

Scientific paper

# Ranking of QSAR Models to Predict Minimal Inhibitory Concentrations Toward *Mycobacterium tuberculosis* for a Set of Fluoroquinolones

Marjan Vračko,<sup>1,\*</sup> Nikola Minovski<sup>1</sup> and Károly Héberger<sup>2</sup><sup>1</sup> Kemijski inštitut/National Institute of Chemistry, Hajdrihova 19, 1000 Ljubljana, Slovenia<sup>2</sup> Chemical Research Center, Hungarian Academy of Sciences, H-1025 Budapest, Pusztaszeri út 59/67, Hungary

\* Corresponding author: E-mail: marjan.vracko@ki.si

Fax: + 386 1 4762 300, Phone: +386 1 4762 315

Received: 31-05-2010

*This paper is dedicated to Professor Milan Randić on the occasion of his 80th birthday*

## Abstract

CP-ANN technique was used to build 54 different QSAR models. The models were built for three sets (assays) of fluoroquinolones considering their antituberculosis activity and using different technical parameters (dimension of network and number of learning epochs). The models served as a reliable basis for ranking by a new powerful method based on sum of ranking differences (SRD). With the applied SRD procedure we can find the optimal ones. The best model can be selected easily for the first assay. Two models can be recommended for the second assay, and no recommended model was found for the assay 3.

**Keywords:** QSAR, modeling, ranking, method comparison, neural networks, prediction of antituberculosis activity

## 1. Introduction

In the age of antibiotics the tuberculosis (TBC) still represents a global health problem. It treats the developing countries, where 95% of cases and 98% deaths occur, and also developed world mostly due its fatal combination with HIV infections.<sup>1</sup> The disease is mainly caused by the microorganism *Mycobacterium tuberculosis* or in some cases by *Mycobacterium fortuitum*, *Mycobacterium smegmatis*, or *Mycobacterium avium-intracellulare complex*. In combat of TBC the chemotherapy plays the major role.<sup>2</sup> Because of the drug resistance, which has been developed by microorganisms, the searching for new drugs represents an important task in this combat. Some fluoroquinolones seem to be promising candidates.<sup>3</sup> In this report we present an analysis of Quantitative Structure-Activity Relationship (QSAR) models, which were developed to estimate the activity of fluoroquinolones against the *Mycobacterium tuberculosis*. The details on data sets, which were used to build the models, are given in Section 3. As the modeling technique we selected the Counter

Propagation Artificial Neural Networks (CP-ANN), which are briefly described in Section 2.<sup>4-6</sup> During the development of a QSAR model many models are constructed and at the end the crucial question arises, how to select the best final model? In this report we applied a novel procedure the Sum of Ranking Differences (SRD) method to rank the models. The ranking was validated by comparison of ranks by random numbers (a kind of simulation test).<sup>7,8</sup> The aim of this study was to find the models, which can be used for further application, i.e., the prediction of activity for new (untested) compounds.

## 2. Methods

### 2.1. Modeling Algorithm

CP-ANN is often used in QSAR applications. Its architecture and learning strategy is described in many textbooks and articles, and therefore, we provide here only its brief description.<sup>4-6</sup> The CP-ANN is a generalization of Self Organizing Map (SOM), which is indeed a mapping from multidimensional descriptor space on two-di-

mensional map of neurons. It is an iterative procedure taking many learning epochs. The dimension of the network and the number of learning epochs must be optimized to get the optimal performance of the model. In our CP-ANN procedure the SOM technique was applied for division of data into training set and the test set.<sup>9,10</sup> In this procedure entire map is divided into several sub-parcels and from each sub-parcel some objects are randomly selected into training set and remaining objects to the test set. It is ensured that both sets cover entire information space equivocally.<sup>11–13</sup> It is important to emphasize that in SOM only independent variables (molecular descriptors) participate in the training. The CP-ANN is a generalization of SOM in a way that it includes additionally the property variables (activity values) in a training procedure.

