ELSEVIER

The inhibition of α-chymotrypsin predicted using theoretically derived molecular properties

Bernd Beck,* Robert C. Glen,†,[1] and Timothy Clark*

**Computer-Chemie-Centrum des Instituts für Organische Chemie der Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany*
†Wellcome Research Laboratories, Beckenham, Kent, U.K.

The structures and molecular properties of 95 aromatic and heteroaromatic ligands previously tested as reversible inhibitors of chymotrypsin catalysis have been calculated using AM1. The properties obtained have been used as input for multiple linear regression analysis and as descriptors for a back-propagation neural network to predict the binding affinity of α-chymotrypsin inhibitors. Using polarizability, molecular shape, electrostatic similarity, dipole moment, ClogP, and the diagonalized quadrupole moments of the ligands, correlation coefficients between calculated and experimental affinities of 0.96 for the training set and 0.89 for the test set were obtained using a neural network. The performance of the multiple linear regression was significantly worse, although useful QSARs were also obtained.

Keywords: α-chymotrypsin, computational chemistry, QSAR, neural networks

INTRODUCTION

The ability of many drugs to interfere with biochemical processes depends on their affinity for specific enzyme active sites. The magnitude of the interaction is a complex relationship that depends on the precise complementarity between the enzyme and the ligand as well as the desolvation of the ligand onto the active site. Conformational changes associated with these interactions may also result in a substantial entropy change on association. There are, however, many cases in which the precise three-dimensional binding conformation of ligands is unknown, so that quantitative structure activity relationships (QSARs) have been derived in order to predict the affinity of novel compounds. Classical QSAR analyses generally use physicochemical substituent constants describing properties of molecules. Recently, more diverse types of data based on computational chemistry methods have been generated.[1]

Here we describe a QSAR study for the prediction of the inhibition strength K_i in the model enzyme system α-chymotrypsin. α-Chymotrypsin catalyzes the hydrolysis of peptide amide bonds on the carboxy side of aromatic amino acids, such as tyrosine, phenylalanine, and tryptophan.[2] Many aromatic compounds inhibit α-chymotrypsin reversibly by binding within the hydrophobic pocket located in the active site, as shown in Fig. 1.

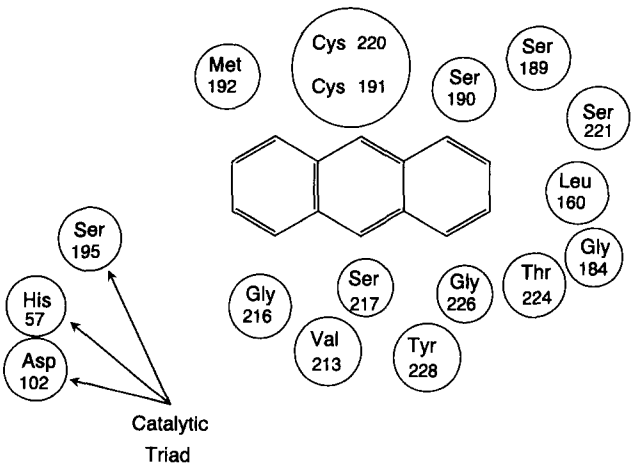

Figure 1. Schematic view of the active site of α-chymotrypsin. An anthracene is shown docked into the cavity to illustrate the pocket size. The catalytic triad is located outside the binding pocket.

Color Plates for this article are on page 142.

Address reprint requests to: Timothy Clark, Computer-Chemie-Centrum, Universität Erlangen-Nürnberg, Nagelbachstrasse 25, D-91052 Erlangen, Germany.
Received 26 March 1996; accepted 19 April 1996.

[1]Present address: Tripos, Inc., 1699 South Hanley Road, St. Louis, Missouri 63144.

