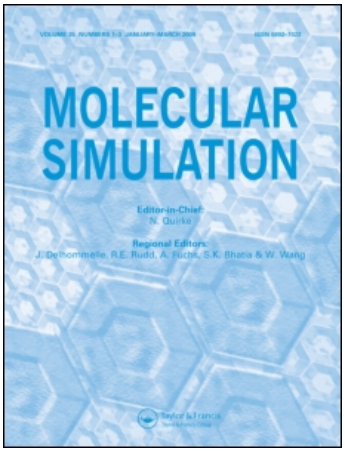

Molecular Simulation
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713644482

Hybrid 3-D QSAR approach using a generalized-regression genetic-neural network

Peter P. Mager[a]
[a] Research Group of Pharmacochemistry, Institute of Pharmacology and Toxicology of the University, Leipzig, Saxony, Germany

Online publication date: 26 October 2010

To cite this Article Mager, Peter P.(2002) 'Hybrid 3-D QSAR approach using a generalized-regression genetic-neural network', Molecular Simulation, 28: 4, 377 — 384
To link to this Article: DOI: 10.1080/08927020290018741
URL: http://dx.doi.org/10.1080/08927020290018741

Molecular Simulation, 2002 VOL. 28 (4), pp. 377–384

HYBRID 3-D QSAR APPROACH USING A GENERALIZED-REGRESSION GENETIC-NEURAL NETWORK

PETER P. MAGER

Research Group of Pharmacochemistry, Institute of Pharmacology and Toxicology of the University, D-04107 Leipzig, Härtelstr. 16-18, Saxony, Germany

(Received May 2001; Accepted June 2001)

The advantage of generalized-regression genetic-neural networks over backpropagation networks is that relevant physicochemical descriptors can be selected, somewhat reminiscent of the results from classical quantitative structure-activity relationships (QSARs). The descriptor matrix is based on a hybrid model of traditional QSAR (use of extra-thermodynamic parameters) and 3-D QSAR. The model was tested using molecular simulation of a series of A_3 adenosine receptor agonists (1,4-dihydropyridine derivatives).

Keywords: Molecular simulation; Generalized-regression genetic-neural network; Backpropagation network; Quantitative structure-activity relationship (QSAR); 3-D QSAR; Comparative molecular field analysis (CoMFA)

1. INTRODUCTION

Adenosine enters cells by facilitated diffusion via a single, nonspecific carrier system that is very sensitive to inhibition by adenosine derivatives. The synthetic analogs modulate neuronal function by receptor-mediated mechanisms. It was found that there are three major subtypes of receptors, called A_1, A_2, and A_3 [1–4]. The receptor types have been cloned and characterized as belonging to a superfamily of receptors with seven transmembrane helices that couple to G proteins [5].

ISSN 0892-7022
DOI: 10.1080/08927020290018741

In this paper, the results obtained by applying a generalized-regression genetic neural network to A_3 receptor agonists will be reported.

2. METHOD

2.1. Synthesis

The synthetic route of the compounds (Fig. 1, Table I) was described previously [2], together with all data that verify the structures.

2.2. Physicochemical Data

To investigate the physicochemical properties, the following constants were determined: The partition coefficients P (octanol/water) at pH 7.4; the basic pK_a values (at $[K^+] + [Na^+] = 0.05$ M, $[Cl^-] = 0.05$ M; 25°C); the mass (MG), and the 1,4 non-valency energy E (1.4 NV) (kcal/mol) determined by geometry optimization using a combined CNDO/MM+ force-field approach [3]. The variables to be studied were:

$$X_1 = \ln E(1.4\,\mathrm{NV}), \quad X_2 = \ln MG, \quad X_3 = \text{basic p}K_a, \quad X_4 = \log P.$$