## 2. 2. Sum of Ranking Differences

The method has recently been introduced in the field of chemistry.<sup>7</sup> The ranking of models is carried out by the following way: first, the data should be arranged in a matrix form, chemical compounds were enumerated in the rows, whereas models to be compared were arranged in the columns. Each matrix element contains the biological activity values (minimal inhibitory concentrations –MIC). The original measured data were also included among the ‘models to be compared’ for checking purposes. The row averages (consensus, ensemble averaging)<sup>13–15</sup> were selected as reference values. Each (individual) model is ranked and compared to known, reference values. Second, the absolute values of the differences between the reference and individual rankings are calculated and summed for each model and, additionally, for experimental values. In such a way the sum of ranking differences, SRD values were calculated for each model and for the experimental values. The closer is the SRD value to zero (i.e. the closer is the ranking to the reference value) the better is the model. The proximity of SRD values shows that the models predict similarly the ordering of MIC values. Groupings of models can also be observed, whereas their distance show dissimilarity from the ordering point of view, i.e. SRD can be considered as a dissimilarity measure (the smaller its value, the more similar to the reference value).<sup>7,8</sup>

## 2. 3. Comparison of Ranks by Random Numbers

Validation of the SRD procedure can be easily carried out by the theoretical distribution of the (random) SRD values corresponding to the number of compounds, using simulated random numbers, where simulated “models” are formed with the same number of compounds. The SRD procedure can be carried out for simulated “models” many times. When the number of simulated random “models” is quite large, (above hundred thousands) the results can be used to characterize the theoretical SRD di-

tribution function with very small error. The theoretical discrete distribution has been calculated earlier using a recursive algorithm if the number of compounds was small ( $n < 14$ ). The normal distribution was applied to approximate the theoretical (random) SRD distribution function for large number of compounds, if  $n > 13$ .<sup>8</sup>

## 3. Data and Models

The data, which consist of chemical structures and inhibitory activities against *Mycobacterium tuberculosis*, were collected from the Internet. The activity is expressed as the Minimal Inhibitory Concentration (MIC ( $\mu\text{g/mL}$ )). Three data sets (assays) were collected from internet applying different searching criteria. Assay 1 consists of 66 compounds and was collected using the searching criterion “fluoroquinolones” in NIAID database.<sup>16</sup> Assay 2 consists of 145 fluoroquinolones selected from NIAID using the criteria “antituberculous agent”. As several structures have the same MIC values we eliminated some structures getting the reduce set of 66 compounds (assay 3). In the next step the molecular descriptors were calculated with DRAGON and CODESSA software.<sup>17,18</sup> All descriptors were imported to CODESSA where the heuristic option was applied for searching of the best ones. On this way, for each assay, ten descriptors have been selected, which have been used in CP-ANN.<sup>19</sup> Applying the SOM technique each of the three assays was divided into the training and the test set. In the assay 1, 48 compounds were selected into training set and 18 into the test set. In the assay 2 and assay 3 the divisions were: training sets 115 and 46 and test sets 30 and 20, respectively. In our ranking procedure 16 models for each assay have been taken into account. For assays 1 and 3 the dimensions  $7 \times 7$ ,  $8 \times 8$ ,  $9 \times 9$ , and  $10 \times 10$  were considered taking 400, 600, 800, and 1000 learning epochs. For assay 2 the learning epochs were the same, on the other hand, the selected dimensions were  $11 \times 11$ ,  $12 \times 12$ ,  $13 \times 13$ , and  $14 \times 14$ . The details of models are given in the reference.<sup>19</sup> In this report we present the ranking of models using the SRD method. The strategy of this method was outlined above.<sup>7,8</sup>

## 4. Results and Discussion

Assay 1: The ranking by SRD for the training and test sets are given in table 1. It is obvious that model A5 is the best one considering the training and test sets. Similarly it is striking how far the simulated random SRDs (the distribution is characterized by percentiles, median, etc.) are from the SRDs of models. Any statistical test would suggest significant difference between them. The SRDs of models, however, are not significantly different from each other. In the test set model A5, A6, A7, and A8 are equivalent according to the SRD ranking. They can be distin-

