ICC
Iranian Chemical Communication
Original Research Article
Payame Noor University http://icc.journals.pnu.ac.ir

QSRR models of veterinary drugs in milk in ultra-performance liquid chromatography coupled to time of flight mass spectrometry

Sharmin Esmaeilpoor[a], Shahnaz Nosratolahy[b], Hadi Noorizadeh[a,*], Zohreh Moghadam[a]

[a]*Department of Chemistry, Payame Noor University, P.O. BOX 19395-4697, Tehran, Iran*

[b]*Education of Ilam, Fazilat Girls High School*

Received: 5 April 2014, Accepted: 17 June 2014, Published: 17 June 2014

Abstract

The veterinary drugs residues are also important pollutants found in milk since veterinary drugs are commonly used in cattle management. Considering the role of milk in human nutrition and its wide consumption throughout the world, it is very important to ensure the milk quality. A quantitative structure–retention relationship (QSRR) was developed using the partial least square (PLS), Kernel PLS (KPLS) and Levenberg-Marquardt artificial neural network (L-M ANN) approach for chemometrics study. Genetic algorithm was employed as a factor selection procedure for PLS and KPLS modeling methods. By comparing the results, GA-KPLS descriptors are selected for L-M ANN. Finally, a model with a low prediction error and a good correlation coefficient was obtained by L-M ANN. This is the first research on the QSRR of veterinary drugs using the chemometrics models.

Keywords: Veterinary drugs; milk; UPLC; TOFMS; correlation coefficient; QSRR.

Introduction

Many classes of antibiotics, such asmacrolides, sulphonamides, quinolones, anthelmintics and tetracyclines, are widely administered to food-producing animals for the purposes of prevention and treatment of several diseases as well as for promoting growth [1]. Residues of drugs in food can endanger con-

*Corresponding author: Hadi Noorizadeh
Tel: 09181432750, Fax: +98 (841) 3382681
E-mail: Hadinoorizadeh@yahoo.com

sumers’ health. Short-term health effects include allergic and toxic reactions, and long-term exposure could result in chronic toxic effects or the development of antibiotic-resistant bacteria in humans. Antimicrobial agents in food, regardless of their minute amounts, can be potentially carcinogenic. Milk consumption has been promoted around the world as it is an inexpensive source of saturated fats, proteins and calcium. It provides the primary source of nutrition for young mammals before they are able to digest other types of food. Milk is known to be a nutritious, wholesome food that is consumed globally by humans. It is recommended for children and elderly women. The presence of antibiotics in milk can be very problematic because their residues can slow or destroy the growth of the fermentation bacteria as well as they can provoke allergic reactions in some hypersensitive individuals. Different studies indicate that low-level doses of antibiotics for long periods could result in bacteria resistance [2].

To protect milk consumers’ health from the presence of residues of veterinary drugs, maximum residue levels (MRLs) of veterinary drugs in food have been set up in the Commission Regulation 37/2010 [3]. The Commission includes MRLs for substances that are normally employed to treat or prevent animal diseases. However, as a consequence of the unavoidable carry-over of coccidiostats in non-target feed, maximum levels (MLs) for residues of these substances in food of animal origin have been set up in the Commission Regulation 124/2009 [4].

The accurate detection of low levels of antimicrobial drug residues in milk is not only of great importance for governmental control laboratories and the dairy industry, but also for farmers to enable them to ensure that contaminated milk from individual cows is not consigned to the bulk tank. Milk may also be cocontaminated with compounds of one of the other four major antimicrobial drug classes: the sulphonamides (e.g., sulphadiazine), tetracyclines (e.g., oxytetracycline), macrolides (e.g., erythromycin) and aminoglycosides (e.g., neomycin) [5].

In order to ensure human food safety, the European Union has set (MRLs) in milk for some antibiotics, such as 10 ~ g/kg for fenbendazol and 100 ~ g/kg for enrofloxacin, although for some antibiotics it is indicated that they cannot be used in animals from which milk is produced for human consumption. These limits require the development of sensitive and specific methods to monitor and determine antibiotic residues in milk [6-8].

Several papers use bioassay techniques for rapid determination of antibiotic residues in food due to cost and simplicity of analyses [9,10]. However, these methods do not distinguish among several classes of antibiotics, and they only provide semiquantitative measurements, so other techniques such as liquid chromatography (LC) [1,11] or capillary electrophoresis [12] are being used to detect veterinary drugs in milk. LC has been the most frequently instrumental technique coupled with UV, photodiode array detection [13,14] and fluorimetric detection [15]. However these conventional detection techniques have been replaced by mass spectrometry (MS) detection, bearing in mind that public health agencies rely on detection by MS for unambiguous confirmation of antibitiocs in foodstuff [16].

Recently, a novel approach represented by LC–TOF MS has been introduced into the analysis of pesticide residues in food. The potential of this technique for both, target and non-target analyses, has been demonstrated in several studies [17–19]. Ongoing developments in instrument design have resulted not only in extending dynamic range allowing improved quantification, but also in high attainable accuracy of mass measurements (typically 2–5 ppm). This, in combination with high spectral resolution (5000–12,000 FWHM, full width at half of maximum), enables the identification of unknowns based on elemental composition.