655 Avenue of the Americas, New York, NY 10010

0263-7855/96/$15.00
PII S0263-7855(96)00041-0

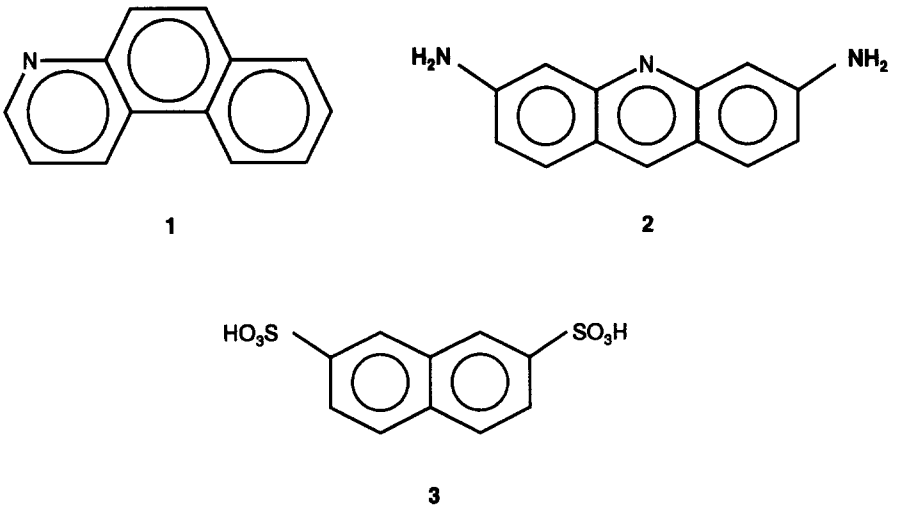

Figure 2. **1,** *Benzo[f]chinoline;* **2,** *2,6-diaminoacridine;* **3,** *2,7-naphthalenedisulfonic acid.*

A survey of 103 simple mono-, bi-, and tricyclic aromatic ring inhibitors has been published.[3] The reported K_i values span four orders of magnitude and range between 63 μM for benzo[*f*]chinoline **(1),** 0.13 mM for 2,6-diaminoacridine **(2),** and 400 mM for 2,7-naphthalenedisulfonic acid **(3).** (see Figure 2).

The starting point for this investigation is the analysis reported by Stewart et al.,[4] in which ligands were docked into the active site of the enzyme with the DOCK software package.[5] The "goodness of fit" was then evaluated according to a Lennard–Jones potential scoring routine and correlated with inhibition of the enzyme. The results suggest that this method is not able to reproduce the experimental K_i values properly.

We have analyzed the same compound set using semiempirical molecular orbital calculations to obtain most of the descriptors. Multiple regression analysis and principal components selection[6,7] were then used to relate the properties to the experimentally determined inhibition strength and to identify which properties are significant. The predictive power of an artificial back-propagation neural network[8] is compared with the regression approach.

METHODS

Calculation of molecular properties

Molecular structures for 95 of 103 compounds were constructed within the SYBYL molecular modeling system[9] on a Silicon Graphics Indigo II. The molecular geometries were optimized using the AM1 Hamiltonian in the semiempirical program package PowerVAMP5.6[10] on an SGI Power Challenge computer. Molecular properties were calculated using PowerVAMP5.6 (AM1) and Daylight Chemical Information software.[11] In all, the 13 properties listed in Table 1[10–13,17,18] were calculated for each compound.

Molecular dipole and quadrupole moments were calculated using the natural atomic orbital–point charge model (NAO–PC)[12,13] included in VAMP5.6. This model describes each heavy atom by nine point charges (including the core charge). Molecular dipoles and quadrupoles are calculated directly from the NAO–PCs using Buckingham's definition[14] within the Born–Oppenheimer approximation[15] for the quadrupoles. To avoid alignment problems, the resulting quadrupole tensor is diagonalized using the Jacobi method.[16] After this step the coordinate system is rotated to

Table 1. Molecular properties calculated for 95 α-chymotrypsin inhibitors

Property	Symbol for property	Reference
Calculated octanol water partition coefficient	ClogP	11
Calculated molar refractivity	CMR	11
Dipole moment components aligned in the diagonalized quadrupole coordinate system	d_x, d_y, d_z	10, 12, 13, 17
Total dipole moment	d_t	10, 12
Eingenvalues of the quardrupole tensor	q_{xx}, q_{yy}, q_{zz}	10, 12, 13, 17
van der Waals surface	WS	10, 11a
van der Waals volume	WV	10, 11a
Mean polarizability	POL	10
Electrostatic similarity	SIM	18