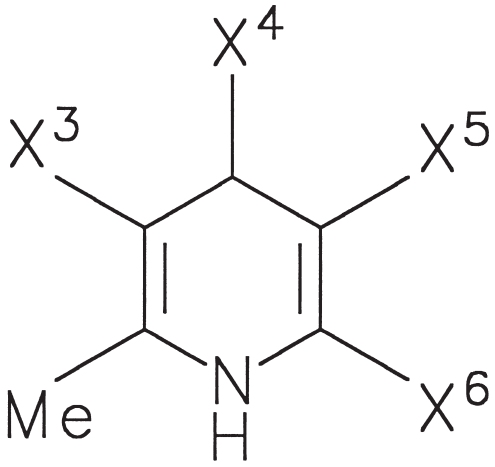

FIGURE 1 Lead structure of 1,4-dihydropyridine and pyridine derivatives.

TABLE I Code of the compounds and substituents of the lead structure (Fig. 1)

Compound	*Substituents*			
	X^3	X^4	X^5	X^6
1	CO_2Me	Me	CO_2Me	Me
2	CO_2Me	Me	$CO_2(CH_2)OMe$	Me
3	CO_2Me	Me	CO_2CH_2Ph	Me
4	CO_2Et	Me	$CO_2CH_2CH_2SPh$	Me
5	CO_2Me	Et	CO_2Et	Me
7	CO_2Me	$CH_2CHCH_3(CH_2)_2CH{=}CMe_2$	CO_2Et	Me
8	CO_2Me	Ph	CO_2Et	Me
9	CO_2Me	Ph-2-NO_2	CO_2Me	Me
10	CO_2Me	Ph-3-NO_2	CO_2Et	Me
11	CO_2Et	Ph-3-NO_2	CO_2Et	Me
12	CO_2Me	Ph-3-NO_2	$CO_2(CH2)_2N(Me)CH_2Ph$	Me
13	CO_2-*i*-Pr	Ph-3-NO_2	$CO_2(CH)_2OMe$	Me
14*	CO_2Me	Ph-3-NO_2	$CO_2(CH_2)_3$-AAA	Me
15	CO_2Me	Ph-4-NO_2	CO_2Et	Me
16	CO_2Me	Ph-2-CF_3	CO_2Et	Me
17	CO_2Me	Ph-2-CF_3	NO_2	Me
18	CO_2Me	Ph-4-OMe	CO_2Et	Me
19	CO_2Me	Ph-3-OMe,4-OH	CO_2Et	Me
20	CO_2Me	OPh-3,4-OCH	CO_2Et	Me
21	CO_2Me	$(CH_2)_2Ph$	CO_2Et	Me
22	CO_2Me	CH=CHPh (*trans*)	CO_2Et	Me
23	CO_2Me	*CH ≡ CHPH*	CO_2Et	Me
24	CO_2Et	Me	CO_2Et	*n*-Pr
25	CO_2Et	Me	CO_2Et	Ph
26	CO_2Et	CH=CHPh (*trans*)	CO_2Et	Ph

**AAA = C–N$(CH_2CH_2)_2$ − 4 − Ph, 4′ − Ph.

The variables are a mixture of extra-thermodynamic parameters (X_3, X_4) that are usually applied in simulating an ordinary quantitative structure-activity relationship (QSAR) analysis, and a 3-D QSAR (X_1) on the other.

2.3. Pharmacological Data

Affinity of cloned human A_3 adenosine receptors expressed in HEK-293 cells was determined using N^6-(4-amino-3-[^{125}I]iodobenzyl)-5′-(*N*-methylcarbamoyl)-adenosine [2]. The affinity constant K_i (nM) was transformed by its natural logarithm

$$Y = \ln K_i$$

to get normalized and variance-stabilized scores.