Prediction of physico–chemical properties of materials based on their molecular structure has been one of the wishes of scientists and engineers for a long time. One of the best methods which have been applied for this purpose is quantitative structure–property relationships (QSPR) [20,21]. Quantitative structure–retention relationships (QSRR) represent statistical models that quantify the relation between the structure of molecules and their chromatographic retention time, allowing the prediction of the RT of novel compounds [22,23]. QSRR on the RT has been reported for different types of compounds [24-26]. The aim of the present study is estimation of ability optimal descriptors calculated with linear regression (the partial least squares (PLS)) and non-linear regressions (the kernel partial least squares (KPLS) and Levenberg- Marquardt artificial neural network (L-M ANN)) in QSRR analysis of retention time (RT) of veterinary drugs in milk. The stability and predictive power of these models was validated using Leave-Group-Out Cross-Validation (LGO CV) and external test set techniques.

Computational

Computer hardware and software

A Pentium IV personal computer (CPU at 3.06 GHz) with the Windows XP operating system was used. The structures of the pesticides were drawn with HyperChem version 7.0. All molecules were preoptimized using molecular mechanics AM1 method in the HyperChem program. Some quantum descriptor such as polarizability and orbital energy of LUMO were calculated by using the HyperChem software. The output files were exported from Dragon for generating descriptors developed by Todeschini *et al* [27]. The GA-PLS, GA-KPLS, L-M ANN, cross validation and other calculations were performed in MATLAB (Version 7.0, Math works, Inc).

Data set

A data set of veterinary drugs residues in raw milk with their RT values were available from the literature reported in this reference [28]. These data were obtained by ultra-performance liquid chromatography coupled to the time of flight mass spectrometry (UPLC–TOFMS). These compounds are included avermectines, benzimidazoles, beta-agonists, beta-lactams, corticoides, macrolides, nitroimidazoles, quinolones, sulfonamides,m tetracyclines and some others.

An Acquity UPLC system coupled to LCT Premier XE (Waters Corp., MA, USA) was employed for all experiments. The chromatography was carried out on a Waters Acquity UPLC BEH C18, 1.7 ~ m 100×2.1 mm column protected with a precolumn VanGuard Acquity UPLC BEH C18, 1.7 ~ m 5×2.1 mm. The mobile phase consisted of 0.1% of formic acid in water and 0.1% of formic acid in MeCN. Mass spectrometry was performed using a LCT Premier XE (Waters, Manchester, UK) equipped with a dual ESI source (lock spray). The system was tuned for optimum sensitivity and resolution using leucine-enkephalin solution at 0.5 ng/ ~ L infused at 5 ~ L/min in positive electrospray ionization mode. The TOF was calibrated daily using sodium formate solution. The system was operated in V mode with acquisition from 50 to 1150 m/z with a scan time of 0.2 s in order to reach the best sensitivity. The name of studied compounds and their experimental RT values for training and test sets are shown in Table 1 and Table 2, respectively.

The root mean square error of prediction (*RMSE*) is a measurement of the average difference between predicted and experimental values at the prediction stage. The *RMSE* can be interpreted as the average prediction error, expressed in the same units as the original response values. The *RMSEP* was obtained using the following formula:

$$RMSE = [\frac{1}{n}\sum_{i=1}^{n}(y_i - y_i^{\wedge})^2]^{\frac{1}{2}} \qquad (1)$$

Where y_i is the experimental RT value of the pesticides in the sample i, $\hat{y}_i$ represents the predicted RT value in the sample i and n is the total number of samples used in the test set.

Table 1. The compounds, structure, retention time (min), calculated and RMSE values by L-M ANN model for training set

No	Compounds	Structure	RT $_{Exp}$	RT $_{Cal}$	RMSE
	Calibration Set				
1	Sulfaguanidine	$C_7H_{10}N_4O_2S$	0.95	0.88	0.007
2	Salbutamol	$C_{13}H_{21}NO_3$	1.48	1.34	0.015
3	Amoxicillin	$C_{16}H_{19}N_3O_5S$	1.50	1.62	0.013
4	Olaquindox	$C_{12}H_{13}N_3O_4$	1.55	1.46	0.010
5	Roxarsone	$C_6H_6AsNO_6$	1.57	1.51	0.007
6	Metronidazole	$C_6H_9N_3O_3$	1.67	1.72	0.006
7	Dimetridazole-hydroxy	$C_5H_7N_3O_3$	1.68	1.61	0.008
8	Thiabendazole-5-hydroxy	$C_{10}H_7N_3OS$	1.70	1.69	0.001
9	Sulfisomidine	$C_{12}H_{14}N_4O_2S$	1.75	1.60	0.016
10	Sulfacetamide	$C_8H_{10}N_2O_3S$	1.77	1.94	0.018
11	Cimbuterol	$C_{13}H_{19}N_3O$	1.79	1.82	0.003
12	Clopidol	$C_7H_7Cl_2NO$	1.82	1.79	0.003
13	Lincomycin	$C_{18}H_{34}N_2O_6S$	1.88	1.89	0.001
14	Cefaclor	$C_{15}H_{14}ClN_3O_4S$	1.89	1.83	0.007
15	Cefquinom	$C_{23}H_{24}N_6O_5S_2$	1.90	1.87	0.003
16	Levamisole	$C_{11}H_{12}N_2S$	1.92	1.81	0.011
17	Ternidazole	$C_7H_{11}N_3O_3$	1.93	1.96	0.003
18	Sulfathiazole	$C_9H_9N_3O_2S_2$	2.00	2.10	0.011
19	Enoxacin	$C_{15}H_{17}FN_4O_3$	2.02	2.03	0.001

