**Response to reviewers**

Thank you for your suggestions. We corrected and improved our manuscript according to your observations and we hope that in this form the manuscript is totally in accordance with your expectation in order to adhere to the style accepted by the Journal of the Serbian Chemical Society.

The new paragraphs introduced in the manuscript are highlighted in yellow. More details and information have been added for clarification in Supplementary Material. Some of the references have been transferred to the section Supplementary Material and consequently those in the manuscript have been renumbered.

For reviewer B

We thank the reviewer for kind and positive feedback.

-The manuscript is considered with an appropriate length. However, there are too many "in line" citations to Schrodinger site. It would be better to cite in references the full software an in text to specify the procedure with a separate reference if exists (for instance the PHASE procedure is again cited in reference 35). This would give a much clear text.

We have remedied this issue accordingly. We added separate references for the software used.

- On the other hand, there are a lot of numerical results, mainly
statistical parameters but no visible supporting data. A table would be
useful at least for one set of data as an example or if there are too many
involved data a supplemental file.

We have incorporated the statistical data into Table 1. Also we added a Supplementary Material where detailed explanations along with graphics and tables were presented.

For reviewer C

We are grateful to the reviewer for useful and interesting remarks, and we tried to address all of them as described below.

 The software operation of pharmacophore generation protocol is very
trivial, a lot of which can be moved to supporting information, but the core idea and details for the method procedure are not described clearly.

The software operation of pharmacophore generation protocol was moved to Supplementary Material. More details and information were added for clarification.

 The section title “EXPERIMENTAL” should not be used for a pure
computational study, because it may mislead readers.

The section title “EXPERIMENTAL” has been replaced by “Computational Methods”.

 The sample set of 26 compounds is too small to perform a statistically significant QSAR modeling, which should be at least more than 30 samples. Considering that the sample set is small, random splitting of training and test sets are not reliable because it may obtain a good result by accident; some other methods such as D-optimal and k-NN may be more useful for the current problem.

Indeed, the Phase software has the option to split randomly the training and test sets, but our approach was a little different. We followed the recommendation of Golbraikh papers [1-3]:

(1) “On the basis of our analysis, we suggest that the test set must include no less than five compounds, whose activities and structures must cover the range of activities (see Figure S1) and structures of compounds from the training set.”

(2) “Ideally, the division into the training and test set must satisfy the following three conditions: (i) All representative compound-points of the test set in the multidimensional descriptor space must be close to those of the training set. (ii) All representative points of the training set must be close to those of the test set. (iii) The representative points of the training set must be distributed within the whole area occupied by the entire dataset.” To verify this in our case we have calculated the similarity between our training and test sets:

1. A. Golbraikh, A. Tropsha, *J. Mol. Graph. Model.* **20** (2002) 269 ([https://doi.org/10.1016/S1093-3263(01)00123-1](https://doi.org/10.1016/S1093-3263%2801%2900123-1))

2. A. Golbraikh, A. Tropsha, *J. Comput. Aided Mol. Des.* **16** (2002) 357 ([https://doi.org/10.1023/A:1020869118689](https://doi.org/10.1023/A%3A1020869118689))

3. A. Golbraikh, M. Shen, Z. Xiao, Y.-D. Xiao, K.-H. Lee, A. Tropsha, *J. Comput. Aid. Mol. Des.* **17** (2003)241 (https://doi.org/10.1023/A:1025386326946)



Figure S1. Histogram of dataset pIC50 distribution.

 Peter Willett [4] demonstrates that the well-known Tanimoto coefficient remains the method of choice for the computation of fingerprint-based similarity. The Tanimoto coefficient was calculated in order to have a quantitative basis, a similarity measure, and to assess the degree of resemblance between the training set and the test set. The Tanimoto similarity was computed using MACCS fingerprints, calculated with KNIME. The test -training pairwise similarity values display a distribution shifted toward high values (0.85–1)(Figure S2).



Figure S2. The distribution of 2D Tanimoto coefficients values for training and test sets

4. P. Willett, *Drug. Discov. Today.* **11** (2006) 1046 [10.1016/j.drudis.2006.10.005](https://doi.org/10.1016/j.drudis.2006.10.005)

In addition, we used the Euclidean Distance algorithm to work out the similarity between each two pairs of compounds by computing the score for each pair of nodes. The zero value for Euclidean Distance means absolute identity. For the training and test sets the distribution of values for Euclidean Distance are similar ((Figure S3).



Figure S3. The distribution of Euclidean Distance values for training and test sets

Furthermore, the median values for the most important 2D properties of molecules were calculated. In our case, the median value for training and test denote that the two sets of molecules are rather similar in character (Table S1).

Table S1. Characteristics (minim, maxim and median values) of the training and test set molecules

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Name | MW | NRS | RBN | RB | XLogP | 2d PSA |
| TRAININGSET | MIN | 370.86 | 3 | 3 | 26 | 1.41 | 68.32 |
| MAX | 482.02 | 4 | 6 | 34 | 4.12 | 90.08 |
| **MEDIAN** | **452.405** | **4** | **5** | **31** | **2.865** | **77.19** |
| TEST SET | MIN | 425.93 | 3 | 4 | 27 | 1.24 | 60.51 |
| MAX | 495.93 | 4 | 6 | 34 | 3.93 | 90.08 |
| **MEDIAN** | **450.94** | **4** | **5** | **31.5** | **2.75** | **77.19** |

MW- molecular weight; NRS - number of ring systems; RBN -rotatable bonds; RB - rigid bonds

We included the abovementioned data in the Supplementary Material.

 The language should be improved substantially. It feels hard to read.

We regret there were problems with the English. The paper has been revised to improve the grammar and readability.