2.4. Neural Network [6]

The selection of relevant descriptors occurred by a generalized-regression genetic-neural network (GRNN) was described elsewhere [7]. The neural network architectures, learning algorithms, and genetic selection procedures were based on the proposed default parameters of the NeuroGenetic Optimizer program (BioComp Systems, Redmond, WA). As special parameters, the following layers were used: one input layer with linear transfer function and 4 nodes (physicochemical descriptors); two hidden layer with sigmoidal and summation transfer functions having each 4 nodes; and one output layer with direct transfer function and 1 node (biological activity). The following net parameters were used: generations run = 10, population size = 42, minimum network training passes for each network = 20, cutoff for network training passes = 50, input neural node influence factor = 0, hidden neural node influence factor = 0, limit on hidden neurons = 8; selection was performed by the top 50% surviving, refilling of the population was done by cloning the

TABLE II Design matrix. Notations see "Methods"

Compound	*A_3 activity*	*Chemical descriptors*			
	Y	X_1	X_2	X_3	X_4
1	3.48	2.17	5.53	0.93	2.31
2	4.13	2.35	5.65	0.74	1.82
3	1.02	2.73	5.75	0.68	3.45
4	1.72	2.22	5.93	0.74	4.62
5	2.61	1.72	5.59	1.03	2.83
6	1.87	2.04	5.68	1.05	3.34
7	2.07	2.18	5.89	0.93	6.61
8	2.48	2.11	5.75	0.18	3.45
9	2.12	2.59	5.85	0.17	2.67
10	2.12	2.32	5.89	−0.28	3.19
11	0.92	1.85	5.92	−0.66	3.70
12	1.18	3.04	6.17	7.24	5.00
13	2.14	2.60	6.04	−1.01	3.74
14	0.64	3.16	6.41	8.62	7.27
15	1.78	2.25	5.89	−0.33	3.19
16	2.45	2.38	5.95	−0.12	6.54
17	3.16	2.45	5.87	−3.00	5.39
18	1.41	2.19	5.84	−1.35	4.67
19	3.47	2.54	5.89	−1.35	3.70
20	1.52	2.52	5.88	−1.35	3.86
21	0.83	2.17	5.84	0.82	4.48
22	−0.40	3.13	6.07	0.52	4.41
23	−0.06	1.96	5.83	−0.42	4.04
24	3.85	1.86	5.73	0.61	4.38
25	1.98	2.22	5.80	−0.53	4.25
26	−2.23	2.61	6.03	−0.94	6.36

survivors, mating was performed by using the TailSwap technique (the system picks up a cut point and exchanges "genetic material" between the cut point and the end of the string of the "parents", essentially swapping tails). Mutations were performed using the random bit exchange technique at a rate of 25%. The squared multiple correlation coefficients (R^2) were estimated and tested by simultaneous statistical inference [8].

3. RESULTS AND DISCUSSION

The substituents of the 1,4-dihydropyridine and pyridine derivatives (Fig. 1) are listed in Table I. The design matrix of the chemical parameters is given in Table II.

The variables X_1 (ln of the 1,4 non-valency energy) and X_4 (log of distribution coefficient) were selected by genetic algorithm, the two other descriptors were omitted from the final network. The squared multiple correlation coefficient $R^2 = 0.956$ is significantly different from zero at the 1% level or less. The correlation plot of the experimentally obtained and theoretically calculated data is illustrated in Fig. 2.

The general role of lipophilicity in drug-receptor interactions is well known from classical quantitative structure-activity relationships (QSARs) of Hansch [9,10]. From comparative molecular field analysis (CoMFA) or 3-D QSAR, the role of energies is known [11]. Significance of the 1,4 non-valency energy (van-der-Waals, electrostatic, hydrogen bond energies) may be interpreted as an indication that a high contribution of 1,4 interaction energy of centers separated only by three bonds, is important to model receptor affinity.

On the other hand, the van-der-Waals energy, electrostatic energy, and lipophilicity are commonly involved in drug-receptor interactions of A_1 and A_2 ligands, while a discrimination of A_1 and A_2 receptor sites is only based on steric features of the substituents [3,4,12,13]. Therefore, at the present-day state of knowledge, it must be concluded that a molecular discrimination of the adenosine receptor subtypes A_1, A_2, and A_3 structures is difficult, if not impossible. Some of the problems have also been attributed to the multisubstrate action of adenosine receptors [13].

Nevertheless, it is a certain relief that GRNN has the ability to approximate structure-activity functions with satisfactory accuracy. Like CoMFA and 3-D QSAR, the outcome of a neural network analysis will be an internally highly self-consistent result.