20	Trimethoprim	$C_{14}H_{18}N_4O_3$	2.05	2.11	0.006
21	Norfloxacin	$C_{16}H_{18}FN_3O_3$	2.07	1.95	0.013
22	Cephacetrile	$C_{13}H_{13}N_3O_6S$	2.08	2.03	0.005
23	Fleroxacin	$C_{17}H_{18}F_3N_3O_3$	2.10	2.22	0.013
24	Carbadox	$C_{11}H_{10}N_4O_4$	2.12	2.20	0.008
25	Oxytetracyclin	$C_{22}H_{24}N_2O_9$	2.14	2.12	0.002
26	Cephradin	$C_{16}H_{19}N_3O_4S$	2.15	2.18	0.003
27	Sulfamerazine	$C_{11}H_{12}N_4O_2S$	2.17	2.30	0.014
28	Cefotaxime	$C_{16}H_{17}N_5O_7S_2$	2.18	1.99	0.020
29	Enrofloxacin	$C_{19}H_{22}FN_3O_3$	2.24	2.29	0.005
30	Albendazole sulfoxide	$C_{12}H_{15}N_3O_3S$	2.32	2.11	0.022
31	Cefazolin	$C_{14}H_{14}N_8O_4S_3$	2.34	2.30	0.004
32	Tetracyclin	$C_{22}H_{24}N_2O_8$	2.35	2.26	0.010
33	Sulfamethazine	$C_{12}H_{14}N_4O_2S$	2.37	2.55	0.019
34	Sulfamethoxypyridazine	$C_{11}H_{12}N_4O_3S$	2.41	2.18	0.025
35	Ipronidazole-hydroxy	$C_7H_{11}N_3O_3$	2.44	2.57	0.014
36	Demeclocyclin	$C_{21}H_{21}ClN_2O_8$	2.45	2.68	0.024
37	Clenbuterol	$C_{12}H_{18}Cl_2N_2O$	2.46	2.70	0.025
38	Difloxacin	$C_{21}H_{19}F_2N_3O_3$	2.47	2.61	0.015
39	Spiramycin I	$C_{43}H_{74}N_2O_{14}$	2.51	2.48	0.003
40	Morantel	$C_{12}H_{16}N_2S$	2.52	2.52	0.000
41	Mebendazole-amine	$C_{14}H_{11}N_3O$	2.57	2.42	0.016
42	Dapsone	$C_{12}H_{12}N_2O_2S$	2.60	2.67	0.008
43	Cefoperazone	$C_{25}H_{27}N_9O_8S_2$	2.62	2.53	0.010
44	Brombuterol	$C_{12}H_{18}Br_2N_2O$	2.64	2.88	0.026
45	Chlortetracycline	$C_{22}H_{23}ClN_2O_8$	2.67	2.79	0.013
46	Flubendazole-amine	$C_{14}H_{10}FN_3O$	2.69	2.64	0.005
47	Cyclobendazole	$C_{13}H_{13}N_3O_3$	2.72	2.85	0.014
48	Carazolol	$C_{18}H_{22}N_2O_2$	2.76	2.63	0.014

49	Pyrimethamine	$C_{12}H_{13}ClN_4$	2.77	2.74	0.004
50	Cinoxacin	$C_{12}H_{10}N_2O_5$	2.79	3.03	0.026
51	Oxibendazole	$C_{12}H_{15}N_3O_3$	2.82	2.92	0.011
52	Sulfaethidol	$C_{10}H_{12}N_4O_2S_2$	2.83	2.67	0.017
53	Cefamandole	$C_{18}H_{18}N_6O_5S_2$	2.84	3.03	0.020
54	Tilmicosin	$C_{46}H_{80}N_2O_{13}$	2.86	2.93	0.008
55	Sulfatroxazole	$C_{11}H_{13}N_3O_3S$	2.88	2.67	0.022
56	Sulfisoxazole	$C_{11}H_{13}N_3O_3S$	2.94	2.77	0.018
57	Ceftiofur	$C_{19}H_{17}N_5O_7S_3$	2.98	3.04	0.007
58	Oxolinic acid	$C_{13}H_{11}NO_5$	2.99	3.07	0.008
59	Sulfabenzamide	$C_{13}H_{12}N_2O_3S$	3.13	3.05	0.009
60	Cephalothin	$C_{16}H_{16}N_2O_6S_2$	3.17	3.22	0.005
61	Sulfadimethoxine	$C_{12}H_{14}N_4O_4S$	3.18	3.15	0.003
62	Erythromycin	$C_{37}H_{67}NO_{13}$	3.19	2.90	0.031
63	Hydrocortisone	$C_{21}H_{30}O_5$	3.21	2.91	0.032
64	Natamycin	$C_{33}H_{47}NO_{13}$	3.22	3.10	0.013
65	Fenbendazole sulfone	$C_{15}H_{13}N_3O_4S$	3.29	2.97	0.034
66	Mebendazole	$C_{16}H_{13}N_3O_3$	3.33	3.25	0.008
67	Benzocaine	$C_9H_{11}NO_2$	3.34	3.46	0.013
68	Penicillin G	$C_{16}H_{18}N_2O_4S$	3.41	3.13	0.029
69	6a-Methylprednisolone	$C_{22}H_{30}O_5$	3.48	3.16	0.034
70	Tiamulin	$C_{28}H_{47}NO_4S$	3.49	3.57	0.008
71	Betamethasone	$C_{22}H_{29}FO_5$	3.55	3.63	0.008
72	Flubendazole	$C_{16}H_{12}FN_3O_3$	3.58	3.23	0.037
73	Sulfanitran	$C_{14}H_{13}N_3O_5S$	3.63	3.28	0.037
74	Troleandomycin	$C_{41}H_{67}NO_{15}$	3.64	3.41	0.024
75	Roxithromycin	$C_{41}H_{76}N_2O_{15}$	3.66	3.72	0.007
76	Leucomycin A1	$C_{40}H_{67}NO_{14}$	3.67	3.35	0.034
77	Oxacillin	$C_{19}H_{19}N_3O_5S$	3.79	3.69	0.011

