Eur. J. Med. Chem. 33 (1998) 647–658

QSAR/QSTR of fluoroquinolones: an example of simultaneous analysis of multiple biological activities using neural network method

Yun Tang*, Kai-Xian Chen, Hua-Liang Jiang, Ru-Yun Ji

State Key Laboratory of Drug Research, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, 294 Taiyuan Road, Shanghai 200031, China

(Received 29 October 1997; accepted 9 March 1998)

Abstract – In this paper, a three-layer three-output-node back-propagation neural network was applied to analyze the QSAR and QSTR of 23 5,8-disubstituted fluoroquinolones simultaneously. Four descriptors, q_{N1}, q_{O4}, σ_{m5}, and MR_8, were selected from a set of quantum chemical indices and physicochemical parameters using PLS method. The results obtained from neural networks showed that the 4 descriptors were correlated significantly with both the antibacterial activity and cytotoxicity and a multiple QSAR model was built. It provided useful information to develop new selectively potent fluoroquinolones with high activity and low toxicity. Meanwhile, the study demonstrated that the multi-output-node neural network was a powerful tool to analyze multiple QSAR and superior to PLS method. © Elsevier, Paris

quantitative structure–activity relationship (QSAR) / quantitative structure–toxicity relationship (QSTR) / neural networks / partial least squares (PLS) / fluoroquinolone

1. Introduction

Quantitative structure–activity relationship (QSAR) analysis plays crucial roles in the elucidation of drug action mechanisms and the design of more potent drugs [1]. When activity is replaced by toxicity in the classical QSAR model, it becomes the so-called quantitative structure–toxicity relationship (QSTR) analysis, which is also useful for the improvement of side-effects of drugs. Antitumor, antibacterial, or antiviral drugs, besides their anti-parasite activities, usually have some cytotoxicities to their host organisms, which limit their clinical applications. How to reduce toxicity while keeping or improving activity, that is, how to improve the selectivity of these drugs, is an urgent problem. If only QSAR is analyzed to improve the activity of drugs, the cytotoxicity may increase as well. However, if QSTR is analyzed at the same time, more useful information could be obtained to instruct the modification of the lead compound. Therefore, a multiple QSAR model is proposed to solve such a problem, in which some parallel properties of the compounds, such as activity and toxicity, are assumed as the dependent variables and correlated with the same set of independent descriptors.

Neural network is a computer-based data processing system derived from the simplified concept of the brain, in which a number of nodes, called processing elements or neurons, are interconnected in a network-like structure [2]. The applications of neural network in QSAR analysis have been explored widely since 1990 [3–8]. This method is believed to be superior to the traditional Hansch approach [9] and gives a great impetus to the development of QSAR. However, except a report of the application of neural networks to multiple responses [10], most of these studies only applied a single-output-node neural network in QSAR analysis, that is, using the biological activity as the single output. If we extend the single-output-node neural network to a multi-output-node one, the above-mentioned multiple QSAR problem then could be analyzed simultaneously, which is very helpful for drug design.

*Corresponding author
Present address: Department of Biosciences at Novum, Center for Structural Biochemistry, Karolinska Institute, S-14157 Huddinge, Sweden.

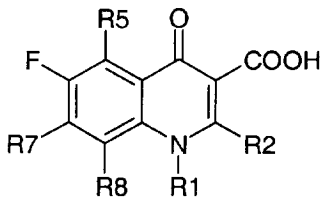


Figure 1. The common structure of fluoroquinolones. Norfloxacin: R1 = C_2H_5, R2 = R5 = R8 = H, R7 = piperazine.

In this study, 23 5,8-disubstituted fluoroquinolones selected from literature [11] served as an example to build a multiple QSAR model using the multi-output-node neural network method. Fluoroquinolones (*figure 1*) are potent antibacterials with a broad spectrum of activities against Gram-positive and Gram-negative bacteria and mycobacteria by inhibiting bacterial topoisomerase II (DNA gyrase) [12]. However, it was reported recently that fluoroquinolones could also inhibit eukaryotic topoisomerases, which could possibly lead to clastogenicity and/or cellular toxicity to the host [13, 14]. Therefore, it is necessary to study the structure-cytotoxicity relationships of fluoroquinolones to develop safer and more potent derivatives. Many QSAR analyses have been performed on fluoroquinolones [15, 16], especially substituent in the 1- or 7-position, which led to the discovery of norfloxacin (cf. *figure 1*). However, except for a few qualitative analyses of the structure-toxicity relationships [11, 17, 18], no quantitative analysis and comparison of the structural features of fluoroquinolones related to the observed cytotoxicity and antibacterial activity were reported. The aim of this paper is to analyze and compare the QSAR and QSTR of fluoroquinolones, and provide an approach to the simultaneous analysis of multiple biological activities.

2. Results

2.1. Variable selection

In the original paper [11], a total of 116 fluoroquinolones with various substituents at the 1-, 5-, 7-, or 8-positions have been tested by Suto et al. for their mammalian cell cytotoxicity (clonogenic assay of cellular survival) and antibacterial activities against both Gram-positive and Gram-negative bacteria. Among the 116 fluoroquinolones, 23 analogs, only changeable at 5-, 8-positions while keeping a cyclopropyl group at 1-position and a 3′-aminopyrrolidinyl group at 7-position constant, were selected as the object of this study (*table I*). Because no structural variables were available in the original paper, the possible structural and property descriptors correlated with both activity and toxicity should be determined above all. This work was carried out on the Silicon Graphics IRIS 4D/310VGX workstation using the SYBYL V6.1 software package [19].

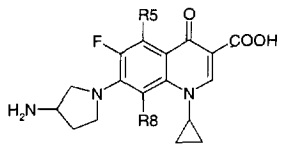
At first, the three-dimensional structure of each compound was built up using normal bond-length and bond-angle in the SYBYL/Base module. After conformational search and energy minimization, each structure was optimized by the semiempirical quantum chemical program MOPAC V6.0 [20] with AM1 (Austin method 1) Hamiltonian. Therefore a series of quantum chemical indices were obtained, such as the HOMO energy E_{HOMO}, the LUMO energy E_{LUMO}, and net atomic charges of nitrogen at 1-position q_{N1}, oxygen at 4-position q_{O4}, the center atom of the group at 5-position q_{C5}, fluorine at 6-position q_{F6}, the terminal-N of the group at 7-position q_{N7}, and the center atom of the group at 8-position q_{C8}. Besides these indices, we also selected a number of physicochemical parameters of the groups at 5,8-position from literature [21], such as hydrophobic constant π_5 and π_8, molar refraction MR_5 and MR_8, electronic constant σ_{m5}, σ_{m8}, σ_{p5}, and σ_{p8}. A total of 16 possible variables were obtained.

