not related to a given substituent position, e.g., the very predictive C10-3 model did not require a descriptor in the 3-position.

The negligible difference in performance between the 6-descriptor and 10-descriptor GNN models set the limit of significant information of the data set. Besides faster computations, another advantage in dealing with a smaller set of input descriptors is that subsequent analysis can be greatly simplified. One can focus on the few key elements that appear to determine the biological activities. For this reason our subsequent work was to be based upon the best 6-descriptor models found by our T6-2 GNN run, which only took 3.3 CPU h.[31]

**Analysis of the Chosen Descriptors in T6-2 GNN Models.** The top-ranking 6-descriptor QSARs emerged from the T6-2 GNN simulation are shown in Table 4 (T6-2 #1−10). As already discussed in the preceding section, the current result represented a major improvement over the MJ 6-descriptor model derived from a pruning procedure. The gain in correlation and predictivity was substantial and significant. Moreover, the T6-2 GNN models were also better than the best MJ 10-descriptor model.

Some interesting similarities and differences emerged when our GA-derived selections were compared with the

MJ set. For substituent position 7, there was agreement with the MJ study that $\pi 7$ was an essential descriptor. It was also evident that a second descriptor from this position was necessary. GNN selected either $F7$ or $\sigma_m 7$. The sets of values from the two descriptors encoding the $R_7$ substituents were highly correlated ($R^2 = 0.96$), and it was appropriate that GNN never included both of them in the same model. Interestingly MJ had the same three descriptors in the final elimination rounds at this position. However, it was not obvious from their procedure how many of the three should be included in the QSAR model.

MJ used three descriptors to encode the chemical environment at position 1. However, the GNN analysis suggested that only one of them, MR1, was the necessary and sufficient determinant at this position. Any additional descriptor at this position seemed superfluous. This suggested that a good steric fit of the $R_1$ substituent with the receptor site is the fundamental factor in binding, and hydrophobic or electronic effects at this position are of lesser importance.

For position 2′, instead of either MR2′ or $F2'$ of the MJ optimal model, $\sigma_m 2'$ was selected by GNN as the most important parameter for some very predictive 6-descriptor models. In a few cases, $\sigma_p 2'$ was also used to provide complementary information. The importance of the Hammett parameters gave support to the conclusion that the principal factor governing receptor affinity at this position was electrostatic in nature.[17−20]

Since positions 1, 2′, and 7 are the ones for which the data set is most complete, we did some additional tests on three important determinants (MR1, $\sigma_m 2'$, and $\pi 7$) at these positions. Four separate 6-2-1 GNN simulations were made using data sets from which any one or all of the three descriptors had been removed, and the results were compared with the top T6-2 model ($R_{cv} = 0.932$). Without the MR1 descriptor, $R_{cv}$ of the most predictive GNN model reduced significantly to 0.871; a data set without description involving $\sigma_m 2'$ yielded a maximum $R_{cv}$ value of 0.925; and without $\pi 7$ it was 0.911. Finally $R_{cv}$ dropped to as low as 0.777 when all three descriptors were absent. This set of results suggested that $\pi 7$ and in particular MR1 were the essential descriptors for obtaining highly predictive models. On the other hand, it appeared that $\sigma_m 2'$ was not indispensable and could be replaced by other electronic parameters (such as $F2'$ as in the MJ model), though its unanimous preference in all of the top T6-2 models suggested that GNN was sufficiently sensitive to select it as the principal determinant that worked best in conjugation with other descriptors.