78	Josamycin	$C_{42}H_{69}NO_{15}$	3.80	3.90	0.011
79	Triflupromazine	$C_{18}H_{19}F_3N_2S$	3.88	4.05	0.018
80	Zeranol	$C_{18}H_{26}O_5$	4.01	4.16	0.016
81	Cloxacillin	$C_{19}H_{18}ClN_3O_5S$	4.03	3.69	0.035
82	Ketoprofen	$C_{16}H_{14}O_3$	4.11	4.07	0.004
83	Praziquantel	$C_{19}H_{24}N_2O_2$	4.17	4.12	0.005
84	Meloxicam	$C_{14}H_{13}N_3O_4S_2$	4.22	4.10	0.013
85	Dicloxacillin	$C_{19}H_{17}Cl_2N_3O_5S$	4.31	4.01	0.032
86	Febantel	$C_{20}H_{22}N_4O_6S$	4.90	5.07	0.018
87	Novobiocin	$C_{31}H_{36}N_2O_{11}$	5.06	5.26	0.021
88	Tolfenamic acid	$C_{14}H_{12}ClNO_2$	5.35	5.59	0.026
89	Abamectin	$C_{48}H_{72}O_{14}$	6.53	6.50	0.003
90	Ivermectin B1a	$C_{48}H_{74}O_{14}$	7.36	7.44	0.009
	Prediction Set				
91	Cefsulodin	$C_{22}H_{20}N_4O_8S_2$	1.45	1.36	0.017
92	Diethylcarbamazine	$C_{10}H_{21}N_3O$	1.60	1.66	0.010
93	Cephapirin	$C_{17}H_{17}N_3O_6S_2$	1.72	1.92	0.036
94	Ronidazole	$C_6H_8N_4O_4$	1.88	2.20	0.058
95	Thiabendazole	$C_{10}H_7N_3S$	1.90	1.87	0.006
96	Ampicillin	$C_{16}H_{19}N_3O_4S$	2.00	1.95	0.009
97	Marbofloxacin	$C_{17}H_{19}FN_4O_4$	2.02	2.34	0.059
98	Pefloxacin	$C_{17}H_{20}FN_3O_3$	2.10	2.04	0.012
99	Pyrantel	$C_{11}H_{14}N_2S$	2.16	2.47	0.056
100	Danofloxacin	$C_{19}H_{20}FN_3O_3$	2.19	2.57	0.069
101	Clenproperol	$C_{11}H_{16}Cl_2N_2O$	2.26	2.19	0.013
102	Sulfameter	$C_{11}H_{12}N_4O_3S$	2.41	2.34	0.012
103	Tulobuterol	$C_{12}H_{18}ClNO$	2.45	2.24	0.038
104	Cefuroxime	$C_{16}H_{16}N_4O_8S$	2.51	2.32	0.035
105	Sulfamonomethoxine	$C_{11}H_{12}N_4O_3S$	2.59	2.83	0.044

106	Cefoxitin	$C_{16}H_{17}N_3O_7S_2$	2.67	2.56	0.021
107	Ticlopidine	$C_{14}H_{14}ClNS$	2.71	3.13	0.076
108	Oxfendazole	$C_{15}H_{13}N_3O_3S$	2.78	2.65	0.024
109	Sulfamethoxazole	$C_{10}H_{11}N_3O_3S$	2.83	2.69	0.025
110	Carbenicillin	$C_{17}H_{18}N_2O_6S$	2.96	3.55	0.107
111	Sulfachlorpyrazine	$C_{10}H_9ClN_4O_2S$	3.15	3.25	0.018
112	Sulfaquinoxaline	$C_{14}H_{12}N_4O_2S$	3.19	3.20	0.001
113	Promethazine	$C_{17}H_{20}N_2S$	3.32	3.30	0.004
114	Albendazole	$C_{12}H_{15}N_3O_2S$	3.36	3.39	0.006
115	Nalidixic acid	$C_{12}H_{12}N_2O_3$	3.54	3.60	0.012
116	Flumequine	$C_{14}H_{12}FNO_3$	3.64	3.92	0.052
117	Virginiamycin M1	$C_{28}H_{35}N_3O_7$	3.82	3.58	0.045
118	Naproxen	$C_{14}H_{14}O_3$	4.13	3.76	0.067
119	Rifaximin	$C_{43}H_{51}N_3O_{11}$	4.53	4.70	0.031
120	Eprinomectin B1a	$C_{50}H_{75}NO_{14}$	6.18	5.76	0.077

Table 2. The compounds, structure, RT, calculated and RMSE values by L-M ANN model for test set

No	Compounds	Structure	RT $_{Exp}$	RT $_{Cal}$	RMSE
		Test Set			
1	Metronidazole-hydroxy	$C_6H_9N_3O_4$	1.48	1.80	0.058
2	Cefadroxil	$C_{16}H_{17}N_3O_5S$	1.62	1.99	0.068
3	Procaine	$C_{13}H_{20}N_2O_2$	1.70	1.73	0.005
4	Dimetridazole	$C_5H_7N_3O_2$	1.88	1.98	0.018
5	Sulfadiazine	$C_{10}H_{10}N_4O_2S$	1.89	2.01	0.022
6	Diaveridine	$C_{13}H_{16}N_4O_2$	1.93	1.84	0.017
7	Minocyclin	$C_{23}H_{27}N_3O_7$	2.00	1.90	0.019
8	Cefalexin	$C_{16}H_{17}N_3O_4S$	2.05	2.23	0.032
9	Sulfapyridine	$C_{11}H_{11}N_3O_2S$	2.08	2.40	0.059
10	Ciprofloxacin	$C_{17}H_{18}FN_3O_3$	2.17	2.26	0.016