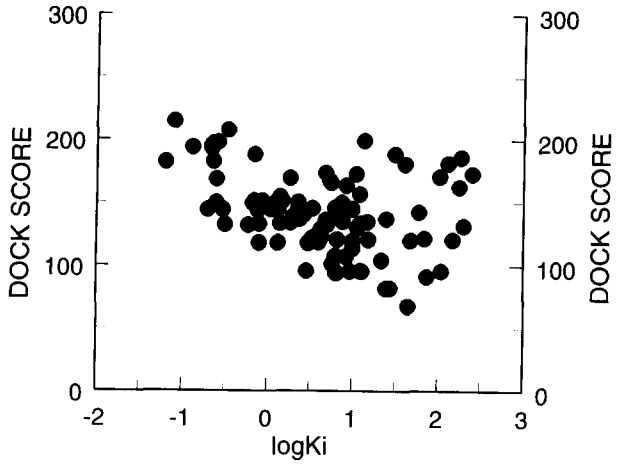

Figure 3. Experimental log K_i values versus the so-called DOCK score. Data taken from Ref. 4.

align the eigenvectors of the quadrupole moment with the Cartesian axes. Finally, the Cartesian components of the dipole moment are calculated within the new coordinate system.[17]

For the calculation of the electrostatic similarity index (SIM)[18] the strongest inhibitor is taken as reference molecule (SIM = 1). In the next step the centers of gravity of the reference and the test molecule are overlaid. Seven starting points are then generated by randomly displacing and twisting the molecular pair. Finally, the electrostatic interactions for each new molecule pair are optimized using the NAO–PCs within a modified "great deluge algorithm" (GDA).[19] This algorithm is similar to the threshold-accepting and simulated annealing methods. The difference lies in the acceptance rule for worse intermediate solutions. For the achievement of best possible performance, a good choice is necessary only for a single parameter. This contrasts with the classical SA algorithm, in which it is necessary to choose carefully a certain sequence of parameters, the so-called annealing schedule.

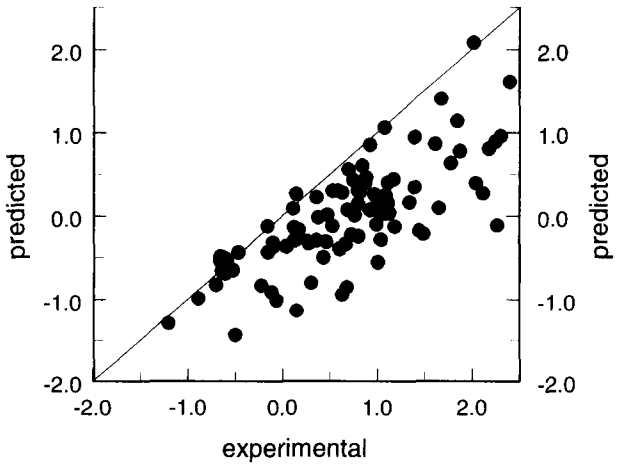

Figure 4. Log K_i versus predicted log K_i using six multipole moments and their square terms as input for a multiple regression analysis.

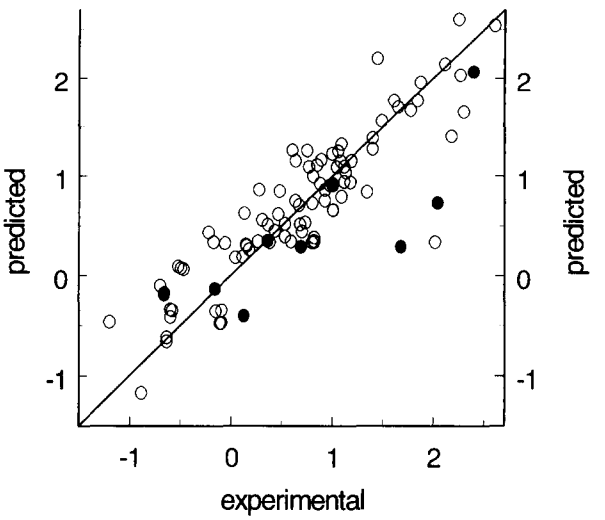

Figure 5. Plot of log K_i against predicted log K_i using a neural net (6:4:1 topology) with six multipole moments as input. Test compounds are denoted by filled circles.

The similarity index (SIM) was calculated as follows:

$$\mathrm{SIM} = 2\frac{E_{12}}{E_1 + E_2} \tag{1}$$

$$E_{12} = \sum_i \sum_{<j} \frac{q_i q_j}{r_{ij}} \tag{2}$$

$$E_1, E_2 = \sum_{i<j}^{N} \frac{q_i q_j}{r_{ij}} \tag{3}$$

where E_{12} is the electrostatic interaction energy between molecules 1 and 2; E_1, E_2 is the electrostatic interaction energy of one molecule with itself; q_i, q_j is the natural atomic orbital–point charges of molecules i and j; and r_{ij} is the distance between point charges, including a 1-Å shift in a fictitious fourth dimension.