The 3 experimentally biological activities were expressed as the concentration forms in μg/mL, that is, median inhibitory concentration (IC_{50}) for the cytotoxicity and minimum inhibitory concentration (MIC) for the two antibacterial activities which are the average activities against 5 Gram-negative and 5 Gram-positive bacteria, respectively. For the multiple QSAR study, we transformed them into corresponding logarithm forms log(1/C) (here C stands for IC_{50} or MIC, in mol/L).

In order to determine which variables were significantly correlated with both activity and toxicity, the PLS method [22] was used as the screening tool. PLS is generally used in handling multiple linear regression problems with latent variables. The latent variables were applied to describe the maximum predictive variance of a data set, which provided maximal fit to the model. By only using several significant latent variables in the procedure, a noise filtering effect was obtained which resulted in an improved predictive ability of PLS. A cross-validation technique [23] was used in PLS to select the optimal number of components, which was then applied to obtain a predictive PLS model. A squared predictive correlation coefficient R^2 was given as a statistical index.

Here, the PLS program installed within the SYBYL QSAR/CoMFA module was used to select the variables from the 16 possible descriptors using leave-one-out

Table I. The compounds used in this study and their observed biological activities (μg/mL). [a]



Compound	R5	R8	Mammalian cell cytotoxicity		Antibacterial activity (Gram$^-$)		Antibacterial activity (Gram$^+$)	
			IC_{50}	log(1/C)	MIC	log(1/C)	MIC	log(1/C)
1	H	F	30	4.07	0.04	6.94	0.03	7.07
2	H	H	72	3.66	0.09	6.57	0.14	6.37
3	H	CF_3	300	3.12	0.20	6.30	0.20	6.30
4	H	Cl	26	4.15	0.03	7.09	0.04	6.96
5	H	OMe	44	3.91	0.07	6.71	0.03	7.08
6	H	NO_2	69	3.74	0.46	5.91	0.92	5.61
7	H	NH_2	260	3.12	0.53	5.82	2.06	5.23
8	H	SMe	31	4.09	0.20	6.28	0.06	6.80
9	NH_2	H	54	3.81	0.04	6.94	0.05	6.84
10	Me	H	14	4.39	0.09	6.58	0.05	6.84
11	Et	H	46	3.89	1.81	5.30	>2.71	5.12
12	OH	H	72	3.68	0.04	6.94	0.09	6.59
13	SMe	H	92	3.61	>1.58	5.38	>3.1	5.09
14	F	F	100	3.57	0.08	6.66	0.37	6.00
15	F	CF_3	190	3.34	25	4.22	25	4.22
16	NH_2	F	15	4.39	0.01	7.56	0.01	7.56
17	NH_2	Cl	59	3.81	0.08	6.68	0.05	6.88
18	NH_2	OMe	100	3.58	0.11	6.53	0.04	6.97
19	Me	CF_3	290	3.15	0.91	5.66	0.79	5.72
20	Me	F	11	4.52	0.02	7.26	0.008	7.66
21	OH	F	43	3.93	0.03	7.09	0.05	6.86
22	OMe	Cl	> 500	2.90	2.08	5.28	12.48	4.50
23	Cl	Cl	> 500	2.90	0.13	6.49	0.15	6.43

[a] All values with the '>' symbol adopt the same values after the '>' symbol.

cross-validated R^2 as the screening criterion. At first one-by-one analysis of each descriptor was performed, then combinatorial analysis of 2, 3 or 4 descriptors was processed. Finally, 4 descriptors with high R^2 values, i.e. net atomic charges of nitrogen at 1-position q_{N1} and oxygen at 4-position q_{O4}, electronic constant of the group at 5-position σ_{m5}, and molar refraction of the group at 8-position MR_8, were selected as the independent variables (*table II*).

2.2. *Neural network analysis*

This work was performed on AST PII486/33 personal computer using the back-propagation neural network program written by ourselves. The basic operation of the back-propagation neural networks was described in detail elsewhere [24, 25]. Here we simply emphasize several points about our program: the three parameters, learning coefficient η, momentum factor α and gain in the sigmoid function β, can take values between 0 and 1, also can take different values between the input:hidden layer and the hidden: output layer to accelerate the convergence, designated by η1 and η2, α1 and α2, β1 and β2, respectively; every input and target output is scaled between 0.1 and 0.9, and the target and actual outputs are rescaled according to their original values to analyze the results after the task is finished; the connected weights are initialized with random values between −0.3 and 0.3; the quality of QSAR is assessed by 2 statistical indices, the correlation coefficient R^2 and the residual variance (RV), respectively.

Table II. The four descriptors selected by PLS method. [a]

Compound	q_{N1}	q_{O4}	σ_{m5}	MR_8
1	−0.146	−0.346	0.00	0.92
2	−0.153	−0.350	0.00	1.03
3	−0.156	−0.338	0.00	5.02
4	−0.151	−0.344	0.00	6.03
5	−0.143	−0.350	0.00	7.87
6	−0.154	−0.340	0.00	7.36
7	−0.144	−0.352	0.00	5.42
8	−0.146	−0.351	0.00	13.82
9	−0.156	−0.398	−0.16	1.03
10	−0.154	−0.357	−0.07	1.03
11	−0.154	−0.360	−0.07	1.03
12	−0.151	−0.393	0.12	1.03
13	−0.158	−0.335	0.15	1.03
14	−0.150	−0.325	0.34	0.92
15	−0.161	−0.319	0.34	5.02
16	−0.150	−0.393	−0.16	0.92
17	−0.155	−0.393	−0.16	6.03
18	−0.148	−0.396	−0.16	7.87
19	−0.160	−0.346	−0.07	5.02
20	−0.149	−0.352	−0.07	0.92
21	−0.145	−0.389	0.12	0.92
22	−0.158	−0.329	0.12	6.03
23	−0.156	−0.324	0.37	6.03

[a] q_{N1}, q_{O4} came from quantum chemical calculations with the AM1 method; σ_{m5}, MR_8 were adopted from literature [20].