MJ pointed out that any interpretation of the choices of the descriptors made at positions 3, 6′, and 8 should be treated with caution because there was only a limited range of substituents in the data set. This was reflected in the GNN result by the fact that there was no consensus as to which descriptors for these three positions were particularly useful. In some cases, descriptors from a given position were not even represented, e.g., the only pattern found by GNN was that a descriptor from position 6′ was useful, though not essential (T6-2 #9 in Table 4 did not contain such descriptors), for obtaining a high correlation. Both the $R_3$ and $R_8$ substituents appeared to be less important, as indicated by a general lack of representation for

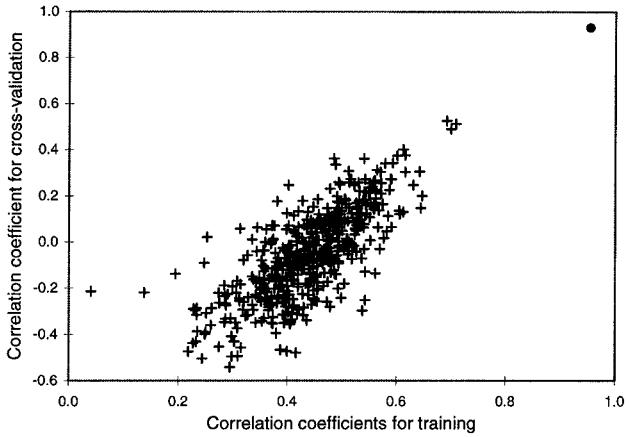


**Figure 7.** Scatter plot for $R_{cv}$ against $R_t$ for the real QSAR (●) and those with randomized activity values (+).

descriptors from these two positions though they did appear in some models. This implies either that the types of substituents at these positions in the data set do not play an important role in the activity determination or that the nature of the descriptors is not appropriate for describing the interactions between the substituents and the receptor subsites. In this regard, it is interesting to note that substitution at the 3-position is important, as has been shown by the study of Blount et al. that the stability of the bioactive conformation is related to the stereospecificity of the substituent at this site.[34] However, because our activity data are derived from either the active enantiomer or the racemic mixture, there is no information on this in the data set. Finally, the selection of GNN descriptors at position 8 disagrees with the conclusion of MJ that the only useful input is $\sigma_p 8$. Other descriptors such as MR8 and $\sigma_m 8$ were also shown to be effective.

**Validation of the Best T6-2 GNN Model.** Three independent tests were made to validate the reliability and the robustness of some of the best T6-2 GNN models. First, a full cross-validation GNN run (C6-2) was performed that directly optimized the values of $R_{cv}$. The top four models from this run are listed in Table 4. The fact that the top three models of T6-2 were among these four suggested that (unlike the C10-3/T10-3 GNN simulations) the sampling by the GA was now sufficient to provide nearly converged sets of optimal models.

The second validation tool was the randomization test that is often employed.[35] The output activities of the 57 BZs were given randomized values, and the resulting data were trained against real input descriptors. The rationale behind this test is that the significance of the real QSAR diminishes if there is a significant correlation between selected descriptors with these randomized response variables. Figure 7 shows a plot of $R_{cv}$ against $R_t$ for 500 such runs (crosses) using T6-2 #1, together with the point (circle) corresponding to the real QSAR. The separation of the true QSAR point from those corresponding to random response variables provides compelling evidence that the QSAR is the result of genuine correlation between the chosen descriptors and activity. Thus, the probability of chance correlation is extremely low.[36,37]

The final measure to verify the usefulness of the chosen descriptors was to determine the cross-correlation among the chosen input descriptors.[20] Table 5 shows the correlation matrix for the six descriptors in

**Table 5.** Correlation ($R^2$) Matrix for the Descriptors in the Best 6-Descriptors GNN Model

| | $\pi$7 | *F*7 | MR1 | $\sigma_m$2 | $\pi$6 | MR8 |
|---|---|---|---|---|---|---|
| $\pi$7 | 1 | | | | | |
| *F*7 | 0.02 | 1 | | | | |
| MR1 | 0.00 | 0.06 | 1 | | | |
| $\sigma_m$2 | 0.01 | 0.00 | 0.02 | 1 | | |
| $\pi$6 | 0.05 | 0.00 | 0.00 | 0.05 | 1 | |
| MR8 | 0.01 | 0.07 | 0.02 | 0.01 | 0.00 | 1 |