Color Plates 1 and 2 show examples for structures overlaid according to their electrostatic similarity. As described

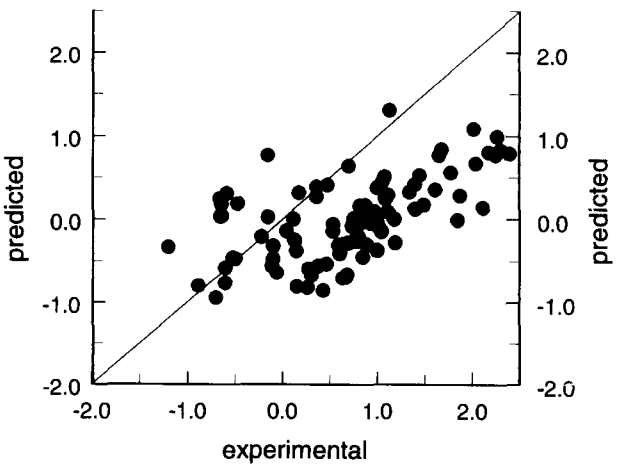

Figure 6. Plot of log K_i against predicted log K_i using Eq. (5).

Table 2. Compounds of the training set and their experimental[a] and predicted log K_i values

Ligand	Log K_i value	
	Experimental	Predicted
1-Aminoacridine	−0.469	−0.723
3,9-Diamino-7-ethoxyacridine	1.130	1.002
2-Styryl-4-aminoquinoline	−0.585	−0.649
3-Aminoacridine	−0.638	−0.693
Acridine	−0.658	−1.071
2-Aminoacridine	−0.658	−0.753
1-Naphthylamine-5-sulfonic acid	1.491	1.562
1-Naphthylamine-4-sulfonic acid	2.267	2.112
5-Azaphenanthrene	−0.638	−0.579
4-Azaphenanthrene	−1.201	−0.626
Naphthalene-2,6-disulfonic acid	2.120	2.120
2-Naphthylamine-1-sulfonate	1.613	1.525
1-Naphthylamine-6-sulfonate	0.681	0.665
2-Naphthylamine-6-sulfonate	1.041	1.082
4-Quinolinecarboxylate	2.017	1.989
Naphthalene-2-sulfonate	0.265	0.588
p-Hydroxybiphenyl	−0.602	−0.767
Benzimidazole-2-carboxylate	0.732	0.790
2-Carboxymethylbenzimidazole	0.748	0.774
Quinoline-4-carboxamide	0.924	0.636
Quinoline-8-sulfonate	2.248	1.979
Quninoline-4-carbinol	1.086	0.438
1,3-Dihydroxynaphthalene	0.146	0.351
2-Aminoquinoline	0.114	0.022
2-Hydroxyquinoline	−0.061	0.154
7-Methylquinoline	−0.155	−0.069
2-Methylquinoline	0.176	−0.112
3-Aminoquinoline	0.362	0.560
2-Naphthylamine	−0.602	−0.475
N-Benzylacetamide	0.875	1.186
Coumarin	0.820	0.778
Propionanilide	0.799	0.574
N-Benzylpropionamide	0.532	0.534
2-Benzimidazol-ethanol	0.130	0.639
N-β-Phenethyl-α-chloroacetamide	0.996	0.875
4-Aminoquinoline	0.041	−0.163
1-Naphthol	−0.699	−0.449
1-Naphthylamine	−0.523	−0.556
8-Hydroxyquinoline	−0.114	−0.143
4-Phenylbutyric acid	1.778	1.992
Ninhydrin hydrate	0.431	0.476
Phthalimidine	0.305	0.674
3-Phenylpropionamide	0.845	1.411
3-Phenylpropionic acid	1.398	1.024
1,3-Indandione	0.380	0.525
Phenoxyethanol	0.887	0.691
Phenylacetamide	1.176	0.926
1-Indanone	0.274	0.337
Acetanilide	1.114	1.210
Phthalide	0.152	0.139
1-Methylindole	−0.097	−0.493
γ-*p*henylbutyramide	1.079	0.971
Quinoxaline	0.699	0.561
Quinoline	−0.222	0.056
Phenylacetic acid	2.301	2.064
N-β-Phenethylacetamide	1.057	1.144
3-Phenyl-1-propanol	0.634	0.724
Benezenesulfonamide	0.634	0.762
N,N-Dimethylaniline	0.532	0.659
Benezenesulfonate	1.845	1.593
N-Ethylaniline	0.820	0.699
Sulfanilamide	1.188	0.987
Benezoic acid	2.176	1.971
Formanilide	0.591	0.814
2-Phenylethanol	0.602	0.851
Indole	−0.097	−0.162
Benzimidazole	0.477	0.396
2,4,6-Trimethylpyridine	1.000	0.876
N-Methylaniline	0.799	0.864
Anisole	0.924	0.983
Benzylamine	1.342	1.029
Benzylalcohol	0.763	0.759
3-Aminopyridine	1.090	0.717
4-Aminopyridine	0.462	0.824
2-Aminopyridine	0.973	0.710
Toluene	1.114	1.031
Phenol	0.806	1.130
Aniline	0.820	0.740
Cyclohexanol	1.875	1.847
Pyridine	1.447	1.545
Benzene	1.398	1.426
Imidazole	1.653	1.686
2,7-Naphthalenedisulfonate	2.602	2.384
Cresol red	0.669	0.808
Fluorescein	1.009	1.050