The 4 independent variables selected with PLS method were regarded as the inputs. The 3 biological activities, i.e. the mammalian cell cytotoxicity $\log(1/IC_{50})$ and the 2 antibacterial activities log(1/MIC) ($Gram^-$ and $Gram^+$ stand for Gram-negative and Gram-positive bacteria, respectively), were regarded as the target outputs. Namely, the input nodes were set to be 4 and the output nodes were 3. The number of hidden nodes was defined by evaluation. Through trial and error, the number of hidden nodes was defined as 4. The configuration of neural network used here was shown in *figure 2*.

The stopped training method [26, 27] was used to train the neural networks. The maximum iterative number was set at 20000, and other parameters were set as following: $\eta 1 = 0.3$, $\eta 2 = 0.25$, $\alpha 1 = 0.75$, $\alpha 2 = 0.70$, $\beta 1 = 1.0$, and $\beta 2 = 1.0$. Under these conditions, a multiple QSAR model was established. *Table III* showed the multiple QSAR study results. The values of R^2 were 0.853, 0.843 and 0.883 for the cytotoxicity, the activities against Gram-negative and Gram-positive bacteria, respectively.

The leave-one-out cross-validation procedure [23] was then carried out to test the predictive ability of the multiple QSAR model just constructed. In this process, one compound was removed from the data set, and the

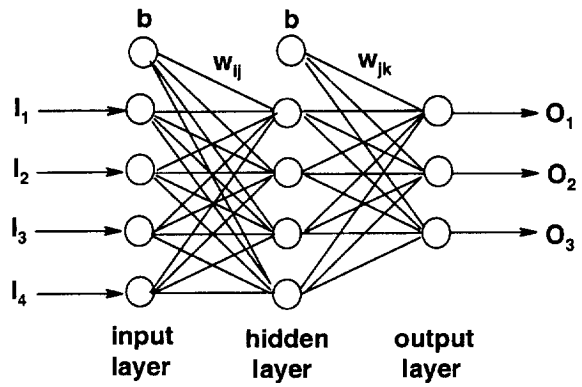


Figure 2. A model of three-layer back-propagation neural network with 4-4-3 configuration used in this study, in which b is bias.

remaining 22 compounds were served as the training set. After trained, the 3 biological activities of the removed compound were predicted from the network. This procedure was repeated 23 times until the predicted biological activities of the entire data set were obtained (*table IV*). The values of R^2 were 0.411, 0.649 and 0.602 for the cytotoxicity, the activities against Gram-negative and Gram-positive bacteria, respectively.

Based on the above learning pattern, we also investigated the relationships between the 4 descriptors and the 3 biological activities [4, 5]. Changing the value of 1 input while keeping the remaining 3 inputs constant at the 1/3 of their maximal ranges, the functional relationships of the two antibacterial activities and cytotoxicity with the changed index were monitored (*figure 3*).

From the multiple QSAR analysis and the graphical relationships of the biological variables with the descriptors, several new analogs were derived. At first some appropriate substituents were obtained for the 5, 8-positions to meet the needs of their parameters, and a semiempirical quantum chemical calculation was performed on each new compound to get the corresponding net atomic charges of N1 and O4. Then the cytotoxicity and antibacterial activities of each analog were predicted using the above trained network. The results were shown in *table V*.

In comparison with the multiple QSAR analysis of the whole data set, another analysis of the data without the 4 compounds with the symbol '>', i.e. compounds **11**, **13**, **22**, and **23**, was performed. Under the same conditions as before-mentioned, for the 19 compounds, the conventional R^2 values were 0.973, 0.923 and 0.952 (*table VI*), while the cross-validated R^2 values were 0.791, 0.856 and

Table III. The fitted biological activities using neural network.

Compound	log(1/IC$_{50}$)			log(1/MIC) (Gram$^-$)			log(1/MIC) (Gram$^+$)		
	Observed value	Fitted value	Residue	Observed value	Fitted value	Residue	Observed value	Fitted value	Residue
1	4.07	4.15	− 0.08	6.94	7.45	− 0.51	7.07	7.54	− 0.47
2	3.66	3.90	− 0.24	6.57	6.23	0.34	6.37	6.00	0.37
3	3.12	3.30	− 0.18	6.30	5.98	0.32	6.30	6.04	0.26
4	4.15	4.08	0.07	7.09	6.84	0.25	6.96	7.17	− 0.21
5	3.91	3.73	0.18	6.71	6.77	− 0.06	7.08	6.94	0.14
6	3.74	3.62	0.12	5.91	5.92	− 0.01	5.61	5.94	− 0.33
7	3.12	3.08	0.04	5.82	5.86	− 0.04	5.23	5.24	− 0.01
8	4.09	3.98	0.11	6.28	6.41	− 0.13	6.80	6.62	0.18
9	3.81	3.73	0.08	6.94	6.80	0.14	6.84	6.81	0.03
10	4.39	4.20	0.19	6.58	6.10	0.48	6.84	6.13	0.71
11	3.89	4.12	− 0.23	5.30	6.14	− 0.84	5.12	6.11	− 0.99
12	3.68	3.77	− 0.09	6.94	6.83	0.11	6.59	6.51	0.08
13	3.61	3.50	0.11	5.38	5.47	− 0.09	5.09	5.08	0.01
14	3.57	3.64	− 0.07	6.66	6.54	0.12	6.00	6.04	− 0.04
15	3.34	2.91	0.43	4.22	4.76	− 0.54	4.22	4.42	− 0.20
16	4.39	4.16	0.23	7.56	7.50	0.06	7.56	7.62	− 0.06
17	3.81	3.87	− 0.06	6.68	6.60	0.08	6.88	6.87	0.01
18	3.58	3.92	− 0.34	6.53	6.68	− 0.15	6.97	6.92	0.05
19	3.15	3.18	− 0.03	5.66	5.85	− 0.19	5.72	5.86	− 0.14
20	4.52	4.35	0.17	7.26	7.24	0.02	7.66	7.38	0.28
21	3.93	3.90	0.03	7.09	7.10	− 0.01	6.86	6.97	− 0.11
22	2.90	2.98	0.08	5.28	4.81	0.47	4.50	4.50	0.00
23	2.90	3.01	− 0.11	6.49	6.50	− 0.01	6.43	6.31	0.12
R^2	0.853			0.843			0.883		
RV	0.031			0.097			0.104		

0.814 (*table VII*), for the cytotoxicity, the activities against Gram-negative and Gram-positive bacteria, respectively. Furthermore, the corresponding biological activities of the 4 removed analogs were predicted using the 19 compounds to train the network (shown in *table VIII*).