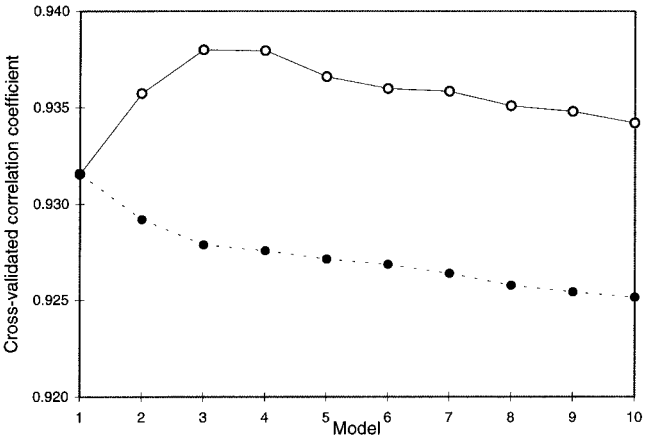


**Figure 8.** Diagram showing the cross-validated correlation coefficient plotted against the *N*-averaged composite model (○) and the *N*th best model (●).

the T6-2 #1 model. The values of correlation coefficients were generally small. This suggested that each of the inputs was independent and furthermore optimally useful in explaining the variance in binding affinities.

**Evaluation of the Top Models.** The original GNN study[1] found that a more reliable set of predictions could be obtained by averaging the outputs of several top-ranking GNN models. The same averaging was performed here, and the averaged predicted activities of up to 10 highest ranking T6-2 GNN models were calculated. The $R_{cv}$ values of these new composite models (open circles) were plotted (Figure 8) in conjugation with the $R_{cv}$ values of the parent GNN models (filled circles). As in the previous case a small gain in predictivity was observed upon averaging, i.e., the $R_{cv}$ value of the composite model peaked when the best three or four models were combined, and with further averaging it decreased gradually but remained greater than any individual GNN model value.

The result indicated that a gain in predictivity with simultaneous averaging of a few high-performance QSARs is a general result, though in the present case the gain was not statistically significant. Nevertheless, it led us to use of at least the best three T6-2 models for the design of new compounds.

**Minimum Perturbation Approach to Lead Optimizations.** A good QSAR provides insights into the underlying physical parameters involved in drug−receptor binding and indicates the key structural elements that are required for an active drug. Such knowledge is of primary interest for the prediction of new related molecules with high potency as candidates for synthesis. In this section we use the results to predict new compounds by a minimum perturbation approach in which we focus on one substitution at a time. More time-consuming approaches that take advantage of the nonlinear and nonadditive attributes of these GNN QSARs can also be applied. For example, an exhaustive screening of all six physicochemical parameters in a given model, or a multivariate experimental design method,[35] can be used to search for improved BZ analogues. However, the clear results from the simple analysis suggest that it should be a useful method for predictions.

The three highly predictive T6-2 GNN models (#1−3) were used for predicting the activity of a number of new BZs. The most active compound (**46**) was used as the template that was subjected to small structural changes. In this study, new substituents at a given position were restricted to those that were not significantly greater in bulk than the largest known substituent from a compound of at least moderate activity. This measure was taken to avoid generating new compounds that have a poor steric fit with the receptor.

The effect of a given substituent on biological activity was monitored by a procedure that had appeared in many neural network studies, including that of MJ.[1,20,38,39] In the current implementation, all but one of the input parameters was kept at the values that corresponded to the substituents of the template and the one remaining input descriptor was varied between the minimum and the maximum of its known range from the parameters in Table 2. The resulting plot provided the functional dependence of the biological activity on the variable descriptor. The same procedure was repeated for all input descriptors and for each of the three models. These functionality plots are shown in Figure 9.

