Neural Network Studies. 1. Estimation of the Aqueous Solubility of Organic Compounds

Nicholas Bodor,* Alan Harget,[1] and Ming-Ju Huang

Contribution from the Center for Drug Discovery, College of Pharmacy, J.H.M. Health Center, University of Florida, Gainesville, Florida 32610. *Received April 22, 1991*

Abstract: A study has been made of the ability of neural networks to estimate the aqueous solubility of a wide range of organic compounds. A training set of 331 compounds was used and the trained neural network tested on a prediction set of 19 unknown compounds. A comparison was made with the results obtained from a previous study using regression analysis (Bodor et al.[2]). On the basis of training set results, the neural network model exceeded the performance of the regression analysis technique and gave a prediction which was sufficiently accurate to estimate the water solubility of an organic compound based entirely on calculated values of selected properties.

Introduction

The aqueous solubility of a drug is a key parameter in determining its biological activity. Phenomenological treatment of drug delivery, transport, and distribution are dependent on knowledge of aqueous solubilities. The aqueous solubility of an organic compound can often be the major factor which controls its bioaccumulation. A knowledge of aqueous solubility is also essential for predicting the general environmental distribution of substances.

A variety of computational methods have been used by the medicinal chemist in order to predict quantitative structure–activity relationships. The methods have been established and their usefulness demonstrated over a long period of time. As part of our ongoing study to develop computer programs which can give assistance to the medicinal chemist, we have investigated the performance of the relatively new technique of neural networks. The studies conducted in this laboratory show that neural networks have the potential to become an important addition to the computational repertoire available to the chemist.

Neural computing offers exciting new techniques that allow computers to be trained to recognize patterns in data of high dimensionality. Unlike conventional computer algorithms, neural networks have a high degree of parallelism in their structure and in the way they process information. In addition, they also have a number of important distinguishing features such as the ability to learn, a resistance to noise, and a high degree of fault tolerance. The worth of neural networks has already been demonstrated in applications as diverse as text-to-speech conversion,[3] natural language processing,[4] and the analysis of sonar data.[5] Given this variety of application, a study has been conducted in this laboratory to determine the scope and performance of neural networks applied to chemical problems. As part of this study we now describe a neural network application for estimating the aqueous solubility of organic compounds.

The aqueous solubility of a compound is of considerable interest to the drug designer since it is an important determinant in drug delivery and the transport and distribution of a drug. Although aqueous solubility is generally easy to determine experimentally, this is not always appropriate and is clearly impossible to achieve for unknown compounds. A theoretical approach which provides a reliable measure of a drug's aqueous solubility could be used by the drug designer to eliminate some of the large number of candidates for a drug from synthesis and experimentation. The consequent saving in time and resources could then be directed to more likely candidates.

In this study we have used a neural network to determine the association between the aqueous solubility of a compound and calculated values of selected properties of the compound. Comparison is made with the results obtained from a previous study[2] using a regression analysis technique.

In the next section we discuss the theoretical basis of neural networks and then describe the results obtained using the neural network approach.

Theory of Neural Networks

A neural network model is composed of a large number of simple processing units arranged into a network structure of a form similar to that given in Figure 1.

The networks consist in general of an input layer, an output layer, and any number of intermediate layers, called hidden layers. Each unit in the network is influenced by those units to which it is connected, the degree of influence being dictated by the values of the links or connections. The overall behavior of the system can be modified by adjusting the values of the connections, or weights, through the repeated application of a learning algorithm. Several algorithms have been described in the literature, but one of the most popular algorithms, and the one used in this study, is the back propagation algorithm.[6]

The output O_j of a unit j at the nth level is a sigmoidal function of its total input and is given by

$$O_j = (1 + \exp(-y_j))^{-1} \tag{1}$$

where

$$y_j = \sum_i O_i w_{ji} + \Theta_j \tag{2}$$

where O_i is the output of the unit i in the previous layer, w_{ji} is an element of the weight matrix and expresses the weight of the connection between the units j and i, and Θ_j is the bias or threshold for unit j.

Training consists of presenting each training set pattern at the input units and iteratively minimizing the difference between the output of the net O_k and the desired (target) pattern t_k.