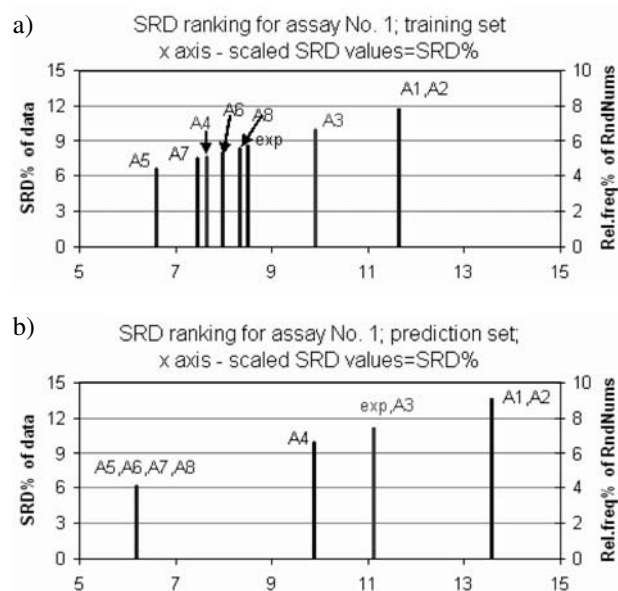
**Table 1:** Sum of ranking differences for assay 1, training (48) and test (18) sets

Name	A5	A7	A4	A6	A8	exp	A3	A1	A2	XX1	Q1	Med	Q3	XX19
SRDtrain	76	86	88	92	96	98	114	134	134	650	720	766	812	882
Name	A5	A6	A7	A8	A4	exp	A3	A1	A2	XX1	Q1	Med	Q3	XX19
SRDtest	10	10	10	10	16	18	18	22	22	80	96	108	120	136

XX1–first icosaille (5%), Q1 – first quartile, Med – median, Q3 – last quartile, XX19 – last icosaille (95%)

guished from each other using more information (larger number of compounds – training set)  $A7 > A4 > A6 > A8$  (where the sign “>” means that the previous model is better than the later one). The model A4 is better for the training set than for the test set. The experimental data were included into the ranking procedure for checking purposes: models A1, A2 and A3 are not better than the measured data, their usage should be avoided.

The raw SRD values for training and test sets cannot be compared directly because the number of compounds was different. Therefore we recalculated the SRD values for the same scale between 0 and 100. Figure 1 visualizes the ranking of models by scaled SRD values.

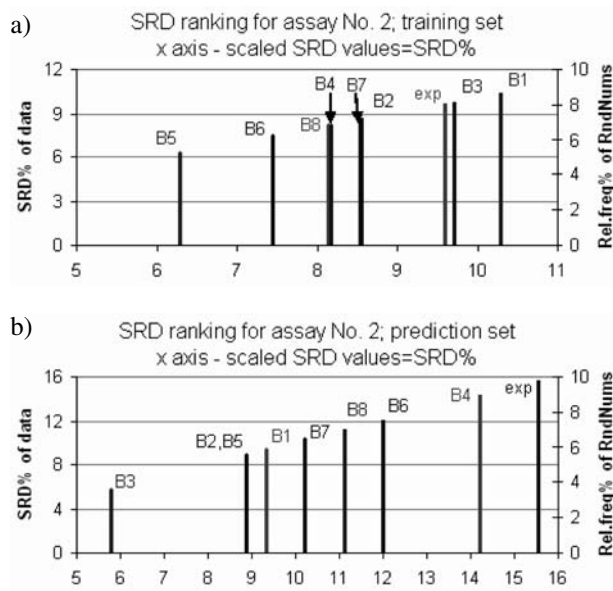


**Figure 1a, b.** Scaled SRD values for predictive models for assay 1 for the training set (a) and the test set (b). The mean and standard deviation of the validation distribution are given in the header.

If the prediction capacity is modeled by the test set the equivalent models are to be used on equal bases. However, there is no doubt that Model A5 is the best one and we recommend its usage.