[a]The K_i [mM] refers to the inhibition constant for the inhibition of the chymotrypsin catalysed hydrolysis of *N*-acetyl-L-valine methyl ester in aqueous solution at pH 7.9.

above, benzo[*f*]chinoline was taken as reference molecule. Color Plates 1 and 2 also show the regions of the largest deviations in the electrostatic potentials of both molecules. The color code is blue/green for negative and red/yellow for positive deviations.

Data analysis

Multiple linear regression was applied to the calculated molecular properties and their square terms and correlated with the measured log K_i values. A back-propagation neural network (SAIC ANSim program, version 2.30[20]) was used. The neural net consists of an input layer, a hidden layer, and an output layer, with each layer comprising a number of neurons. The neurons in each layer are connected, with adjustable weights, to all the neurons in the adjacent layers.

Table 3. Compounds in the test set and their experimental and predicted log K_i values

Ligand	Log K_i value	
	Experimental	Predicted
2,6-Diaminoacridine	-0.886	-0.266
1-Azaphenanthrene	-0.155	-0.319
1-Naphthylamine-8-sulfonate	2.398	1.816
4-Methylquinoline	0.362	-0.178
N-Benzyl-α-chloroacetamide	0.690	0.582
Isoquinoline	-0.495	-0.126
β-Phenethylamine	1.681	0.717
Benzamide	1.000	1.076
7-Azaindole	0.124	0.560
2-Hydroxypyridine	2.041	2.170

The variables were range scaled from −0.5 to 0.5 prior to analysis by back-propagation. A test set containing 10 compounds was extracted at random and the network trained on the remaining 85 compounds. More details on neural networks are given elsewhere.[8,21]

RESULTS

The results reported for α-chymotrypsin–ligand interactions using the DOCK program[5] (version 1.1) are given in Ref. 4. The DOCK score is based on the sum of the van der Waals contacts between ligand and receptor. The energy evaluation function used in determining the DOCK score considers only van der Waals contacts and ignores important interactions, such as hydrogen bonding and electrostatic forces. Figure 3 shows a plot of the DOCK score versus the measured log K_i (data taken from Ref. 4). There is no linear correlation.

Our first approach used only the six calculated multipole moments and their square terms as descriptors. The most significant result, obtained with a nine-term equation [Eq. (4)] (standard errors on the coefficients as quoted), is shown in Figure 4.

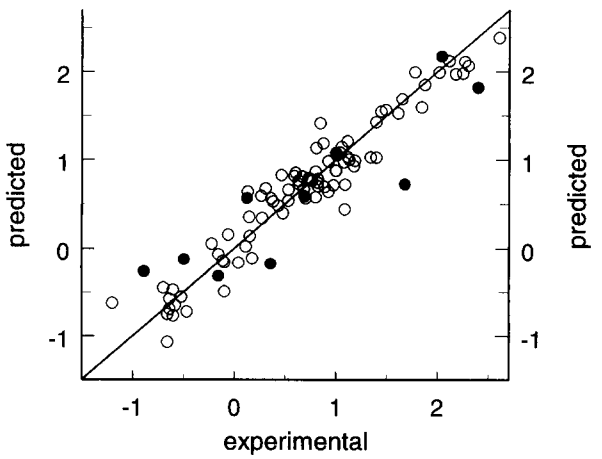

Figure 7. Log K_i *versus predicted log* K_i *using a 9:4:1 back-propagation neural net.*