The basic equations for the training of the network can be shown to be[6]

$$\Delta w_{kj} = \eta \delta_k O_j \tag{3}$$

where η is a learning factor and δ_k is given by

$$\delta_k = (t_k - O_k) f'(y_k) \tag{4}$$

and $f'(y_k)$ the derivative of the function is

$$f'(y_k) = \mathrm{d}f(y_k) = O_k(1 - O_k) \tag{5}$$

* Author to whom correspondence should be addressed.

(1) On leave of absence from the University of Aston, Birmingham, B4 7PA, England.

(2) (a) Bodor, N.; Gabanyi, Z.; Wong, C.-K. *J. Am. Chem. Soc.* **1989**, *111*, 3783. (b) Bodor, N.; Huang, M.-J. *J. Pharm. Sci.* In press. (c) Bodor, N.; Huang, M.-J. Submitted for publication.

(3) Sejnowski, T.; Rosenberg, C. *Complex Syst.* **1987**, *1*, 75.

(4) Allen, R. *Proc. IEEE First Int. Conf. Neural Networks* **1987**, *2*, 335.

(5) Gorman, R. P.; Sejnowski, T. *Neural Networks* **1988**, *1*, 75.

(6) Rumelhart, D. E.; McClelland, J. L. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*; Bradford Books, MIT Press: Cambridge, 1986.

0002-7863/91/1513-9480$02.50/0

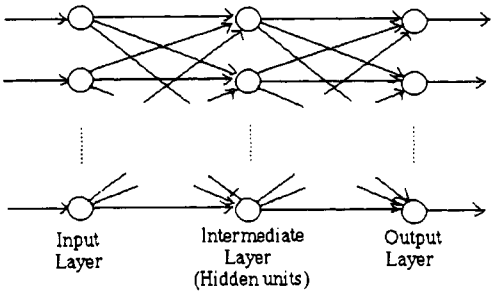

Figure 1. A three-layer neutral network.

Weight correction starts with the output unit (eq 3) and is back propagated toward the input.

The δ term for the hidden units for which there is no specified target is determined recursively in terms of the δ terms of the units to which it directly connects and the weights of those connections, that is

$$\delta_j = f'(y_j)\sum_k \delta_k w_{kj} = O_j(1 - O_j)\sum_k \delta_k w_{kj} \qquad (6)$$

Training continues until either the error between actual outputs and target outputs for the entire training set is acceptably low or a predetermined number of iterations is exceeded. If convergence is attained then the neural network has the ability to determine input patterns of the training set to the overall desired accuracy.

A computer program was designed for a neural network using the back propagation algorithm and implemented in Pascal. The program allows an arbitrary number of hidden layers, and units within each layer, to be defined and incorporates both a training and predictive capability.

Methods

The methods used are essentially as described in our previous work on octanol-water partition coefficients (Log P)[2a,b] and aqueous solubilities (Log W),[2c] where more details are given. First, the molecule is sketched on our IBM PC/AT computer terminal with the ChemCAD software, which generates a starting geometry. After converting the coordinate data file to the format of AMPAC[7] input files, geometry optimization is performed by the semi-empirical AM1 method.[7] From the AM1 optimum geometry and the atomic van der Waals radii, the molecular volume, surface area and ovality are calculated by numerical integration techniques,[2] which we now briefly review.

To calculate molecular volume and surface area we treat the atoms in a molecule as spheres with radii equal to the appropriate van der Waals radii. The molecular volume is taken to be the sum of spherical atomic volumes, but the volume common to more than one atom is counted only once. Likewise the molecular surface area is the sum of spherical atomic surface areas but the common surface area only counted once. The following algorithms are used.

For the volume, we generate three-dimensional cubic grids of points centered at each atomic nucleus. The dimension of the Lth cube is such that it completely encloses a sphere of radius r_L, where r_L is the van der Waals radius of the Lth atom. Points in a grid are considered to belong to atom L if (1) they are within r_L of the center of atom L and (2) they are not within the van der Waals radii of any previously considered atoms. If the number of points satisfying conditions is n and the number satisfying just the first condition is n_t, the Lth atomic volume is given by

$$V_L = {}^4/_3\pi r_L{}^3(n/n_t)$$

With this definition of atomic volume, the total molecular volume is just the sum of atomic volumes.