Assay 2: The ranking by SRD for the training and test sets are given in table 2. Model B5, which is the best one for training set, is not consistently the best one in this case. As the number of compounds are larger than in case

of assay 1 the difference between simulated random SRDs (the distribution is characterized by percentiles, median, etc.) and SRDs of models are larger and could be identified by statistical tests as in previous case: the SRDs of models are not significantly different from each other. In the test set the model B3 is the best one, it is clearly distinguished from the other models. In the training set the same model is worse than the experimental one showing that the split of training and test sets are far from being optimal. The experimental values are ranked to the latest position, showing a clear over-fitting effect for all of the models on the test set. The SRD ranking shows:  $B5 > B6 > B8 > B4 > B7 > B2$  for the training set whereas the same



**Figure 2a, b.** Scaled SRD values for predictive models for assay 2 for the training set (a) and the test set (b). The mean and standard deviation of the validation distribution are given in the header.

for the test set is slightly different:  $B3 > B2 > B5 > B1 > B7 > B8 > B6 > B4$ . The scaled SDR values are shown in Figure 2.

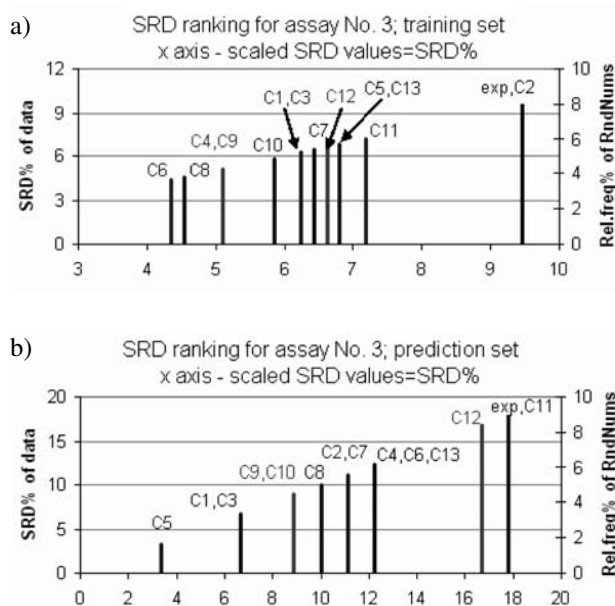
The models of assay 2 should be handled with caution. Models B5 and B8 could be recommended for further applications. On the other side, the models B1, B3, B4, and B6 are not recommended.

Assay 3: The ranking by SRD for the training and test

**Table 2:** Sum of ranking differences for assay 2, training (115) and test (30) sets

Name	B5	B6	B8	B4	B7	B2	Exp	B3	B1	XX1	Q1	Med	Q3	XX19
SRDtrain	416	492	538	540	564	566	634	642	680	3802	4132	4396	4596	4926
Name	B3	B2	B5	B1	B7	B8	B6	B4	exp	XX1	Q1	Med	Q3	XX19
SRDtest	26	40	40	42	46	50	54	64	70	240	272	300	322	358

For notation see Table 1.



**Figure 3a, b.** Scaled SRD values for predictive models for assay 3 for the training set (a) and the test set (b). The mean and standard deviation of the validation distribution are given in the header.

**Table 3:** Sum of ranking differences for assay 3, training (46) and test (20) sets

Name	C6	C8	C4	C9	C10	C1	C3	C7	C12	C5	C13	C11	exp	C2	XX1	Q1	Med	Q3	XX19
SRDtrain	46	48	54	54	62	66	66	68	70	72	72	76	100	100	588	662	704	746	810
Name	C5	C1	C3	C9	C10	C8	C2	C7	C4	C6	C13	C12	exp	C11	XX1	Q1	Med	Q3	XX19
SRDtest	6	12	12	16	16	18	20	20	22	22	22	30	32	32	90	108	120	132	150

For notation see Table 1.

sets are given in table 3. It is difficult to select the best models consistently for the training and test sets. Most models are better than the experimental data (except C2 for training set and C11 for the test set) what indicates the over-fitting effect. The models C6, C8 and C9 might be selected as acceptable ones. The difference between simulated random SRDs (the distribution is characterized by percentiles, median, etc.) is significant. On the other side, the differences among SRD values of models are less significant, however, some groupings can be observed. The best models for training and test sets are different indicating that the split of training and test sets are far from being optimal. The scaled SRD values for assay 3 are shown in Figure 3. The comparison of models for assay 3 shows that no unambiguous selection of the optimal model can be made.