$$\log K_i = (0.045 \pm 0.039)d_x + (0.073 \pm 0.052)d_z + (0.598 \pm 0.0999)q_{xx} + (0.051 \pm 0.013)d_x^2 + (0.030 \pm 0.009)d_y^2 + (0.071 \pm 0.024)d_z^2 + (0.041 \pm 0.006)q_{xx}^2 - (0.007 \pm 0.003)q_{zz}^2 + (2.249 \pm 0.299) \quad (4)$$

$r = 0.67 \quad F = 8.57$

The correlation coefficient of 0.67 is very low, so that this equation is neither really predictive nor useful.

The neural net trained with the six multipole moments achieved a correlation coefficient of 0.89 (standard deviation of 0.34) for the training set (containing 85 compounds) and 0.81 (0.44) for the test set (containing 10 compounds) (Figure 5).

These results are better than those obtained by the regression analysis or the DOCK program. To improve the predictive power of the neural net, new descriptors such as the similarity index (SIM), the empirically calculated partition coefficient (ClogP), or those providing a description of the molecular shape (WS, WV) were included. Multiple linear regression on all parameters shown in Table 1, allowing some of them to appear as squared terms, gave a significant nine-term equation, Eq. (5), whose performance is shown in Figure 6.

$$\log K_i = (-0.180 \pm 0.088)q_{zz} + (0.058 \pm 0.010)\mathrm{WV} - (0.298 \pm 0.030)\mathrm{POL} - (0.0002 \pm 0.00004)\mathrm{WS}^2 + (0.013 \pm 0.0058)q_{zz}^2 - (2.868 \pm 1.238)\mathrm{SIM}^2 + (0.00017 \pm 0.00009)\mathrm{WV}^2 + (3.347 \pm 1.313)\mathrm{SIM} + (0.1139 \pm 0.068) \quad (5)$$

$r = 0.813 \quad F = 20.97$

Using more than nine terms does not improve the result.

The back-propagation network (8–4–1 architecture) trained with the properties used in Eq. (5) gave a correlation coefficient of 0.90 (standard deviation 0.32) for the training set and 0.86 (0.41) for the test set.

A better result was obtained using a 9:4:1 neural net. The nine input parameters (POL, WV, WS, SIM, d_t, q_{xx}, q_{yy}, q_{zz}, and ClogP) are found using the principal component selection within the program Arthur[7] applied to the calculated properties. We thus obtained correlation coefficients of 0.96 for the training set (Table 2) and 0.89 for the test set (Table 3).

The standard deviations are 0.22 and 0.39 for training and test sets, respectively. The plot of measured versus predicted log K_i is shown in Figure 7. The biggest deviations within the test set are for β-phenethylamine and 2,6-diaminoacridine.

The correlation coefficient for all 95 compounds shown in Figure 7 is 0.95, with a standard deviation of 0.25. The neuron interconnection weights are given in Table 4.

CONCLUSIONS

A series of diverse molecular properties including polarizability, molecular shape, electrostatic similarity, total dipole moment, ClogP and the quadrupole moments appears to be useful in the prediction of affinities for the α-chymotrypsin test system. The application of a neural network gives better results than multiple regression. This combination of molecular properties may be useful in other enzyme or receptor