For the surface area, a set of cubic grids is again generated. Grid points are tested for their closeness to atomic surfaces. A grid point is considered to be on an atomic surface if (1) it is within a present threshold of the spherical surface of an atom and (2) it is not on the surface of any previously considered atom. Again, if n is the total number of grid points satisfying both conditions and n_t is the number satisfying only the first, the Lth atomic surface area is given by

$$S_L = 4\pi r_L{}^2(n/n_t)$$

(7) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902.

Table I. Selected Properties of the Organic Compounds

parameter	definition
S	molecular surface
I_{alkane}	indicator variable for alkenes
D	calculated dipole moment
Q_N	square root of the sum of the squared charges on nitrogen atoms
Q_O	square root of the sum of the squared charges on oxygen atoms
$Q_N{}^2$	sauare of Q_N
$Q_O{}^2$	square of Q_O
$Q_N{}^4$	square of $Q_N{}^2$
$Q_O{}^4$	square of $Q_O{}^2$
V	molecular volume
S^2	square of the molecular surface
C	constant
O	ovality of the molecule
O^2	square of the ovality
ABSHQ	sum of the absolute values of atomic charges on hydrogen atoms
ABSCQ	sum of the absolute values of atomic charges on carbon atoms
I_{amine}	indicator variable for aliphatic amines
NH	number of N–H single bonds in the molecule

With this definition, the molecular surface area is just the sum of atomic surface areas. From the volume and surface, we obtain the ovality, the ratio of the actual surface to the minimum surface:

$$O = S/(4\pi(3V/4\pi)^{2/3})$$

where S is the molecular surface and V is the molecular volume.

Using the AM1 results and the calculated molecular volume, surface area, and ovality we generated a total of 56 molecular descriptors[2] which contain the calculated properties in various forms and combinations computed for each of 331 compounds. Some of these are geometrical (volume, surface area, and ovality), others depend on the structural formula, and others (charges and dipole moment) may be termed quantum chemical descriptors. The descriptors contribute in different ways to a description of solvation phenomena. The most important are the geometrical descriptors. When a solute molecule enters water it creates a hole in the water structure and the energetics of forming this hole are strongly dependent on volume, surface area, and ovality. The charge density parameters describe hydrogen bonding, while the dipole moment is an overall descriptor of the solute–solvent interaction. The other descriptors refine the description of the solvation.

In our regression analysis of log W, we have tested the importance of all these descriptors and found that a 17-parameter function can describe solubility of organic solutes. The 17 parameters seem to be the most important subset of the 56 descriptions. Accordingly our neural network approach began with these.

Results and Discussion

The training set was composed of 331 organic compounds, identical to that used in a recent study of ours,[2] A wide range of organic molecules was used, including hydrocarbons, halohydrocarbons, multiple-substituted benzenes, polynuclear aromatics, ethers, alcohols, aldehydes, ketones, esters, nitrile, and nitro compounds. For each compound properties including the dipole moment and charges were calculated using the semi-empirical AM1 method,[7] and the molecular surface area, volume, and ovality were calculated using the same algorithms as before.[2] The experimental aqueous solubilities were taken from the compilation of Hansch et al.,[8] Hine and Mookerjee,[9] Mackay et al.,[10] and Kamlet et al.[11] with additional results from 21 other sources.[12-32] The values of the solubilities are for atmospheric

(8) Hansch, C.; Quinlan, J. E.; Lawrence, G. L. *J. Org. Chem.* **1968**, *33*, 347.
(9) Hine, J.; Mookerjee, P. K. *J. Org. Chem.* **1975**, *40*, 292.
(10) Mackay, D.; Bobra, A.; Shiu, W. Y.; Yalkowsky, S. H. *Chemosphere* **1980**, *9*, 701.
(11) Kamlet, M. J.; Doherty, R. M.; Abraham, M. H.; Carr, P. W.; Doherty, R. F.; Taft, R. W. *J. Phys. Chem.* **1987**, *91*, 1996.
(12) (a) Chiou, C. T.; Schmedding, D. W.; Manes, M. *Environ. Sci. Technol.* **1982**, *16*, 4. (b) Chiou, C. T.; Porter, P. E.; Schmedding, D. W. *Environ. Sci. Technol.* **1983**, *17*, 227. (c) Chiou, C. T.; Block, J. H. *Partition Coefficient Determination and Estimation*; Dunn, W. J., III, Block, J. H., Pearlman, R. S., Eds.; Pergamon Press: New York, 1986; pp 37–58.
(13) Hermann, R. B. *J. Phys. Chem.* **1972**, *76*, 2754.

pressure and 25 °C. The accuracy and reliability of experimental log W values is most likely not better than 0.1–0.2 log units, if one considers the analytical problems or the effect of even relatively small impurities in the compounds tested, which would affect primarily the compounds with low solubility.