The only curves with definite minima were the three corresponding to MR1 (Figure 9a). All three indicated that the most potent compounds would have a value of MR1 within the range of 1.0−3.0. The only substituent with a qualifying MR value is H. This finding is consistent with Haefely's analysis that the affinity of the derivatives carrying a side chain at this position was always smaller than that of the parent compound.[10] On this basis the template was unmodified at this position. A similar situation was found for the substituent at position 8. The plot (Figure 9b) suggests that optimal activity is reached using substituents that have low values of both MR8 and $\sigma_m$8. While it was possible to select new substituents that had either MR or $\sigma_m$ values lower than those of H in the template, the relatively flat dependence of the curves suggested that the any change in predicted activity would be minimal. The additional synthetic effort associated with a regiospecific substitution at that position appears not to be justified. At position 3 the methyl group was kept from the template. This was done because of the lack of representation at this position in the three GNN QSARs, i.e., there is no indication of explicit influence from any substituent at this position.

The plot for 2′-position (Figure 9b) suggested that increases in the Hammett constants would enhance the predicted receptor affinity. Thus, the template was modified with a substituent that had higher values of $\sigma_m$2′ and $\sigma_p$2′ than those of Cl. Three relatively simple substituents, CN, $NO_2$, and $SO_2F$, were considered. Their substituent values are listed in Table 2.

An examination of Figure 9a showed that predicted activity would be also improved by decreases in substituent parameters ($\pi$6′, MR6′, and $\sigma_p$6′) at the 6′-position. In the template this position was occupied by

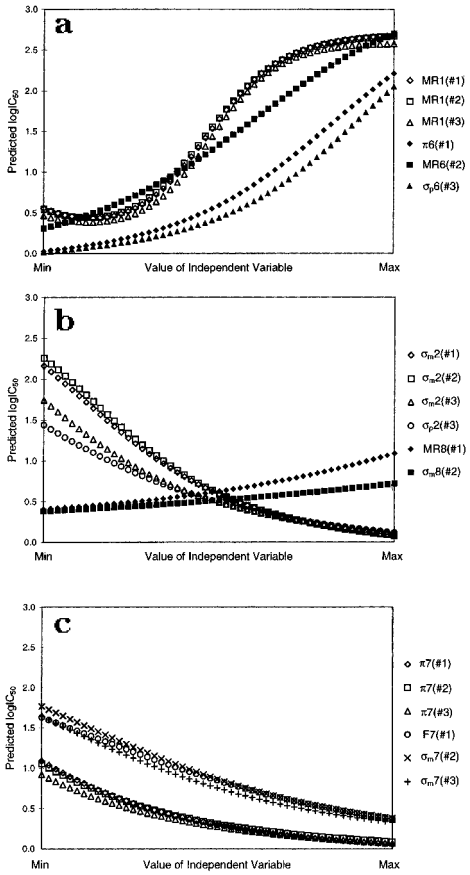


**Figure 9.** Predicted activity as a function of the descriptors that have been chosen by GNN. The ranges of each physiochemical descriptor type shown in the figures are based on the 22 substituents used in the MJ study (see Table 2). Their minimum and maximum values are $\mu$ (−4.13 to 1.53), $\pi$ (−1.34 to 1.98), MR (0.92 to 19.62), $F$ (−0.07 to 0.67), $R$ (−0.68 to 0.20), $\sigma_m$ (−0.16 to 0.71), and $\sigma_p$ (−0.66 to 0.78). The scale is linear between min and max for each descriptor.

a hydrogen atom with $\pi = 0.0$, MR = 1.03, and $\sigma_p = 0.00$. However, because there is no functional group that has lower substituent parameters in all three categories than those of hydrogen, substituents with optimized properties in one or two categories were considered. The effect on predicted receptor affinity upon substitution with $NH_2$ (smaller $\pi6'$ and $\sigma_p6'$ values) and $F$ (smaller MR6′ value) was examined. Furthermore, because of equivalency of the 2′- and 6′-positions in the phenyl ring, it might be worthwhile to try the three new 2′ substituents (CN, $NO_2$, and $SO_2F$) as well.

The last, and perhaps the key, position for modification was position 7. It is the position for which the most data are available. The dependency plots of $\pi7$, $F7$, and $\sigma_m7$ (Figure 9c) were consistent with published analyses that increases in both lipophilicity and polar effects in this position increased the receptor affinity.[9,17,19,20] The challenge was to find some simple substituents that were in the optimal ranges. This was not trivial because lipophilicity and polarity are naturally anticorrelated, especially for groups that are small enough. Three substituents were considered to be suitable candidates: $CH_2CF_3$, $SO_2F$, and $OCF_3$.