11	Sulfamoxole	$C_{11}H_{13}N_3O_3S$	2.28	2.34	0.011
12	Sulfamethizole	$C_9H_{10}N_4O_2S_2$	2.37	2.37	0.000
13	Sarafloxacin	$C_{20}H_{17}F_2N_3O_3$	2.44	2.94	0.091
14	Xylazine	$C_{12}H_{16}N_2S$	2.45	2.74	0.053
15	Mebendazole-5-hydroxy	$C_{16}H_{15}N_3O_3$	2.62	2.58	0.007
16	Sulfachlorpyridazine	$C_{10}H_9ClN_4O_2S$	2.69	3.34	0.119
17	Doxycyclin	$C_{22}H_{24}N_2O_8$	2.78	3.13	0.063
18	Sulfadoxine	$C_{12}H_{14}N_4O_4S$	2.82	2.76	0.010
19	Albendazole sulfone	$C_{12}H_{15}N_3O_4S$	2.87	3.27	0.073
20	Ipronidazole	$C_7H_{11}N_3O_2$	2.98	3.38	0.072
21	Prednisolone	$C_{21}H_{28}O_5$	3.17	3.00	0.031
22	Tylosin	$C_{46}H_{77}NO_{17}$	3.30	3.00	0.056
23	Bromhexine	$C_{14}H_{20}Br_2N_2$	3.41	3.38	0.005
24	Dexamethasone	$C_{22}H_{29}FO_5$	3.58	3.37	0.039
25	Penicillin V	$C_{16}H_{18}N_2O_5S$	3.67	3.39	0.051
26	Fenbendazole	$C_{15}H_{13}N_3O_2S$	3.86	4.63	0.140
27	Nafcillin	$C_{21}H_{22}N_2O_5S$	4.13	3.98	0.028
28	Flunixin	$C_{14}H_{11}F_3N_2O_2$	4.31	3.93	0.070
29	Emamectin B1a	$C_{49}H_{75}NO_{13}$	5.11	5.91	0.146
30	Doramectin	$C_{50}H_{74}O_{14}$	6.84	7.44	0.110

Determination of molecular descriptors

Molecular descriptors are defined as numerical characteristics associated with chemical structures. The molecular descriptor is the final result of a logic and mathematical procedure which transforms chemical information encoded within a symbolic representation of a molecule into a useful number applied to correlate physical properties. The Dragon software was used to calculate the descriptors in this research and a total of molecular descriptors, from 18 different types of theoretical descriptors, were calculated for each molecule. Since the values of many descriptors are related to the bonds length, bonds angles and etc., the chemical structure of every molecule must be optimized before calculating its molecular descriptors. For this reason, the chemi-

cal structure of the 150 studied molecules were drawn with Hyperchem software and saved with the HIN extension. To optimize the geometry of these molecules, the AM1 geometrical optimization was applied. After optimizing the chemical structures of all compounds, the molecular descriptors were calculated using Dragon. A wide variety of descriptors have been reported in the literature, having been used in QSRR analysis.

Nonlinear model

Artificial neural network

A three-layer back propagation artificial neural network ANN with a sigmoid transfer function was used in the investigation of feature sets. The descriptors from the calibration set were used for the model generation whereas the descriptors from the prediction set were used to stop the overtraining of network. The descriptors from the test set were used to verify the predictivity of the model. Before training the networks, the input and output values were normalized with autoscaling of all data [29-32]. The goal of training the network is to minimize the output errors by changing the weights between the layers.

$$\Delta W_{ij,n} = F_n + r\Delta W_{ij,n-1} \qquad (2)$$

In this, ΔW_{ij} is the change in the weight factor for each network node, is the momentum factor, and F is a weight update function, which indicates how weights are changed during the learning process. The weights of hidden layer were optimized using the Levenberg-Marquardt algorithm, a second derivative optimization method [33].

Levenberg-Marquardt Algorithm

In Levenberg-Marquardt algorithm, the update function, F_n, is calculated using equations.

$$F_0 = -g_0 \qquad (3)$$

$$g = J^T e \qquad (4)$$

$$F_n = -[J^T \times J + \sim I]^{-1} \times J^T \times e \qquad (5)$$

Where g is gradient and J is the Jacobian matrix that contains first derivatives of the network errors with respect to the weights, and e is a vector of network errors. The parameter μ is multiplied by some factor ( ) whenever a step would result in an increased e and when a step reduces e, μ is divided by [34].

Results and discussion

Linear model

Results of the GA-PLS model

The best model is selected on the basis of the highest square correlation coefficient leave-group-out cross validation (R^2), the least root

mean squares error (*RMSE*) and relative error (*RE*). These parameters are probably the most popular measure of how well a model fits the data. The best GA-PLS model contains twelve selected descriptors in four latent variables space. The R^2, mean *RE* and *RMSE* for training and test sets were (0.826, 0.731), (12.26, 20.48) and (0.034, 0.132), respectively. The predicted values of RT are plotted against the experimental values for training and test sets in Fig 1. In general, the number of components (latent variables) is less than the number of independent variables in PLS analysis. The PLS model uses higher number of descriptors that allow the model to extract better structural information from descriptors to result in a lower prediction error.