2.3. Comparison with PLS analysis

The PLS method can be used to analyze multiple QSAR by means of several dependent variables, too. The selected 4 descriptors were served as the independent variables, and the 3 biological activities as the dependent ones. Using 4 optimal components, the cross-validated R^2 for all the 23 analogs were poor, which were only 0.188, 0.258 and 0.282 for the cytotoxicity, the activities against Gram-negative and Gram-positive bacteria, respectively. After the graphical analysis of the results, compound **7** was removed from the data set, which was responsible for the poor results. The quality of the multiple QSAR model for the other 22 compounds was then improved significantly while still using 4 optimal components. For the cytotoxicity, the activities against Gram-negative and Gram-positive bacteria, the conventional R^2 values were 0.604, 0.615 and 0.650, while the cross-validated R^2 values were raised to 0.338, 0.407 and 0.467, respectively. The 3 regression equations were shown in the following:

$$\log(1/IC_{50}) = 12.975 - 0.044\ MR_8 - 1.450\ \sigma_{m5} + 52.186\ q_{N1} + 3.002 q_{O4}$$
$$n = 22,\ R^2 = 0.604,\ s = 0.286,\ F = 6.485$$

$$\log(1/MIC)(Gram^-) = 18.442 - 0.055\ MR_8 - 0.600\ \sigma_{m5} + 93.816\ q_{N1} - 6.959\ q_{O4}$$
$$n = 22,\ R^2 = 0.615,\ s = 0.493,\ F = 6.795$$

$$\log(1/MIC)(Gram^+) = 21.374 - 0.024\ MR_8 - 1.606\ \sigma_{m5} + 111.662\ q_{N1} - 5.984\ q_{O4}$$
$$n = 22,\ R^2 = 0.650,\ s = 0.554,\ F = 7.908$$

Table IV. The predicted biological activities by leave-one-out method using neural network.

Compound	log(1/IC_{50})			log(1/MIC) (Gram$^-$)			log(1/MIC) (Gram$^+$)		
	Observed value	Predicted value	Residue	Observed value	Predicted value	Residue	Observed value	Predicted value	Residue
1	4.07	4.14	− 0.07	6.94	7.05	− 0.11	7.07	7.18	− 0.11
2	3.66	3.92	− 0.26	6.57	6.19	0.38	6.37	6.12	0.25
3	3.12	3.36	− 0.24	6.30	6.02	0.28	6.30	5.81	0.49
4	4.15	3.43	0.72	7.09	6.59	0.50	6.96	6.57	0.39
5	3.91	3.89	0.02	6.71	6.76	− 0.05	7.08	6.98	0.10
6	3.74	3.52	0.22	5.91	6.16	− 0.25	5.61	6.06	0.45
7	3.12	3.75	− 0.63	5.82	6.56	− 0.74	5.23	6.67	− 1.44
8	4.09	3.77	0.32	6.28	6.68	− 0.40	6.80	6.88	− 0.08
9	3.81	3.55	0.26	6.94	6.70	0.24	6.84	6.80	0.04
10	4.39	3.88	0.51	6.58	6.54	0.04	6.84	6.43	0.41
11	3.89	3.94	− 0.05	5.30	6.63	− 1.33	5.12	6.71	− 1.59
12	3.68	3.95	− 0.27	6.94	6.79	0.15	6.59	6.80	− 0.21
13	3.61	3.03	0.56	5.38	5.45	− 0.07	5.09	5.12	− 0.03
14	3.57	3.33	0.24	6.66	6.65	0.01	6.00	6.39	− 0.39
15	3.34	2.70	0.64	4.22	5.09	− 0.87	4.22	4.23	− 0.01
16	4.39	4.43	− 0.04	7.56	7.22	0.34	7.56	7.60	− 0.04
17	3.81	4.10	− 0.29	6.68	6.78	− 0.10	6.88	7.13	− 0.25
18	3.58	3.74	− 0.16	6.53	6.49	0.04	6.97	6.45	0.52
19	3.15	2.76	0.39	5.66	5.72	− 0.06	5.72	5.70	0.02
20	4.52	4.48	0.04	7.26	7.41	− 0.15	7.66	7.36	0.30
21	3.93	3.84	0.09	7.09	6.69	0.40	6.86	6.69	0.17
22	2.90	2.73	0.17	5.28	5.28	0.00	4.50	5.15	− 0.65
23	2.90	3.19	− 0.29	6.49	5.72	0.77	6.43	5.31	1.12
R^2	0.411			0.649			0.602		
RV	0.126			0.217			0.354		

Similarly, the data set without compound **7** was also analyzed by neural networks under the same conditions as before-mentioned. For the 22 compounds, the conventional R^2 values were 0.918, 0.860 and 0.847 (*table IX*), while the cross-validated R^2 values were 0.768, 0.799 and 0.808 (*table X*), for the cytotoxicity, the activities against Gram-negative and Gram-positive bacteria, respectively.

3. Discussion

3.1. QSAR and QSTR of fluoroquinolones

The possible structural and property descriptors were obtained from two ways, quantum chemical indices and physicochemical parameters. Because the common structure of the 23 analogs was rigid and planar, and earlier QSAR studies also showed that groups at positions 1, 3, 4, 5, 6, 7, and 8 might be related to the pharmacophore, so the distance and angle indices could be ignored and only electronic indices were selected. Meanwhile, the physicochemical parameters including hydrophobic, electronic and steric parameters may serve as a supplement to the quantum chemical indices. All the 16 descriptors were possible to correlate with both activity and toxicity.

Earlier QSAR analyses of fluoroquinolones indicated that hydrophobic, electronic and steric properties of the compounds played the same important roles to the antibacterial activity [16]. The key features to their binding with macromolecular target were electronic and steric properties of substituents at 1-, 6-, 7-, 8-positions. Domagala showed that a halogen (F or Cl) at 8-position improved oral absorption and activity against anaerobes, an alkylated pyrrolidine or piperazine at 7-position increased serum half-life and potency against Gram-positive bacteria, and both of a cyclopropyl group at 1-position and an amino substituent at 5-position improved overall potency [17]. Suto et al. demonstrated that cytotoxicity was not controlled exclusively by any of substituent at the 1-, 5-, 7- or 8-position of the quinolone common structure (*figure 1*), but rather by the combination of these substituents. They also found that quinolo-