In our previous study we found that 17 molecular properties were sufficient to describe the aqueous solubility of the given range of organic compounds. These same properties have been retained for this study and are listed and described in Table I. The identical nature of the data allows us to compare the results given by the neural network with those obtained from the regression analysis approach.

A three-layer neural network was used comprising 17 input neurons and one output neuron with the number of hidden units varied in order to determine the optimum architecture. Several methods have been employed to decrease the training time of the back-propagation algorithm. The one we have adopted in this study uses the method of Rumelhart et al.[33] which involves adding a term to the weight adjustment which is a portion, called the momentum, of the previous weight change. It was found that this improved the learning rate while enhancing the stability of the process. It is well-known that the performance of a neural network depends on many variables, including the number of training examples, the number of hidden units, and the degree of homology between the training and testing sets. If the size of the training set is too small then the network can readily describe all of the correct outputs because of the large capacity of the weights. The resulting network is then accurate on the training set but gives poor prediction on the testing set. With a training set of 331 compounds and 17 parameters we believe that associations captured by the neural network at the end of the training phase will be significant and meaningful.

In order to determine the optimum number of hidden units several training sessions were conducted with different numbers of units in the hidden layer. The learning factor, η, and momentum coefficient values tested were based on other previous studies in our and other research groups. In a related study we have tested values of between 0.5 and 1.0 for the momentum coefficient and 0.25 and 0.35 for η. We have found the following parameter values give the best performance as reflected by the quickest convergence:

$$\eta = 0.25 \qquad \text{momentum coefficient} = 0.9$$

which were used in the present study. The input data were normalized to give values between 0.0 and 1.0. Training continued until an overall error of 0.001 between target and output values was obtained. It was found that a hidden layer consisting of 18 units gave the best results in terms of the overall error and rate of convergence. This architecture was adopted for all subsequent training and testing sessions.

The results obtained for the training set are shown in Table II together with the experimental values and those results calculated by regression analysis. The standard deviation with the neural network approach was calculated to be 0.23 which was

Table II. The Experimental and Estimated log W Using Neural Networks and Regression Analysis (W in mol/L)[a]