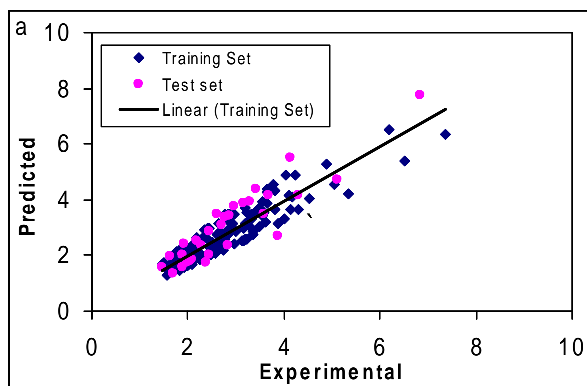

Figure 1. Plots of predicted retention time against the experimental values

Nonlinear model

Results of the GA-KPLS model

In this paper, a radial basis kernel function, k(x,y)= exp(||x-y||^2/c), was selected as the kernel function with $c = rm\dagger^2$ where r is a constant that can be determined by considering the process to be predicted (here r was set to be 1), m is the dimension of the input space and $\dagger^2$ is the variance of the data [35, 36]. It means that the value of c depends on the system under the study. The 10 descriptors in 7 latent variables space chosen by GA-KPLS feature selection methods were contained. The R^2, mean *RE* and *RMSE* for training and test sets were (0.833, 0.764), (11.92, 17.09) and (0.031, 0.105), respectively. The Q^2 for training and test sets GA-PLS and GA-KPLS models are (0.839, 0.717) and (0.847, 0.759), respectively. It can be seen from these results that statistical results for GA-KPLS model are superior to GA-PLS method. The plots of the residuals versus the experimental RT values obtained by the GA-KPLS modeling, is demonstrated in Figure 2.

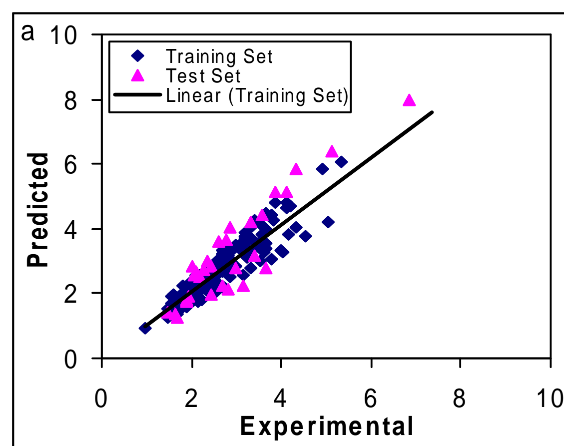

Figure 2. Plots of predicted RT versus the experimental values

Results of the L-M ANN model

With the aim of improving the predictive performance of nonlinear QSRR model, L-M ANN modeling was performed. The networks were generated using the ten descriptors appearing in the GA-KPLS models as their inputs and RT as their output. For ANN generation, data set was separated into three groups: calibration, prediction (training) and test sets. All molecules were randomly placed in these sets. A three-layer network with a sigmoid transfer function was designed for each ANN. Before training the networks, the input and output values were normalized between -1 and 1. The network was then trained using the training set by the back propagation strategy for optimization of the weights and bias values. The proper number of nodes in the hidden layer was determined by training the network with different number of nodes in the hidden layer. The root-mean-square error (RMSE) value measures how good the outputs are in comparison with the target values. It should be noted that for evaluating the overfitting, the training of the network for the prediction of RT must stop when the RMSE of the prediction set begins to increase while RMSE of calibration set continues to decrease. Therefore, training of the network was stopped when overtraining began. All of the above mentioned steps were carried out using basic back propagation, conjugate gradient and Levenberge-Marquardt weight update functions. It was realized that the RMSE for the training and test sets are minimum when three neurons were selected in the hidden layer. Finally, the number of iterations was optimized with the optimum values for the variables. It was realized that after 18 iterations, the RMSE for prediction set were minimum. The *RMSE*, mean relative error and R^2 for calibration, prediction and test sets were (0.013, 4.63, 0.978), (0.031, 7.11, 0.933) and (0.049, 9.45, 0.891), respectively. Comparison between these values and other statistical parameter reveals the superiority of the L-M ANN model over other model. In Tables 1 and 2, the predicted and RMSE values of RT obtained by the L-M ANN model are presented. The key strength of neural networks, unlike regression analysis, is their ability to carry out flexible mapping of the selected features by manipulating their functional dependence implicitly. The statistical parameters reveal the high predictive ability of L-M ANN model. The whole of these data clearly displays a significant improvement of the QSRR model consequent to nonlinear statistical treatment. The Q^2 for calibration, prediction and test sets was (0.977, 0.940, 0.905), respectively.

Plot of predicted RT versus experimental RT values by L-M ANN for training and test sets are shown in Figure 3a and Figure 3b. Obviously, there is a close agreement between the experimental and predicted RT and the data represent a very low scattering around a straight line with respective slope and intercept close to one and zero. As can be seen in this section, the L-M ANN is more reproducible than other models for modeling the RT of veterinary drugs in milk.