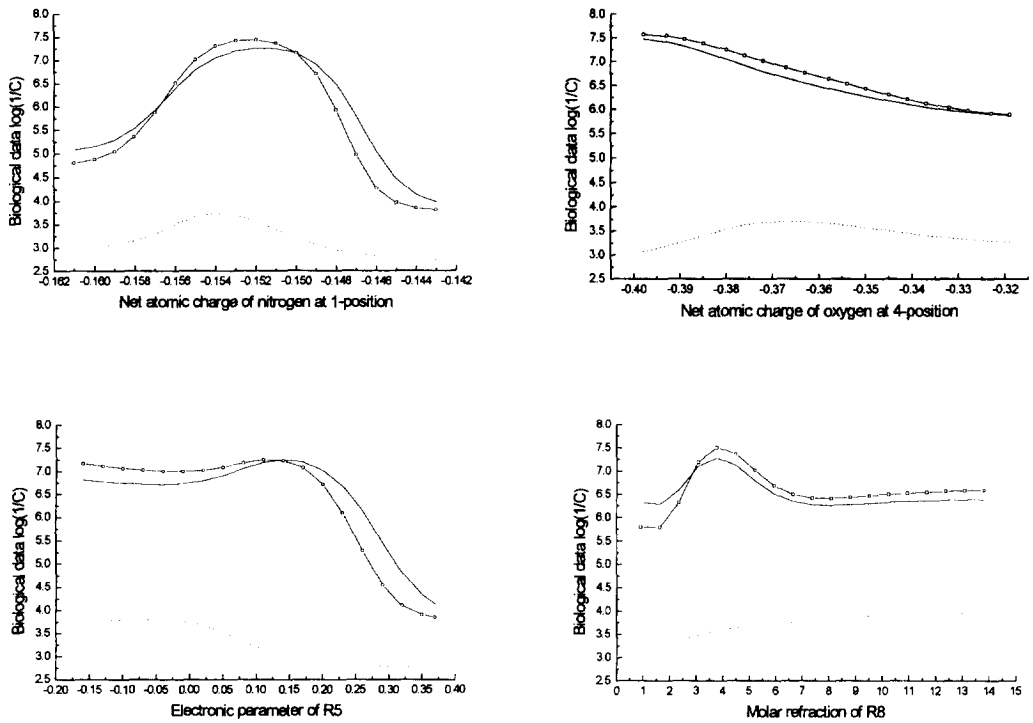


Figure 3. The functional relationships of the 3 biological activities log(1/C) with the 4 descriptors. The mammalian cell cytotoxicity is presented by a dotted line, while the activities against Gram-positive and Gram-negative bacteria are presented by solid lines with and without open circles, respectively.

nes were more cytotoxic if pyrrolidines replaced piperazines at the 7-position and/or halogens replaced hydrogen at the 8-position [11]. In the present study, we got the similar results. That is, 4 descriptors selected from 16 ones were composed of 3 electronic descriptors and 1 steric one, which were used to model the electrostatic and van der Waals interactions. The hydrophobic parameter π_5 was also shown to be important to the antibacterial activities, but it was not explicitly related with the toxicity, so was abandoned.

Tables III and *IV* showed the fitted and predicted biological activities of all the 23 compounds calculated using neural network method. The R^2 of the predicted biological activities were somewhat low. The principal reason may be that some experimental data was expressed with the symbol '>' and introduced system errors. When the 4 compounds with the symbol '>' were removed from the data set, under the same conditions, the correlation coefficient R^2 were improved singnificantly for the remaining 19 compounds (*tables VI* and *VII*). The other reason might be that either cytotoxicity or antibacterial activities of the compounds were related to other descriptors separately. For example, the hydrophobic parameter π_5 was important to the antibacterial activities, but it was not explicitly related with the toxicity. The range of the observed toxicity was narrow (only 1.62),

Table V. The new analogs derived from the multiple QSAR analysis and their predicted biological activities.

Compound	R5	R8	q_{N1}	q_{O4}	σ_{m5}	MR_8	$\log(1/IC_{50})$	log(1/MIC) (Gram$^-$)	log(1/MIC) (Gram$^+$)
New1	OH	NO	−0.153	−0.391	0.12	5.20	2.83	6.28	5.49
New2	OH	Cl	−0.151	−0.388	0.12	6.03	2.85	6.96	6.57
New3	OH	CF_3	−0.156	−0.384	0.12	5.02	2.91	7.24	7.17

Table VI. The fitted biological activities using neural network (the 4 compounds with '>' symbol removed).

Compound	log(1/IC_{50})			log(1/MIC) ($Gram^-$)			log(1/MIC) ($Gram^+$)		
	Observed value	Fitted value	Residue	Observed value	Fitted value	Residue	Observed value	Fitted value	Residue
1	4.07	4.08	− 0.01	6.94	6.95	− 0.01	7.07	7.00	0.07
2	3.66	3.70	− 0.04	6.57	6.53	0.04	6.37	6.38	− 0.01
3	3.12	3.28	− 0.16	6.30	6.16	0.14	6.30	5.93	0.37
4	4.15	4.21	− 0.06	7.09	6.79	0.30	6.96	6.98	− 0.02
5	3.91	3.89	0.02	6.71	6.96	− 0.25	7.08	7.12	− 0.04
6	3.74	3.65	0.09	5.91	5.96	− 0.05	5.61	5.96	− 0.35
7	3.12	3.04	0.08	5.82	5.67	0.15	5.23	5.42	− 0.19
8	4.09	4.09	0.00	6.28	6.42	− 0.14	6.80	6.58	0.22
9	3.81	3.84	− 0.03	6.94	6.85	0.09	6.84	6.78	0.06
10	4.39	4.22	0.17	6.58	6.99	− 0.41	6.84	7.02	− 0.18
12	3.68	3.63	0.05	6.94	6.94	0.00	6.59	6.89	− 0.30
14	3.57	3.57	0.00	6.66	6.39	0.27	6.00	6.20	− 0.20
15	3.34	3.34	0.00	4.22	4.32	− 0.10	4.22	4.14	0.08
16	4.39	4.43	− 0.04	7.56	7.36	0.20	7.56	7.50	0.06
17	3.81	3.85	− 0.04	6.68	6.64	0.04	6.88	6.73	0.15
18	3.58	3.56	0.02	6.53	6.77	− 0.24	6.97	6.85	0.12
19	3.15	3.10	0.05	5.66	5.95	− 0.29	5.72	5.55	0.17
20	4.52	4.53	− 0.01	7.26	7.40	− 0.14	7.66	7.56	0.10
21	3.93	3.97	− 0.04	7.09	6.80	0.29	6.86	6.88	− 0.02
R^2	0.973			0.923			0.952		
RV	0.005			0.042			0.034		

which should be responsible for the toxicity to have the lowest R^2 value. However, these statistical indices were enough to enable the multiple QSAR model just constructed to provide reliable predictions of the three biological activities for new derivative fluoroquinolones.