Based on the above analysis, 11 new compounds (**58**−**68**) deriving from single substituent replacement on the template were proposed. Their biological activities were predicted using the best three T6-2 models, and the results are listed in Table 6. In all cases, replacements of $NO_2$ by any of the three designed substituents at the 7-position led to an increase in predicted receptor affinity. This indicated that a good compromise between lipophilicity and polarity had been achieved. More pronounced effect on the predicted biological activity was observed with the compounds that had new 2′ substituents. This was particularly true for $NO_2$ and $SO_2F$ derivatives for which very high activities were predicted by all three GNN models.

On the other hand, substitution at the 6′-position led to mixed results. For example, a $NH_2$ substitution (**64**) would lead to an increase in receptor affinity according to two predictions but a significant decrease from the third. In the other four cases, GNN models indicated that a decrease in receptor affinity was the more probable outcome. This seemed to suggest that a second ortho substitution would have a detrimental effect on the predicted activity. From published data the exact

**Table 6.** New BZs with Designed Substituents[a]

| | substituents | | | | | | predicted activity | | |
|---|---|---|---|---|---|---|---|---|---|
| ID | $R_7$ | $R_1$ | $R_{2'}$ | $R_{6'}$ | $R_3$ | $R_8$ | #1 | #2 | #3 |
| template **46** | $NO_2$ | H | Cl | H | $CH_3$ | H | 0.440 | 0.442 | 0.385 |
| **58** | $OCF_3$ | H | Cl | H | $CH_3$ | H | **0.384** | **0.433** | **0.350** |
| **59** | $CH_2CF_3$ | H | Cl | H | $CH_3$ | H | **0.365** | **0.371** | **0.293** |
| **60** | $SO_2F$ | H | Cl | H | $CH_3$ | H | **0.304** | **0.314** | **0.276** |
| **61** | $NO_2$ | H | CN | H | $CH_3$ | H | **0.237** | **0.235** | **0.114** |
| **62** | $NO_2$ | H | $NO_2$ | H | $CH_3$ | H | **0.135** | **0.133** | **0.047** |
| **63** | $NO_2$ | H | $SO_2F$ | H | $CH_3$ | H | **0.090** | **0.089** | **0.011** |
| **64** | $NO_2$ | H | Cl | $NH_2$ | $CH_3$ | H | **0.099** | 0.785 | **0.058** |
| **65** | $NO_2$ | H | Cl | F | $CH_3$ | H | 0.503 | **0.435** | 0.439 |
| **66** | $NO_2$ | H | Cl | CN | $CH_3$ | H | **0.242** | 0.874 | 1.382 |
| **67** | $NO_2$ | H | Cl | $NO_2$ | $CH_3$ | H | **0.332** | 0.984 | 1.650 |
| **68** | $NO_2$ | H | Cl | $SO_2F$ | $CH_3$ | H | 0.462 | 1.133 | 1.944 |
| **69** | $OCF_3$ | H | CN | H | $CH_3$ | H | **0.229** | **0.275** | **0.119** |
| **70** | $OCF_3$ | H | $NO_2$ | H | $CH_3$ | H | **0.143** | **0.186** | **0.057** |
| **71** | $OCF_3$ | H | $SO_2F$ | H | $CH_3$ | H | **0.103** | **0.143** | **0.021** |
| **72** | $CH_2CF_3$ | H | CN | H | $CH_3$ | H | **0.219** | **0.232** | **0.092** |
| **73** | $CH_2CF_3$ | H | $NO_2$ | H | $CH_3$ | H | **0.137** | **0.153** | **0.038** |
| **74** | $CH_2CF_3$ | H | $SO_2F$ | H | $CH_3$ | H | **0.098** | **0.115** | **0.006** |
| **75** | $SO_2F$ | H | CN | H | $CH_3$ | H | **0.150** | **0.153** | **0.066** |
| **76** | $SO_2F$ | H | $NO_2$ | H | $CH_3$ | H | **0.073** | **0.074** | **0.013** |
| **77** | $SO_2F$ | H | $SO_2F$ | H | $CH_3$ | H | **0.039** | **0.040** | **-0.016** |