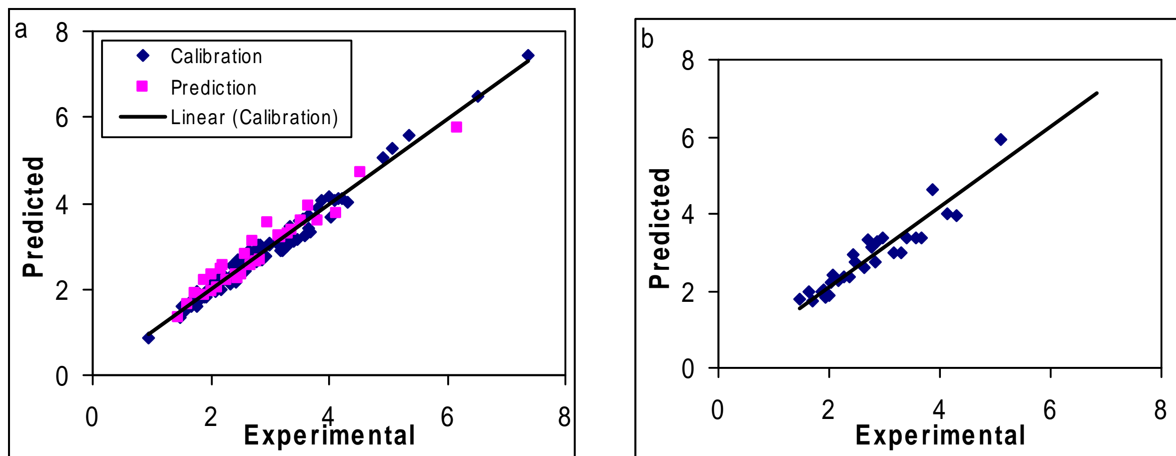

Figure 3. Plot of predicted RT obtained by L-M ANN against the experimental values (a) for training set and (b) test set

Model validation and statistical parameters

The accuracy of proposed models was illustrated using the evaluation techniques such as leave group out cross-validation (LGO-CV) procedure, validation through an external test set. In addition, chance correlation procedure is a useful method for investigating the accuracy of the resulted model, by which one can make sure if the results were obtained by chance or not. Cross validation is a popular technique used to explore the reliability of statistical models. Based on this technique, a number of modified data sets are created by deleting one or a small group (leave-some-out) of objects in each case. For each data set, an input–output model is developed, based on the utilized modeling technique. Each model is evaluated by measuring its accuracy in predicting the responses of the remaining data (the ones or group data that have not been utilized in the development of the model). In particular, the LGO-CV procedure was utilized in this study. A QSRR model was then constructed on the basis of this reduced data set and subsequently used to predict the removed data. This procedure was repeated until a complete

set of predicted was obtained. The statistical significance of the screened model was judged by the correlation coefficient (Q^2). The predictive ability was evaluated by the cross validation coefficient (Q^2 or R^2_{cv}). The accuracy of cross validation results is extensively accepted in the literature considering the Q^2 value. In this sense, a high value of the statistical characteristic ($Q^2 > 0.5$) is considered as proof of the high predictive ability of the model.

The data set should be divided into three new sub-data sets, one for calibration and prediction (training) and the other one for testing. The calibration set was used for model generation. The prediction set was applied dealing with overfitting of the network whereas test set which its molecules have no role in model building was used for the evaluation of the predictive ability of the models for external set [37].

In the other hand, by means of training set, the best model is found and then, its prediction power is checked by test set as an external data set. In this work, 60% of the database was used for calibration set, 20% for prediction set and 20% for test set [38, 39], randomly (in each running program, from all 150 components, 90 components are in calibration set, 30 components are in prediction set and 30 components are in test set).

The result clearly displays a significant improvement of the QSRR model consequent to non-linear statistical treatment and a substantial independence of model prediction from the structure of the test molecule. In the above analysis, the descriptive power of a given model has been measured by its ability to predict RT of unknown drug molecules.

Conclusion

The GA-PLS, GA-KPLS and L-M ANN models were applied for the prediction of the RT values of veterinary drugs in milk. High correlation coefficients and low prediction errors confirmed the good predictability of models. All methods seemed to be useful although a comparison between these methods revealed the slight superiority of the L-M ANN over other models. Application of the developed model to a testing set of 30 compounds demonstrates that the new model is reliable with good predictive accuracy and simple formulation. The QSRR procedure allowed us to achieve a precise and relatively fast method for determination of RT of different series of pesticides to predict with sufficient accuracy the RT of new compounds.

Acknowledgements

The authors thank Payame Noor University for their partial support on this work.

References

[1] A.A.M. Stolker, U.A.Th. Brinkman, *J. Chromatogr. A*, **2005**, *1067*, 15-24.

[2] H.P. Tang, C. Ho, S.S.L. Lai, *Rapid Commun. Mass Spectrom*, **2006**, *20*, 2565-2572.

[3] European Parliament and the Council of the European Union. Commission Regulation N° 124/2009 of 10 February **2009** setting maximum levels for the presence of coccidiostats or histomonostats in food resulting from the unavoidable carry-over of these substances in non-target feed. L 40.

[4] European Parliament and the Council of the European Union. Commission Regulation (EU) No 37/2010 of 22 December **2009** on pharmacologically active substances and their classification regarding maximum residue limits in foodstuffs of animal origin. L 15.

[5] S.L. Stead, H. Ashwin, S.F. Richmond, M. Sharman, P.C. Langeveld, *Int. Dairy J*, **2008**, *18*, 3–11.

[6] D. L. Borneman, S. Ingham, *Food Polic*, **2014**, *46*, 150-156.

[7] P. Anttila, T. Hyötyläinen, A. Heikkilä, M. Jussila, J. Finell, M. Kulmala, M.-L. Riekkola, *J. Sep. Sci.*, **2005**, *28*, 337-348.