case		measured	regression[b]	neural networks	ref
1	1-hexene	−3.23	−3.12	−3.02	9
2	2-heptene	−3.82	−3.75	−3.58	8
3	2-hexen-4-ol	−0.40	−0.68	−0.34	8
4	ethyl isopropyl ether	−0.55	−0.50	−0.51	8
5	hexane	−3.96	−3.97	−3.71	9
6	*p*-xylene	−2.83	−3.17	−3.15	14
7	benzoic acid	−0.78	−0.45	−0.83	12
8	phenylacetic acid	−0.91	−1.13	−1.23	10
9	nitrobenzene	−1.80	−1.66	−2.07	9
10	dipropyl ether	−1.44	−1.11	−1.24	11
11	propyl formate	−0.51	−0.34	−0.32	9
12	2-hexanone	−0.78	−0.90	−0.62	8
13	nitroethane	−0.24	−0.25	−0.13	9
14	isopentyl formate	−1.52	−1.41	−1.43	18
15	methyl benzoate	−1.53	−1.21	−1.75	9
16	ethyl benzoate	−2.22	−1.77	−2.05	11
17	1,4-dimethylnaphthalene	−4.16	−4.73	−4.64	10
18	1-ethylnaphthalene	−4.20	−4.61	−4.24	26
19	benzaldehyde	−1.21	−0.83	−0.96	9
20	acetophenone	−1.34	−0.98	−1.36	9
21	hexanal	−1.30	−1.19	−1.08	18
22	cyclohexanone	0.01	0.07	−0.23	11
23	acetonitrile	0.53	0.12	0.56	11
24	propionitrile	0.33	−0.09	0.26	11
25	butyronitrile	−0.33	−0.57	−0.22	11
26	prednisolone	−3.18	−3.11	−3.22	22
27	hydrocortisone	−2.97	−3.59	−3.03	22
28	phenol	−0.08	0.07	−0.12	25
29	*o*-cresol	−0.65	−0.55	−0.83	25
30	iodobenzene	−2.78	−3.00	−2.94	12
31	1,1,2-trichloroethane	−1.46	−1.64	−1.18	9
32	2-chloropentane	−2.63	−2.66	−2.56	9
33	hexachloroethane	−4.47	−4.37	−4.71	31
34	chloropentafluoroethane	−3.49	−3.08	−3.64	9
35	methyl chloride	−1.00	−1.05	−1.03	9
36	dichloromethane	−0.81	−1.18	−0.95	30
37	2,4′-PCB	−5.07	−4.87	−5.49	24
38	3-chloropropene	−1.28	−1.35	−1.28	9
39	*o*-bromocumene	−4.19	−4.16	−4.14	9
40	1,1,1,2-tetrachloroethane	−2.18	−2.56	−2.29	18
41	1,2-dibromobenzene	−3.50	−3.46	−3.51	22
42	1,3-dibromobenzene	−3.38	−3.60	−3.57	22
43	1,4-difluorobenzene	−1.97	−2.57	−2.11	22
44	1,4-diiodobenzene	−5.25	−4.92	−5.07	22
45	1,3,5-trichlorobenzene	−4.44	−3.84	−3.97	22
46	1,3,5-tribromobenzene	−5.60	−5.04	−5.56	22
47	1,2,3,4-tetrachlorobenzene	−4.25	−4.11	−4.35	27
48	1,2,4,5-tetrafluorobenzene	−2.38	−3.16	−2.63	28
49	1,2,4,5-tetrachlorobenzene	−4.96	−4.56	−5.11	27
50	pentachlorobenzene	−5.28	−5.07	−5.34	29
51	1-fluoro-4-iodobenzene	−3.13	−3.68	−3.34	28
52	1-chloro-3-fluorobenzene	−2.35	−2.47	−2.46	28
53	1-bromo-3-fluorobenzene	−2.67	−2.83	−2.79	28
54	1-bromo-4-chlorobenzene	−3.63	−3.57	−3.40	22
55	1-bromo-4-iodobenzene	−4.56	−4.47	−4.55	22
56	1-chloro-4-iodobenzene	−4.03	−3.99	−3.91	22
57	betamethasone	−3.77	−3.57	−3.82	22
58	progesterone	−4.42	−4.32	−4.51	22
59	propylamine	1.52	1.42	1.61	11
60	diethylamine	1.03	0.96	0.94	11

[a] The above 60 compounds were selected randomly from the 331 compounds used in the studies. The complete Table II is available as supplementary material. [b] Reference 2c.

(14) Sutton, C.; Calder, J. A. *J. Chem. Eng. Data* **1975**, *20*, 320.
(15) Banerjee, S.; Yalkowsky, S. H.; Valvani, S. C. *Environ. Sci. Technol.* **1980**, *14*, 1227.
(16) Hollifield, H. H. *Bull. Environ. Contam. Toxicol.* **1979**, *23*, 579.
(17) May, W. E.; Wasik, S. P.; Freeman, D. H. *Anal. Chem.* **1978**, *50*, 997.
(18) Deno, N. C.; Berkheimer, H. E. *J. Chem. Eng. Data* **1960**, *1*, 1.
(19) Parmelee, H. M. *Refrig. Eng.* **1953**, *61*, 1341.
(20) Park, T.; Rettich, T. R.; Battino, R.; Peterson, D.; Wilhelm, E. *J. Chem. Eng. Data* **1982**, *27*, 324.
(21) Price, L. C. *Am. Assoc. Pet. Geol. Bull.* **1976**, *60*, 213.
(22) Yalkowsky, S. H.; Valvani, S. C. *J. Pharm. Sci.* **1980**, *69*, 912.
(23) Valvani, S. C.; Yalkowsky, S. H.; Roseman, T. J. *J. Pharm. Sci.* **1981**, *70*, 502.
(24) Hutzinger, O.; Safe, S.; Zitko, V. *The Chemistry of PCB's*; *CRC*: Cleveland, OH, 1974. Wallnofer, P. R.; Koniger, M.; Hutzinger, O. *Analabs, Res. Notes* **1973**, *13*, 14.
(25) Valvani, S. C.; Yalkowsky, S. H. *Solubility and Partitioning in Drug Design: Physical Chemical Properties of Drugs*; Yalkowsky, S. H., Sinkula, A. A., Valvani, S. C., Eds.; Marcel Dekker, Inc.: New York and Basel, 1980; pp 201–229.
(26) Yalkowski, S. H.; Valvani, S. C.; Mackay, D. *Residue Rev.* **1983**, *85*, 43–55.
(27) Miller, M. M.; Wasik, S. P.; Huang, G.-L.; Shiu, W.-Y.; Mackay, D. *Environ. Sci. Technol.* **1985**, *19*, 522.
(28) Baker, R. J.; Donelan, B. J.; Peterson, L. J.; Acree, W. E., Jr.; Tsai, C.-c. *Phys. Chem. Liq.* **1987**, *16*, 279.
(29) Garten, C. T., Jr.; Trabalka, J. R. *Environ. Sci. Technol.* **1983**, *17*, 590.
(30) McConnell, G.; Ferguson, D. M.; Pearson, C. R. *Endeavour* **1975**, *34*, 13.
(31) Mackay, D.; Shiu, W. Y. *J. Phys. Chem. Ref. Data* **1981**, *10*, 1175.
(32) Miller, M. M.; Ghodbane, S.; Wasik, S. P.; Tewari, Y. B.; Martire, D. E. *J. Chem. Eng. Data* **1984**, *29*, 184.
(33) Rumelhart, D.; Hinton, G.; Williams, R. *Nature* **1986**, *323*, 533.