## 5. Conclusions

The SRD method was applied to select the optimal QSAR model. This analysis was performed on training and test sets. For the first assay we found one model, which could be selected as the best one and recommended for further applications. For the second assay we could select two models, which could be recommended for further usage. The SRD analysis shown some weak points of models, like over-fitting and not optimal training/test set division. In the third assay the different models were found as the best ones for training and test sets. It follows that no unambiguous selection for an optimal model can be made.

## 6. Acknowledgement

Authors thank Agency of Research of R. Slovenia (ARRS) for the financial support through the Grants P1-0017 and 1000-07-310016. One author (KH) thanks for the support by TÉT No. 8/2010 (Croatian–Hungarian).

## 7. References

1. C. Dye, S. Scheele, P. Dolin, V. Pathania, M. C. Raviglione, *JAMA- J. Am. Med. Assoc.* **1999**, 282, 677–686.
2. Y. L. Janin, *Bioorg. Med. Chem.* **2007**, 15, 2479–2513.
3. K. Drlica, X. Zhao, *Microbiol. Mol. Biol. Rev.* **1997**, 61, 377–392.
4. J. Zupan, J. Gasteiger, Neural networks in chemistry and drug design. Weinheim Wiley-VCH, Weinheim, **1999**, p.380.
5. J. Zupan, M. Novič, J. Gasteiger, *Chemometr. Intell. Lab. Syst.* **1995**, 27, 175–187.
6. M. Vračko, *Curr. Comput.-Aid. Drug Des.* **2005**, 1, 73–78.
7. K. Héberger, *TrAC Trends Anal. Chem.* **2010**, 29, 101–109.
8. K. Héberger, K. Kollár-Hunek, *J. Chemometr.* (in press, **2010**).

9. V. Simon, J. Gasteiger, J. Zupan, *J. Am. Chem. Soc.* **1993**, *115*, 9148–9159.
10. I. Valkova, M. Vračko, S. C. Basak, *Anal. Chim. Acta* **2004**, *509*, 179–186.
11. J. T. Leonard, K. Roy, *QSAR Comb. Sci.* **2006**, *25*, 235–251.
12. M. Kotnik, M. Oblak, J. Humljan, S. Gobec, U. Urleb, T. Solmajer, *QSAR Comb. Sci.* **2004**, *23*, 399–405.
13. P. Gramatica, P. Pilutti, E. Papa, *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1794–1802.
14. T. Hastie, R. Tibshirani, J. Friedman, in: *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. (2<sup>nd</sup> edition), Springer, New York, **2009**, pp 261–294.
15. P. Gramatica, E. Giani, E. Papa, *J. Mol. Graphics Modell.* **2007**, *25*, 755–766.
16. NIAID-Division of AIDS, HIV/OI/TB Therapeutics Database <http://chemdb2.niaid.nih.gov>. (accessed: May 27, **2010**).
17. Talete, srl. DRAGON for Windows (Software for Molecular Descriptor Calculations), version 5.4. **2006**.
18. A. R. Katritzky, V. S. Lobanov, M. Karelson, CODESSA, Training Manual. Gainsville: s.n., **1995**.
19. N. Minovski, M. Vračko, T. Šolmajer, Quantitative structure-activity relationship study of antitubercular fluoroquinolones. *Mol. Divers.* (in press **2010**) DOI: 10.1007/s11030-010-9238.

## Povzetek

V prispevku smo analizirali 54 QSAR modelov, ki so bili narejeni s protitočnimi nevronske mreže (CP-ANN). Modeli so bili narejeni na treh nizih fluorokinolonov kot potencialnih antituberkuloznih učinkovinah. Modele smo razvili ob upoštevanju različnih tehničnih parametrov, kot sta dimenzija nevronske mreže in število učnih epoh. Na nizu modelov smo uporabili pred kratkim predlagano metodo za hierarhično urejanje modelov (SRD), s katero lahko poiščemo najbolj zanesljive modele. Z metodo SRD smo za prvi niz izbrali en model, za drugi niz smo predlagali dva modela, za tretji niz pa nismo našli modela, ki bi bil boljši kot ostali modeli.