[a] Predicted activities that are higher than that of the template compound are shown in bold.

mechanism that makes the di-ortho derivatives less active is unknown, although it has been suggested that the second substituent might be in bad steric contact with its receptor subsite;[20] or if Loew's hypothesis of N4 cationic receptor site was correct,[17] the steric hindrance from the di-ortho-substituted C-ring was so large that the near-planar geometry that was required for charge conjugation to N4 could not be attained.

In the preliminary analysis it was predicted that three new substituents at the 7-and 2′-positions of BZs would exert a positive influence on receptor affinity. In the last set of designed compounds (**69**−**77**), the nine different combinations for the two positions were examined. Their structures and predicted activity are shown in Table 6. All of them were predicted to be highly active by the three GNN models, and the standard deviation of error were no greater than 0.08 in all cases. The first six (**58**−**63**) and the final nine (**69**−**77**) compounds are candidates for synthesis.

## Concluding Discussion

The GNN method for QSAR has been improved and tested by a new application. We have introduced a SCG neural network simulator that makes use of pseudo-second-derivative information. The superior ability of the SCG optimizer in dealing with complex multivariate systems was demonstrated by a multiple-neural network simulation benchmark. It showed that there are significant improvements in the speed of convergence and the stability of the solution over the commonly used SD-based neural networks. Further, a new GNN protocol that is both effective and unbiased was proposed to derive highly predictive QSARs. This resulted in a dramatic speedup in GNN simulations that made possible a more extensive study of the effect of neural network topology.

Using a combination of genetic algorithm and neural network technology, some highly predictive QSARs that are superior to the optimal model of MJ have been discovered. By trying a large number of neural network inputs on this data set, it is found that QSARs with significantly fewer inputs than the MJ model are capable of providing as good correlation and predictions. The significance of the chosen descriptors has been discussed, and they are compared with previous analyses on the BZs. The best 6-descriptor GNN QSAR model has been further analyzed, and they have provided a basis for the design of novel BZ analogues.

The most active compound in the data set has been used as a design template, and a minimum perturbation approach has been applied to suggest new compounds. A number of substituents were proposed by consideration of functional dependence plots. Upon optimizations of the properties of the 7 and 2′ substituents, several new compounds were predicted as highly potent by the GNN models.

The improved GNN method has been validated by several statistical measures. Comparison with the result from a full cross-validation GNN run suggests that the final QSAR comes close to having optimal predictivity. By a series of simulations with randomized response variable, we are able to show that the correlation between the chosen descriptors and the activity is real and not chance related. Finally, a cross-correlation analysis of the input matrix showed that no pair of inputs is correlated and that there is no redundant information.

This study has shown the general utility of the GNN methodology in dealing with data sets of high dimensionality.

**Acknowledgment.** We thank John-Marc Chandonia for his advice in the construction of the SCG neural network simulator and the referees for their careful reviews. This work was performed on Hewlett Packard 735/125, Silicon Graphics Indigo R4000/R3000, and DEC Alpha workstations. This project is supported in part by a grant from the National Institutes of Health.

**Supporting Information Available:** Complete cross correlation matrix for the 42 QSAR parameters by position of substitution (3 pages). Ordering information is given on any current masthead page.