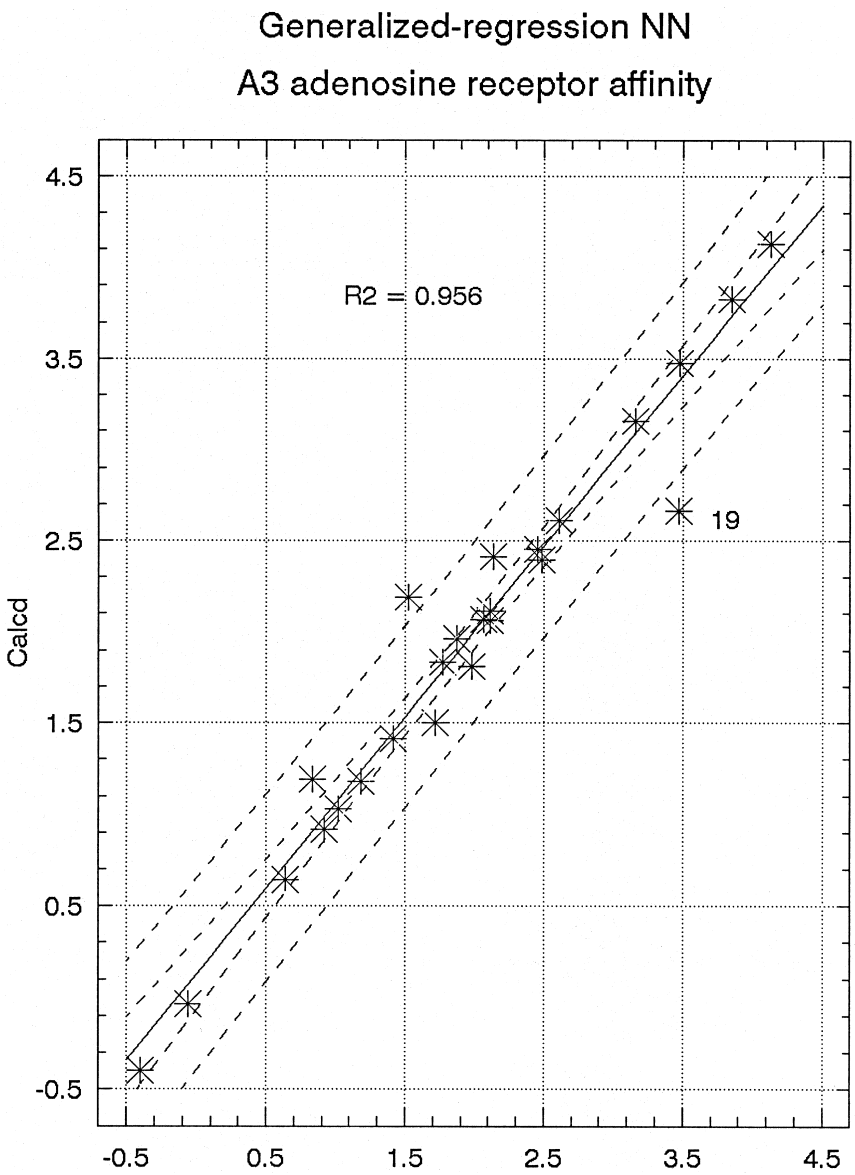

FIGURE 2 Plot of the experimentally obtained and theoretically calculated data.

4. CONCLUSIONS

A combination of the genetic neural network with the generalized-regression approach was tested using a series of A_3 adenosine receptor agonists. An advantage of genetic neural over backpropagation networks is that the physicochemical descriptors can be selected, somewhat reminiscent of the results from traditional QSAR.

However, the selection is made without the manifold assumptions that must approximately be satisfied for hypothesis testing of statistically based QSAR approaches. On the other hand, the molecular-simulation design matrix may be regarded as hybrid model of traditional QSAR (e.g. use of the distribution coefficient) and CoMFA resp. 3-D QSAR (e.g., use of energies). Nevertheless, subsequent research is needed to examine the role of the generalized-regression genetic-neural networks in molecular simulation of drugs.