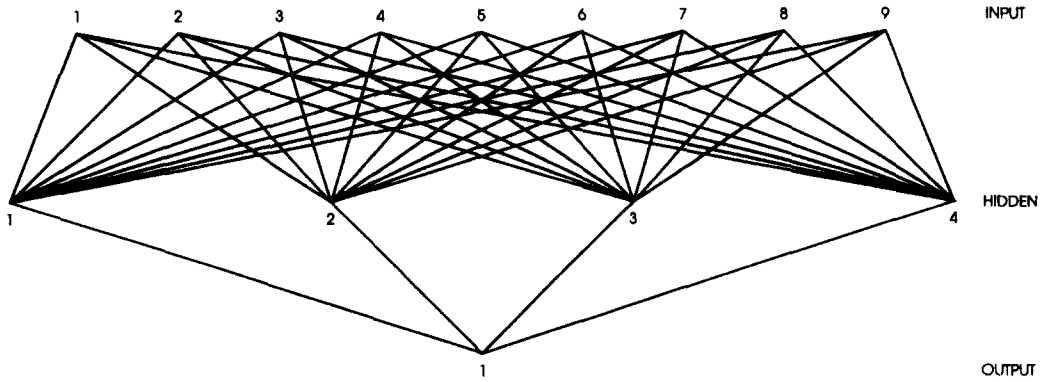

Table 4. Neuron interconnection weights

	POL (1)	WV (2)	SIM (3)	WS (4)	d_t (5)	q_{xx} (6)	q_{yy} (7)	q_{zz} (8)	ClogP (9)
H1	15.672	-1.697	-2.986	2.008	-1.986	-3.416	-3.829	1.184	1.129
H2	-13.148	-11.845	-3.431	-4.903	-9.394	-3.988	-2.143	6.713	0.618
H3	0.693	4.732	-1.383	3.136	11.333	-6.539	-9.309	7.558	-1.640
H4	-0.812	10.442	-0.264	-2.652	-4.578	5.009	-29.882	-2.110	-11.240
	H1		H2		H3		H4		
Output	-12.089		15.076		-8.874		-6.660		
Biases									
Layer 2	12.222		-0.779		-1.217		-0.746		
Layer 3	6.194								

molecules for the prediction of affinity in cases in which a training set of molecules already exists.

ACKNOWLEDGMENT

B.B. thanks the Wellcome Foundation, Ltd. (Beckenham, Kent, U.K.) for financial support.

REFERENCES

1 Hyde, R.M., and Livingstone, D.J. *J. Comput.-Aided Mol. Design* 1988, **2,** 145
2 Blow, D.M. *Acc. Chem. Res.* 1976, **9,** 145
3 Wallace, R.A., Kurz, A.N., and Niemann, C. *Biochemistry,* 1963, **2,** 824
4 Stewart, K.D., Bently, J.A., and Cory, M. DOCKing ligands into receptors: The test case of α-chymotrypsin. *T. C. Methods* 1990, **3,** 713
5 Des Jarlais, R.L., Sheridan, R.P., Seibel, G.L., Dixon, J.S., Kuntz, I.D., and Venkatarghaven, R. *J. Med. Chem.* 1988, **31,** 722
6 Draper, N.R. and Smith, H. *Applied Regression Analysis,* 2nd Ed. John Wiley & Sons, New York, 1981
7 B & B Associates. *Arthur 81,* version 3. B & B Associates, Seattle, Washington, 1990
8 Rumelhart, D.E., and McClelland, J.L. *Parallel Distributed Processing,* Vols. 1 and 2. MIT Press, Cambridge, Massachusetts, 1986
9 Tripos Associates. *SYBYL 6.2.* Tripos Associates, St. Louis, Missouri, 1994
10 Rauhut, G., Alex, A., Chandrasekhar, J., Steinke, T., Sauer, W., Beck, B., and Clark, T. *POWERVAMP,* version 5.6. Oxford Molecular, Ltd., Sandford-on-Thames, Oxford, England, 1995
11 *Pomona89 Physico-Chemical Database & Medchem Software,* version 3.54. Daylight Chemical Information Systems, Inc., Claremont, California, 1989
11a Pascual-Ahuir, J.L., Silla, E., and Tunon, I. *J. Comput. Chem.* 1994, **15,** 1127
12 Rauhut, G., and Clark, T. *J. Comput. Chem.* 1993, **14,** 503
13 Beck, B., Rauhut, G., and Clark, T. *J. Comput. Chem.* 1994, **15,** 1064
14 Buckingham, A.D. *Q. Rev.* 1959, **13,** 189
15 Born, M., and Oppenheimer, J.R. *Ann. Phys.* 1924, **84,** 357
16 Press, W.H. *Numerical Recipes.* University Press, Cambridge, 1986, pp 346–349
17 Beck, B., Rauhut, G., Glen, R.C., and Clark, T. Unpublished results
18 Hutter, M., and Clark, T. Unpublished results
19 Dueck, G. *J. Comput. Phys.* 1993, **104,** 86
20 *SAIC ANSim. SAIC,* San Diego, California, 1989
21 Pao, Y.-H. *Adaptive Pattern Recognition and Neural Networks.* Addison-Wesley, Reading, Massachusetts, 1989