Table III. Predicted Results for log W Using Neural Networks and Regression Analysis

	expt	est(NN)[a]	est(RA)[b]	ref
1. 4-heptanol	-1.40	-1.40	-1.61	11
2. menthone	-2.35	-2.72	-2.03	18
3. 1,1-diphenylethylene	-4.52	-5.23	-5.28	18
4. *p*-cresol	-0.81	-0.53	-0.44	25
5. testosterone	-4.08	-4.66	-4.49	22
6. 2,4,4'-PCB	-6.24	-5.96	-5.66	26
7. dexamethasone	-3.59	-3.47	-3.58	22
8. 4-chloronitrobenzene	-2.85	-1.83	-2.66	11
9. 2,5-PCB	-5.06	-5.75	-5.45	27
10. 2,6-PCB	-5.21	-5.52	-5.35	27
11. 2,4,6-PCB	-6.06	-6.24	-5.88	27
12. fluorene	-4.92	-4.43	-4.78	10
13. pyrene	-6.17	-6.04	-6.39	10
14. indan	-3.04	-3.10	-3.27	10
15. 3-methylpyridine	0.04	-0.01	-0.17	11
16. isoquinoline	-1.45	-1.11	-1.24	11
17. tetrahydrofuran	0.48	0.59	0.74	11
18. cortisone	-3.27	-2.95	-3.55	25
19. 2-naphthol	-2.25	-2.08	-1.61	25

[a] Estimation of log W using neural networks which gives a standard deviation 0.43. The standard deviation is 0.37 if we leave out the 4-chloronitrobenzene. [b] Estimation of log W using regression analysis which gives a standard deviation 0.36.

found to be superior to that obtained with the regression analysis approach, 0.30. The results clearly demonstrate that the neural network has captured the association between the selected properties of an organic compound and its aqueous solubility.

The trained neural network was tested on its ability to predict the aqueous solubility of an unknown set of organic compounds, that is, the compounds were not members of the original training set and indeed in some cases were quite unrelated to the original members. The test set should therefore provide a severe test of the neural network's predictive ability. Care should be taken in interpreting the results, however, since strictly the neural network should only be applied to predicting those compounds containing the particular substituents found in the training set. The results obtained are shown in Table III together with the values predicted by the regression analysis technique. Again the performance of the neural network is very satisfactory and compares favorably with that given by the regression analysis method. The neural network gives a predicted aqueous solubility superior to that obtained by regression analysis in 9 of the 19 cases. The poor value predicted for 4-chloronitrobenzene is probably due to the omission from the training set of any chloronitro compound which would reduce the credence attached to the predicted value.