From *figure 3*, we can see that all the 4 descriptors possessed irregularly nonlinear relationships with the antibacterial activities or cytotoxicity. The two activities have very similar relationships with the descriptors. Approximate trends exist between the cytotoxicity and antibacterial activities, which agree with the qualitative analysis of Suto et al. [11]. One difference between the activities and the toxicity is that the toxicity increases with the increase of q_{O4} when q_{O4} is lower than − 0.37, whereas the antibacterial activities decrease. Therefore, when an electron-contributing group is introduced into 5-position, the net atomic charge of O_4 becomes more negative, which is favorable to the activity. However, if σ_{m5} is more negative, the toxicity may increase too. Similarly, a medium-sized electron-withdrawing group at 8-position is favorable to the activity. In order to obtain selective derivatives with high-activity and low-toxicity, the best ranges of the 4 descriptors should be: q_{N1}: − 0.148 ~ − 0.150, q_{O4}: − 0.39 ~ − 0.40, σ_{m5}: 0.10 ~ 0.20, and MR_8: 3 ~ 5.

Based on these analyses, 3 new compounds were derived from the multiple QSAR model and the 3 biological activities were predicted for each compound (*table V*). We can see from *table V*, although the antibacterial activities of the new compounds might not excess the highest activity of the data set, the toxicity could be lower than the lowest toxicity of the original ones. Then the differences between the activities and the toxicity were larger than those of the old ones, that is to say, the selectivity of the new compounds would be better. If these 3 compounds were synthesized chemically, these predictions might be verified experimentally. We also predicted the 3 biological activities of the 4 compounds with the symbol '>' using the other 19 compounds to train the neural network (*table VIII*). *Table VIII* showed the toxicity and the activities of compounds **13** and **22** were consistent with the experimental data, but the activities of compounds **11** and **23** were deviated from the observed data a little. The same cases were also found in *tables IV* and *X*. Maybe compounds **11** and **23** also should be responsible for the results.

The results above could make us better understand some experimental facts. If R_8 remains constant, while an $-NH_2$ group is introduced into the 5-position, then $\sigma_{m5} = -0.16$, $q_{O4} = -0.39 \sim -0.40$. These derivatives

Table VII. The predicted biological activities by leave-one-out method using neural network (the 4 compounds with '>' symbol removed).

Compound	log(1/IC_{50})			log(1/MIC) (Gram$^-$)			log(1/MIC) (Gram$^+$)		
	Observed value	Predicted value	Residue	Observed value	Predicted value	Residue	Observed value	Predicted value	Residue
1	4.07	4.08	− 0.01	6.94	6.99	− 0.05	7.07	7.02	0.05
2	3.66	3.90	− 0.24	6.57	6.87	− 0.30	6.37	6.78	− 0.41
3	3.12	3.29	− 0.17	6.30	5.95	0.35	6.30	5.85	0.45
4	4.15	3.95	0.20	7.09	6.63	0.46	6.96	6.53	0.43
5	3.91	3.76	0.15	6.71	6.52	0.19	7.08	6.86	0.22
6	3.74	3.71	0.03	5.91	6.30	− 0.39	5.61	6.01	− 0.40
7	3.12	3.57	− 0.45	5.82	6.49	− 0.67	5.23	6.30	− 1.07
8	4.09	4.07	0.02	6.28	6.33	− 0.05	6.80	6.65	0.15
9	3.81	3.80	0.01	6.94	6.84	0.10	6.84	6.99	− 0.15
10	4.39	4.08	0.31	6.58	6.92	− 0.34	6.84	6.97	− 0.13
12	3.68	3.72	− 0.04	6.94	6.92	0.02	6.59	6.50	0.09
14	3.57	3.54	0.03	6.66	6.51	0.15	6.00	5.91	0.09
15	3.34	2.95	0.39	4.22	4.54	− 0.32	4.22	4.50	− 0.28
16	4.39	4.46	− 0.07	7.56	7.35	0.21	7.56	7.51	0.05
17	3.81	3.87	− 0.06	6.68	6.72	− 0.04	6.88	6.77	0.11
18	3.58	3.63	− 0.05	6.53	6.53	0.00	6.97	6.71	0.26
19	3.15	3.07	0.08	5.66	5.47	0.19	5.72	5.53	0.19
20	4.52	4.36	0.16	7.26	7.17	0.09	7.66	7.28	0.38
21	3.93	3.77	0.16	7.09	6.97	0.12	6.86	6.71	0.15
R^2	0.791			0.856			0.814		
RV	0.037			0.079			0.131		

could occupy high activity. When a $-CH_3$ group lies in the 5-position, a satisfactory activity will be obtained, too. However, these compounds perhaps occupy higher toxicity. Similarly, if R_5 remains constant, a halogen (F or Cl) at the 8-position could enhance the activity but also the toxicity. In addition, we found that the size of R5 or R8 group is important to the activity, which are located between two groups, such as >C=O and –F. If their size is large enough to make the aryl plane deform, the activity will decrease.

The interaction mechanism of fluoroquinolones with their macromolecular target is complicated, and has not been completely understood yet [28]. According to our results, such an interaction mode was suggested as following: fluoroquinolone may interact with the DNA gyrase-DNA complex; the main body of the compound intercalates into the base pairs of DNA in dimer form with N1 and O4 to form electrostatic interactions with DNA, while R7 group extends towards the minor or major groove of DNA to interact with the DNA gyrase. R5 and R8 groups form electrostatic and van der Waals interactions with the base pairs of DNA, respectively, which make contributions to the binding affinity to DNA. The planarity of the compound is important when inter-

Table VIII. The predicted biological activities of the 4 compounds with '>' symbol by leave-one-out method using neural network.

Compound	log(1/IC_{50})			log(1/MIC) (Gram$^-$)			log(1/MIC) (Gram$^+$)		
	Observed value	Predicted value	Residue	Observed value	Predicted value	Residue	Observed value	Predicted value	Residue
11	3.89	4.10	− 0.21	5.30	6.85	− 1.55	< 5.12	6.90	− 1.78
13	3.61	3.40	0.21	< 5.38	5.29	0.09	< 5.09	5.03	0.06
22	< 2.90	3.00	− 0.10	5.28	5.45	− 0.17	4.50	4.72	− 0.22
23	< 2.90	2.97	− 0.07	6.49	5.39	1.10	6.43	5.18	1.25

Table IX. The fitted biological activities of neural network (compound **7** removed).