## References

(1) So, S.-S.; Karplus, M. Evolutionary optimization in quantitative structure-activity relationship: an application of genetic neural network. *J. Med. Chem.* **1996**, *39*, 1521−1530.
(2) Selwood, D. L.; Livingstone, D. J.; Comley, J. C.; Hudson, A. T.; Jackson, P.; Jandu, K. S.; Rose, V. S.; Stables, J. N. Structure-activity relationships of antifilarial antimycin analogues: a multivariate pattern recognition study. *J. Med. Chem.* **1990**, *33*, 136−142.
(3) Rogers, D. R.; Hopfinger, A. J. Application of genetic function approximation to quantitative structure-activity relationships and quantitative structure-property relationships. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 854−866.
(4) Luke, B. T. Evolutionary programming applied to the development of quantitative structure-activity relationships and quantitative structure-property relationships. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1279−1287.
(5) Wikel, J. H.; Dow, E. R. The use of neural networks for variable selection in QSAR. *Bioorg. Med. Chem. Lett.* **1993**, *3*, 645−651.
(6) Möhler, H.; Okada, T. Benzodiazepine receptor: demonstration in the central nervous system. *Science* **1977**, *198*, 849−851.
(7) Costa, T.; Rodbard, D.; Pert, C. B. Is the benzodiazepine receptor coupled to a choride anion channel? *Nature* **1979**, *277*, 315−317.
(8) Tallman, J. F.; Paul, S. M.; Skolnick, P.; Gallager, D. W. Receptor for the age of anixety: pharmacology of the benzodiazepines. *Science* **1980**, *207*, 274−281.
(9) Borea, P. A.; Bonora, A. Brain receptor binding and lipophilic character of benzodiazepines. *Biochem. Pharmacol.* **1983**, *32*, 603−607.
(10) Haefely, W.; Kyburz, E.; Gerecke, M.; Möhler, H. Recent advances in the molecular pharmacology of benzodiazepine receptors and in the structure-activity relationships of their agonists and antagonists. *Adv. Drug Res.* **1985**, *14*, 165−322.
(11) Villar, H. O.; Uyeno, E. T.; Toll, L.; Polgar, W.; Davies, M. F.; Loew, G. H. Molecular determinants of benzodiazepine receptor affinities and anticonvulsant activities. *Mol. Pharmacol.* **1989**, *36*, 589−600.
(12) Pritchett, D. B.; Sontheimer, H.; Shivers, B. D.; Ymer, S.; Kettenmann, H.; Schofield, P. R.; Seeburg, P. H. Importance of a novel $GABA_A$ receptor subunit for benzodiazepine pharmacology. *Nature* **1989**, *338*, 582−585.
(13) Bianucci, A. M.; Da Settimo, A.; Da Settimo, F.; Primofiore, G.; Martini, G.; Giannaccini, G.; Lucacchini, A. Benzodiazepine receptor affinity and interaction of some N-(Indo-3-ylglyoxylyl)-amine derivatives. *J. Med. Chem.* **1992**, *1992*, 2214−2220.
(14) Wong, G.; Koehler, K. F.; Skolnick, P.; Gu, Z.-Q.; Ananthan, S.; Schönholzer, P.; Hunkeleer, W.; Zhang, W.; Cook, J. M. Synthetic and computer-assisted analysis of the structural requirements for selective, high-affinity ligand binding to diazepam-insensitive benzodiazepine receptors. *J. Med. Chem.* **1993**, *36*, 1820−1830.
(15) Bunin, B. A.; Ellman, J. A. A general and expedient method for solid-phase synthesis of 1,4-benzodiazepine derivatives. *J. Am. Chem. Soc.* **1992**, *114*, 10997−10998.
(16) Plunkett, M. J.; Ellman, J. A. Solid-phase synthesis of structurally diverse 1,4-benzodiazepine derivatives using the Stille coupling reaction. *J. Am. Chem. Soc.* **1995**, *117*, 3306−3307.
(17) Loew, G. H.; Nienow, J. R.; Poulsen, M. Theoretical structure-activity studies of benzodiazepine analogues. *Mol. Pharmacol.* **1984**, *26*, 19−34.
(18) Ghose, A. K.; Crippen, G. M. Modeling the benzodiazepine receptor binding site by the general three-dimensional structure-directed quantitative structure-activity relationship method REMOTEDISC. *Mol. Pharmacol.* **1990**, *37*, 725−734.