In conclusion, a neural network model has been applied to the prediction of the aqueous solubility of organic compounds and the usefulness of the model clearly demonstrated. The predictive capability of neural networks has been demonstrated on a number of unknown organic compounds. It has been shown in this study that neural networks give a superior performance to that given by a regression analysis technique. While this work was in progress a paper was published[34] describing an application of the neural network approach to estimating quantitative structure-activity relationships. This work confirms the conclusions derived in this study that neural networks can determine such relationships with a performance exceeding that of linear multiregression analysis. Clearly, the neural network approach would seem to have great potential for determining quantitative structure-activity relationships and as such be a valuable tool for the medicinal chemist.

Supplementary Material Available: Listing of complete experimental and estimated log W values (6 pages). Ordering information is given on any current masthead page.

(34) Aoyoma, T.; Suzuki, Y.; Ichikawa, H. *J. Med. Chem.* **1990**, *33*, 2583.

Computational Studies on FK506: Conformational Search and Molecular Dynamics Simulation in Water

Julianto Pranata and William L. Jorgensen*

Contribution from the Department of Chemistry, Yale University, New Haven, Connecticut 06511.
Received May 6, 1991

Abstract: Computational investigations have been undertaken to elucidate the conformational characteristics and the hydration of the immunosuppresant FK506. The calculations made use of the AMBER/OPLS molecular mechanics force field, augmented with some newly developed parameters particularly for the α-ketoamide torsion. A conformational search on FK506 using an internal coordinate Monte Carlo method found 21 distinct energy minima within 12 kcal/mol of the lowest energy structure. The minima include structures with both cis and trans conformations of the amide bond. A 200-ps molecular dynamics simulation in water then provided information on the dynamical behavior of the cis isomer of FK506 as well as its hydration. Two conformations of the macrocyclic ring are sampled during the simulation, and some exocyclic groups undergo rapid conformational changes. Considerable flexibility is also observed near the amide functionality, which is in the binding region of FK506. The hydration of FK506 shows interesting variations owing to differences in the steric environments of potential hydrogen-bonding sites. In the critical binding region, there are on average 5 hydrogen bonds between water molecules and FK506.

FK506, rapamycin, and cyclosporin A (CsA) are immunosuppresive agents that act by blocking the signal transduction pathways that lead to T lymphocyte activation.[1] FK506 and rapamycin are structurally similar and appear to bind to the same receptor, FKBP,[2] while the structurally unrelated CsA, a cyclic undecapeptide, binds to a different receptor, cyclophilin.[3] Both receptors have been shown to be peptidyl-prolyl cis-trans isomerases (rotamases).[4,5] FK506 and rapamycin inhibit the rotamase activity of FKBP, but not of cyclophilin; likewise CsA inhibits the rotamase activity of cyclophilin, but not of FKBP.[4]

(1) Schreiber, S. L. *Science* **1991**, *251*, 283.
(2) Bierer, B. E.; Mattila, P. S.; Standaert, R. F.; Herzenberg, L. A.; Burakoff, S. J.; Crabtree, G.; Schreiber, S. L. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 9231. Dumont, F. J.; Melino, M. R.; Staruch, M. J.; Koprak, S. L.; Fischer, P. A.; Sigal, N. H. *J. Immunol.* **1990**, *144*, 1418. Fretz, H.; Albers, M. W.; Galat, A.; Standaert, R. F.; Lane, W. S.; Burakoff, S. J.; Bierer, B. E.; Schreiber, S. L. *J. Am. Chem. Soc.* **1991**, *113*, 1409.
(3) Handschumacher, R. E.; Harding, M. W.; Rice, J.; Drugge, R. J.; Speicher, D. W. *Science* **1984**, *226*, 544. Handschumacher, R. E.; Harding, M. W. *Transplantation* **1988**, *46*, 29S.
(4) Harding, M. W.; Galat, A.; Uehling, D. E.; Schreiber, S. L. *Nature* **1989**, *341*, 758. Siekierka, J. J.; Hung, S. H. Y.; Poe, M.; Lin, C. S.; Sigal, N. H. *Nature* **1989**, *341*, 755.
(5) Fischer, G.; Wittmann-Liebold, B.; Lang, K.; Kiefhaber, T.; Schmid, F. X. *Nature* **1989**, *226*, 544. Takahashi, N.; Hayano, T.; Suzuki, M. *Nature* **1989**, *226*, 473.

0002-7863/91/1513-9483$02.50/0