Compound	log(1/IC_{50})			log(1/MIC) ($Gram^-$)			log(1/MIC) ($Gram^+$)		
	Observed value	Fitted value	Residue	Observed value	Fitted value	Residue	Observed value	Fitted value	Residue
1	4.07	4.17	−0.10	6.94	7.47	−0.53	7.07	7.64	−0.57
2	3.66	3.94	−0.28	6.57	6.45	0.12	6.37	6.47	−0.10
3	3.12	3.02	0.10	6.30	6.07	0.23	6.30	6.09	0.21
4	4.15	4.07	0.08	7.09	6.78	0.31	6.96	6.92	0.04
5	3.91	3.84	0.07	6.71	6.72	−0.01	7.08	6.91	0.17
6	3.74	3.84	−0.10	5.91	5.87	0.04	5.61	5.85	−0.24
8	4.09	3.99	0.10	6.28	6.34	−0.06	6.80	6.46	0.34
9	3.81	3.84	−0.03	6.94	6.67	0.27	6.84	6.62	0.22
10	4.39	4.13	0.26	6.58	6.17	0.41	6.84	5.99	0.85
11	3.89	4.10	−0.21	5.30	6.11	−0.81	5.12	6.11	−0.99
12	3.68	3.70	−0.02	6.94	6.89	0.05	6.59	6.62	−0.03
13	3.61	3.67	−0.06	5.38	5.35	0.03	5.09	4.92	0.17
14	3.57	3.62	−0.05	6.66	6.63	0.03	6.00	6.04	−0.04
15	3.34	3.40	−0.06	4.22	4.52	−0.30	4.22	4.26	−0.04
16	4.39	4.35	0.04	7.56	7.39	0.17	7.56	7.53	0.03
17	3.81	3.93	−0.12	6.68	6.87	−0.19	6.88	6.98	−0.10
18	3.58	3.84	−0.26	6.53	6.79	−0.26	6.97	6.98	−0.01
19	3.15	3.13	0.02	5.66	5.88	−0.22	5.72	5.67	0.05
20	4.52	4.51	0.01	7.26	7.01	0.25	7.66	7.10	0.56
21	3.93	3.95	−0.02	7.09	6.88	0.21	6.86	7.05	−0.19
22	2.90	3.03	−0.13	5.28	4.92	0.36	4.50	4.65	−0.15
23	2.90	2.84	0.06	6.49	6.49	0.00	6.43	6.19	0.24
R^2	0.918			0.860			0.847		
RV	0.017			0.088			0.134		

calating into DNA. When the planarity of the compound is affected by the R5 or R8 group, the antibacterial activity will be reduced. The hydrophobicity of R5 group will also affect the ability of the compound to penetrate the cellular membrane.

3.2. Neural network is powerful to analyze multiple QSAR and superior to PLS

In this paper we presented an example of multiple QSAR model built by neural network method, which shows that a multi-output-node neural network is a powerful tool to solve the multiple QSAR problem. Especially, it can be done simultaneously rather than separately.

It is important to build a multiple QSAR model with a series of analogs, in which parallel properties could be analyzed and compared, such as activity and toxicity in this case. More useful information could then be obtained under the same conditions, which will direct the development of selectively potent analog with high-activity and low-toxicity. However, if the structure-activity and structure-toxicity relationships are studied separately, it is rather difficult to make comparison among results under the same conditions, while a longer time is required. In practice, it is possible to build a multiple QSAR model by means of statistical methods with multiple dependent variables, such as neural network and PLS.

The neural network method is a new information-processing system with many advantages, such as parallel-processing, error-tolerating, nonlinear and self-taught properties. It is robust to perform nonlinear mapping between some physicochemical parameters and some corresponding biological activities implicitly due to the presence of hidden layers. However, sometimes it is possible to fall into the trap of overfitting or overtraining in analysis [26]. In order to overcome the danger of overfitting or overtraining in neural network analysis, we combined the PLS method with the neural network method, that is, using PLS to select the independent variables and using neural network to analyze the relationships between the descriptors and the biological activities.

Table X. The predicted biological activities by leave-one-out method using neural network (compound **7** removed).

Compound	log(1/IC_{50})			log(1/MIC) (Gram$^-$)			log(1/MIC) (Gram$^+$)		
	Observed value	Predicted value	Residue	Observed value	Predicted value	Residue	Observed value	Predicted value	Residue
1	4.07	4.32	−0.25	6.94	7.11	−0.17	7.07	7.21	−0.14
2	3.66	3.95	−0.29	6.57	6.44	0.13	6.37	6.60	−0.23
3	3.12	3.33	−0.21	6.30	5.95	0.35	6.30	5.97	0.33
4	4.15	3.81	0.34	7.09	6.65	0.44	6.96	6.61	0.35
5	3.91	3.85	0.06	6.71	6.87	−0.16	7.08	6.83	0.25
6	3.74	3.67	0.07	5.91	6.01	−0.10	5.61	6.08	−0.47
8	4.09	4.10	−0.01	6.28	6.45	−0.17	6.80	6.52	0.28
9	3.81	3.94	−0.13	6.94	6.65	0.29	6.84	6.61	0.23
10	4.39	3.99	0.40	6.58	6.14	0.44	6.84	6.04	0.80
11	3.89	3.98	−0.09	5.30	6.25	−0.95	5.12	6.18	−1.06
12	3.68	3.77	−0.09	6.94	6.94	0.00	6.59	6.70	−0.11
13	3.61	3.41	0.20	5.38	5.46	−0.08	5.09	5.02	0.07
14	3.57	3.69	−0.12	6.66	6.47	0.19	6.00	6.04	−0.04
15	3.34	2.85	0.49	4.22	4.53	−0.31	4.22	4.03	0.19
16	4.39	4.34	0.05	7.56	7.31	0.25	7.56	7.67	−0.11
17	3.81	3.70	0.11	6.68	6.86	−0.18	6.88	7.13	−0.25
18	3.58	3.82	−0.24	6.53	6.82	−0.29	6.97	6.86	0.11
19	3.15	3.26	−0.11	5.66	5.94	−0.28	5.72	5.90	−0.18
20	4.52	4.37	0.15	7.26	7.11	0.15	7.66	7.45	0.21
21	3.93	3.73	0.20	7.09	6.97	0.12	6.86	6.90	−0.04
22	2.90	3.06	−0.16	5.28	4.96	0.32	4.50	4.87	−0.37
23	2.90	3.07	−0.17	6.49	5.76	0.73	6.43	5.36	0.83
R^2	0.768			0.799			0.808		
RV	0.048			0.127			0.168		