[8] J. Pol, B. Hohnova, M. Jussila, T. Hyotylainen, *J. Chromatogr. A*, **2006**, *1130*, 64-75.

[9] B.G. Knecht, A. Strasser, R. Dietrich, E. Märtlbauer, R. Niessner, M.G.Weller, *Anal. Chem*, **2004**, *76*, 646-653.

[10] J. Lamar, M. Petz, *Analytica Chimica Acta*, **2007**, *586*, 296-309.

[11] G. Balizs, A. Hewitt, *Analytica Chimica Acta*, **2003**, *492*, 105-116.

[12] F.J. Lara, A.M. García-Campana, F. Alés-Barrero, J.M. Bosque-Sendra, L.E. GarcíaAyuso, *Anal. Chem.*, **2006**, *78,* 7665-7673.

[13] M.A. García-Mayor, R.M. Garcinu˜no, P. Fernández-Hernando, J.S. Durand-Alegría, *J. Chromatogr. A*, **2006**, *1122*, 76-89.

[14] E.A. Christodoulou, V.F. Samanidou, *J. Sep. Sci.*, **2007**, *30*, 2421-2429.

[15] R.C. Rodríguez-Díaz, J.M. Fernández-Romero, M.P. Aguilar-Caballos, A. Gómez-Hens, *J. Agric. Food Chem.*, **2006**, *54*, 9670-9678.

[16] N.A. Karaseva, T.N. Ermolaeva, *Talanta*, **2014**, *120*, 312-317.

[17] Y. Zhou, Sh. X Huang, J.X. Pu, J. R Li, L. Sh. Ding, D. F. Chen, *J. Pharm. Biomed. Anal.*, **2011**, *56*, 916– 927.

[18] G. Mirivel, V. Riffault, J. Galloo, *J. Chromatogr. A*, **2009**, *1216*, 6481–6489.

[19] G. Du, H.Y. Zhao, Q.W. Zhang, G.H. Li, F.Q. Yang, Y. Wang, Y.C. Li, Y.T. Wang, *J. Chromatogr. A*, **2010**, *1217*, 705–714.

[20] P. R. Duchowicz, M. A. Giraudo, E. A. Castro, A. B. Pomilio, *Chem. Intell. Lab. Sys.*, **2011**, *107*, 384-390.

[21] M. Goodarzi, T. Chen, M. P. Freitas, *Chem. Intell. Lab. Sys.*, **2010**, *104*, 260-264.

[22] T. Liu, I.A. Nicholls, T. Öberg, *Analytica Chimica Acta*, **2011**, *702*, 37-44.

[23] R. Kaliszan, P. Wiczling, M. J. Markuszewski, M.A. Al-Haj, *J. Chromatogr. A*, **2011**, *1218*, 5120-5130.

[24] C.I. De Matteis, D. A. Simpson, S. W. Doughty, M. R. Euerby, P. Nicholas Shaw, D. A. Barrett, *J. Chromatogr. A*, **2010**, *1217*, 6987-6993.

[25] K. Héberger, I. G. Zenkevich, *J. Chromatogr. A*, **2010**, *1217*, 2895-2902.

[26] A. A. D'Archivio, A. Incani, F. Ruggieri, *J. Chromatogr. A*, **2011**, *13*, 456-467.

[27] R. Todeschini, V. Consonni, A. Mauri, M. Pavan, DRAGON-Software for the calculation of molecular descriptors; Version 3.0 for Windows, **2003**.

[28] D. Ortelli, E. Cognard, Ph. Jan, P. Edder, *J. Chromatogr. B*, **2009**, *877*, 2363–2374.

[29] E. Sarıpınar, N. Geçen, K. ahin, E. Yanmaz, *Euro. J. Med. Chem.*, **2010**, *45*, 4157-4168.

[30] S. Sagrado, M.T.D. Cronin, *Analytica Chimica Acta*, **2008**, *609*, 169-174.

[31] E. A. Hernández-Caraballo, F. Rivas, A.G. Pérez, L.M. Marcó-Parra, *Analytica Chimica Acta*, **2005**, *533*, 161-168.

[32] V. K. Gupta, H. Khani, B. Ahmadi-Roudi, Sh. Mirakhorli, E. Fereyduni, Sh. Agarwal, *Talanta*, **2011**, *83*, 1014-1022.

[33] T. Bolan a, Š. Cerjan-Stefanovi , M. Regelja, H. Regelja, S. Lon ari , *J. Chromatogr. A*, **2005**, 1085, 74-85.

[34] M. Arab Chamjangali, M. Beglari, G. Bagherian, *J. Molecul. Graph. Model*, **2007**, *26*, 360-367.

[35] R. Jia, Zh. Mao, Y. Chang, Sh. Zhang, *Chem. Intell. Lab. Sys.*, **2010**, *100*, 91-98.

[36] M. Jalali-Heravi, A. Kyani, *Euro. J. Med. Chem.*, **2007**, *42*, 649-659.

[37] H. Noorizadeh, M. Noorizadeh, *Med Chem. Res.*, **2011**, *18*, 324-337.

[38] H. Chen, *Analytica Chimica Acta*, **2008**, *609*, 24–36.

[39] O. Deeb, *Chem. Intell. Lab. Sys.*, **2010**, *104*, 181-194.

[40] Validation of Analytical Procedures; Methodology, International Conference on Harmonization (ICH): Text and Methodology Q2 (R 1): Complementary Guideline on Methodology dated 06 November 1996: incorporated in November **2005**, London.