(19) Greco, G.; Novellino, E.; Silipo, C.; Vittoria, A. Study of benzodiazepines receptor sites using a combined QSAR-CoMFA approach. *Quant. Struct.-Act. Relat.* **1992**, *11*, 461−477.
(20) Maddalena, D. J.; Johnston, G. A. R. Prediction of receptor properties and binding affinity of ligands to benzodiazepine/GABAA receptors using artificial neural networks. *J. Med. Chem.* **1995**, *38*, 715−724.
(21) They also introduced a nonphysical random variable as a descriptor for each substituent position. These random data input sets were not chosen in their final model.
(22) Hansch, C.; Leo, A. *A substituent constants for correlation analysis in chemistry and biology*; John Wiley & Sons, Inc.: New York, 1979.
(23) Lien, E. J.; Guo, Z.-R.; Li, R.-L.; Su, C.-T. Use of dipole moment as a parameter in drug-receptor interaction and quantitative structure-activity relationship studies. *J. Pharm. Sci.* **1982**, *71*, 641−655.
(24) Rumelhart, D. E.; Hinton, G. E.; Williams, R. J. *Learning internal representations by error propagation*; Parallel distributed processing Vol. 1; MIT Press: Cambridge, MA, 1986.
(25) Hertz, J.; Krogh, A.; Palmer, R. G. *Introduction to the theory of neural computation*; Addison-Wesley Publishing Co.: Redwood City, CA, 1991.
(26) Møller, M. F. A scaled conjugate gradient algorithm for fast supervised learning. *Neural Networks* **1993**, *6*, 525−533.
(27) Chandonia, J.-M.; Karplus, M. The importance of larger data sets for protein secondary structure prediction with neural networks. *Protein Sci.* **1996**, *5*, 768−774.
(28) Holland, J. H. *Adaption in natural and artificial systems*; The University of Michigan Press: Ann Arbor, MI, 1975.
(29) Michalewicz, Z. *Genetic algorithms + Data structures = Evolution Programs*, 1st ed.; Artificial Intelligence Vol. XVII; Springer-Verlag: Berlin, 1992.
(30) Cartwright, H. M. *Applications of artificial intelligence in chemistry*; Oxford University Press: Oxford, 1993.
(31) A number of computers with different speeds have been used to perform calculations in this study. The time measurements given in the text correspond to either real CPU time or one that has been appropriately scaled for a Hewlett-Packard 735/125 workstation.
(32) The number of new generations was 50 plus zeroth generation pool = 51, and there were 200 individuals. The sampling size was thus 200 × 51 = 10 200.
(33) The total number of distinct 10-descriptor models chosen from a set of 42 descriptors is 42!/10!(42−10)! = 1 471 442 973.
(34) Blount, J. F.; Fryer, R. I.; Gilman, N. W.; Todaro, L. J. Quinazolines and 1,4-benzodiazepines. 92. Conformational recognition of the receptor by 1,4-benzodiazepines. *Mol. Pharmacol.* **1983**, *24*, 425−428.
(35) van de Waterbeemd, H. *Chemometric methods in Molecular Design*; Methods and principles in medicinal chemistry Vol. 2; VCH: Weinheim, 1995.
(36) Topliss, J. G.; Edwards, R. P. Chance factors in studies of quantitative structure-activity relationships. *J. Med. Chem.* **1979**, *22*, 1238−1244.
(37) Livingstone, D. J.; Manallack, D. T. Statistics using neural networks: Chance effects. *J. Med. Chem.* **1993**, *36*, 1295−1297.
(38) Andrea, T. A.; Kalayeh, H. Applications of neural networks in quantitative structure-activity relationships of dihydrofolate reductase inhibitors. *J. Med. Chem.* **1991**, *34*, 2824−2836.
(39) So, S.-S.; Richards, W. G. Application of neural networks: quantitative structure-activity relationships of the derivatives of 2,4-diamino-5-(substituted-benzyl)pyrimidines as DHFR inhibitors. *J. Med. Chem.* **1992**, *35*, 3201−3207.

JM960536O