The variable selection is important to the performance of neural network. Some methods have been used to select the variable, such as evolutionary algorithms and pruning algorithms [29]. Here we used the traditional PLS method. The aim of variable selection was to determine the most important variables and allow theoretically a better generalization by neural network method. Because there were 3 dependent variables in this study, it would be complicated to use some evolutionary algorithms. Some descriptors had very poor cross-validated R^2 for all the 3 biological activities, so these ones were pruned from the list. Then some different combinations of the remaining descriptors were used to test and some ones were pruned again. Finally 4 descriptors were selected with high correlation coefficient R^2 values for all the 3 biological activities and low chance factors between themselves [30]. According to the rule of the number of connected weights close to the number of samples, 3 or 4 input nodes are reasonable because only a small number of samples are available [6].

Then, when we trained the network, the stopped training method was used to prevent overfitting or overtraining [26, 27]. The weights were initialized with small random values within ± 0.3, while a slow learning rate was used and parameters $\eta 1$ and $\eta 2$, $\alpha 1$ and $\alpha 2$ took different values for the input: hidden payer and hidden:output layer separately. All these were also beneficial to the convergence of the error function as well as the prevention of overfitting or overtraining.

Comparing the results from PLS analysis with those from neural network analysis, they are consistent in the relationships of biological activities with descriptors, but the results of neural network are much better than that of PLS. In PLS analysis, all the 3 biological activities are in inverse proportions to either MR_8 or σ_{m5}, while in direct proportion to q_{N1}. The toxicity is in direct proportion to q_{O4} while the two antibacterial activities are in inverse proportions to q_{O4}. However, one compound had to be removed from the data set in PLS analysis and the PLS results seems only to be a part of neural network results. Obviously, the 3 biological activities have nonlinear relationships rather than linear ones with the 4 descriptors, which might be the reason leading to the poor PLS results. The data set without compound **7** was also

analyzed by neural network. The results were improved significantly (*tables IX* and *X*), which indicated that compound **7** should be responsible for the poor results.

In short, this study demonstrated that neural network is a powerful tool and superior to PLS in solving the multiple QSAR problem.

Acknowledgements

We would like to thank Dr. Jia-Hui Lin for her help in the manuscript. This work was supported by the National Key Research Project of China, No. 85–722–16–02.

References

[1] Fujita T., Quant. Struct. Act. Relat. 16 (1997) 107–112.

[2] Livingstone D.J., Manallack D.T., Tetko I.V., J. Comput. Aided Mol. Des. 11 (1997) 135–142.

[3] Aoyama T., Suzuki Y., Ichikawa H., J. Med. Chem. 33 (1990) 2583–2590.

[4] Andrea T.A., Kalayeh H., J. Med. Chem. 34 (1991) 2824–2836.

[5] So S.S., Richards W.G., J. Med. Chem. 35 (1992) 3201–3207.

[6] Tetko I.V., Luik A.I., Poda G.I., J. Med. Chem. 36 (1993) 811–814.

[7] Manallack D.T., Ellis D.D., Livingstone D.J., J. Med. Chem. 37 (1994) 3758–3767.

[8] So S.S., Karplus M., J. Med. Chem. 39 (1996) 5246–5256.

[9] Hansch C., Acc. Chem. Res. 2 (1969) 232–239.

[10] Livingstone D.J., Ford M.G., in: Ford M.G., Greenwood R., Brooks G.T., Franke R. (Eds.), Bioactive Compound Design: Possibilities for Industrial Use, Bios, Oxford, 1996, pp. 99–107.

[11] Suto M.J., Domagala J.M., Roland G.E., Mailloux G.B., Cohen M.A., J. Med. Chem. 35 (1992) 4745–4750.

[12] Von Rosenstiel N., Adam D., Drugs 47 (1994) 872–901.

[13] Moreau N.J., Robaux H., Baron L., Tabary X., Antimicrob. Agents Chemother. 34 (1990) 1955–1960.

[14] Elsea S.H., Osheroff N., Nitiss J.L., J. Biol. Chem. 267 (1992) 13150–13153.

[15] Chu D.T.W., Fernandes P.B., Antimicrob. Agents Chemother. 33 (1989) 131–135.

[16] Bryskier A., Chantot J.F., Drugs 49 (Suppl. 2) (1995) 16–28.

[17] Domagala J.M., J. Antimicrob. Chemother. 33 (1994) 685–706.

[18] Klopman G., Fercu D., Li J.Y., Rosenkranz H.S., Jacobs M.R., Res. Microbiol. (Paris) 147 (1996) 86–96.

[19] SYBYL 6.1, Tripos Associates Inc., St. Louis, 1995.

[20] Stewart J.J.P., J. Comput. Aided Mol. Des. 4 (1990) 1–105.

[21] Skagerberg B., Bonelli D., Clementi S. et al., Quant. Struct. Act. Relat. 8 (1989) 32–38.

[22] Frank I.E., Kalivas J.H., Kowalski B.R., Anal. Chem. 55 (1983) 1800–1804.

[23] Cramer III R.D., Bunce J.D., Patterson D.E. et al., Quant. Struct. Act. Relat. 7 (1988) 18–25.

[24] Gasteiger J., Zupan J., Angew. Chem. Int. Ed. Engl. 32 (1993) 503–527.

[25] Manallack D.T., Livingstone D.J., in: Van de Waterbeemd H. (Ed.), Chemometrics in Molecular Design, VCH, Weinheim, 1995, pp. 293–318.

[26] Tetko I.V., Livingstone D.J., Luik A.I., J. Chem. Inf. Comput. Sci. 35 (1995) 826–833.

[27] Sarle W.S., in: Proceedings of the 27th Symposium on the Interface of Computing Science and Statistics, 1995, pp. 352–360.

[28] Hooper D.C., Drugs 49 (Suppl. 2) (1995) 10–15.

[29] Tetko I.V., Villa A.E.P., Livingstone D.J., J. Chem. Inf. Comput. Sci. 36 (1996) 794–803.

[30] Topliss J.P., Edwards R.P., J. Med. Chem. 22 (1979) 1238–1244.