References

[1] Colotta, V., Catarzi, D., Varano, F., Cecchi, L., Filacchioni, G., Martini, C., Trincavelli, L. and Lucacchini, A. (2000) "Synthesis and structure-activity relationships of a new set of 2-arylpyrazolo[3,4-c]quinoline derivatives as adenosine receptor antagonists", *J. Med. Chem.* **10**, 3118–3124.

[2] van Rhee, A.M., Jiang, J., Melman, N., Olah, M.E., Stiles, G.L. and Jacobson, K.A. (1996) "Interaction of 1,4-dihydropyridine and pyridine derivatives with adenosine receptors: selectivity of A_3 receptors", *J. Med. Chem.* **39**, 2980–2989.

[3] Mager, P.P., Richter, M. and Reinhardt, R. (1995) "Molecular simulation applied to 2-(N′-alkylidenehydrazino)- and 2-(N′-aralkylidenehydrazino)adenosine A_2 agonists", *Eur. J. Med. Chem.* **30**, 15–25.

[4] Mager, P.P. (1998) "Neural network approaches applied to selective A_{2a} adenosine receptor agonists", *Med. Chem. Res.* **8**, 277–290.

[5] Fredholm, B.B., Gerwins, P., Johansson, B., Parkinson, F.E. and van der Ploeg, I. (1993) "Signalling via G-Protein coupled receptors—using adenosine receptors as an example", *Drug Design and Discovery* **9**, 189–197.

[6] Schneider, G. (2000) "Neural networks are useful tools for drug design", *Neural Netw.* **13**, 15–60.

[7] Mager, P.P. and Reinhardt, R. (1999) "A comparison of backpropagation and generalized-regression genetic-neural network models", *Drug Design and Discovery* **16**, 49–53.

[8] Mager, P.P. (1991) Design statistics in Pharmacochemistry (Wiley, New York).

[9] Fujita, T. (1990) "The extrathermodynamic approach to drug design", *Compr. Med. Chem.* **4**, 497–560.

[10] Hansch, C. and Leo, A. (1995) Exploring QSAR: Fundamentals and Applications in Chemistry and Biology (Am. Chem. Soc., Washington, DC).

[11] Cramer, R.D., Patterson, D.E. and Abraham, D.J. (1988) "Comparative molecular field analysis (CoMFA)", *J. Am. Chem. Soc.* **110**, 5959–5967.

[12] Mager, P.P., Reinhardt, R., Richter, M., Walther, H. and Rockel, B. (1995) "Molecular simulation of 8-styrylxanthines", *Drug Design and Discovery* **13**, 89–107.

[13] Mager, P.P. (1994) "A_2 agonists: structure-activity relationships of 2-(cycloalkylalkynyl)adenosine derivatives", *Eur. J. Med.* **29**, 369–380.

[14] Mager, P.P. (1996) "A random numbers experiment to simulate resample model evaluation", *J. Chemometrics* **10**, 221–240.

APPENDIX: TECHNICAL NOTE

A proposal to test the predictive model power by sequential resampling was not realized. First, cross-validation with, e.g. by using 80% of the data as training and 20% as test set, was not made because it was shown by a random-number simulation experiment that cross-validation requires approximately *equal* sample sizes of the subgroups [14]. Second, subsamples of one-leaving-out procedures must be taken *randomly* from the same finite population; there can be no difference between the subsamples other than for sampling errors, because all possible subsamples are taken from a population in such a way that they have the same probability of being selected [14]. The gain in precision is illusory in most cases. Sequential resampling may be useful for diagnostic statistics which prove the assumptions of the underlying theory of hypothesis testing, however. Also, the root-mean squared error and related measures were not applied because their significance cannot be examined by exact significance tests. As far, genetic algorithms of an evaluation of chemical descriptors may be an alternative to the selection procedures of traditional QSAR.