Journal of the Serbian Chemical Society
JSCS-info@shd.org.rs • www.shd.org.rs/JSCS
J. Serb. Chem. Soc. 85 (3) S165–S170 (2020) *Supplementary material*

SUPPLEMENTARY MATERIAL TO

In silico studies on smoothened human receptor and its antagonists in search of anticancer effects

ANA BOROTA*, SORIN AVRAM, RAMONA CURPAN, ALINA BORA, DANIELA VARGA, LILIANA HALIP and LUMINITA CRISAN**

"Coriolan Dragulescu" Institute of Chemistry, Romanian Academy, 24 Mihai Viteazul Avenue, RO-300223, Timisoara, Romania

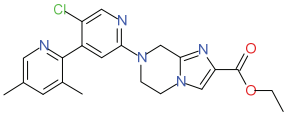
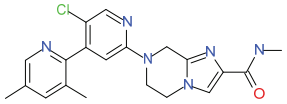
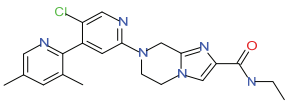
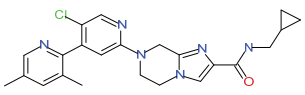
J. Serb. Chem. Soc. 85 (3) (2020) 335–346

PHARMACOPHORE GENERATION PROTOCOL

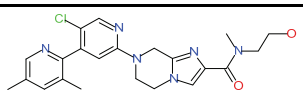
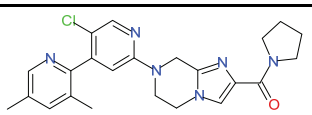
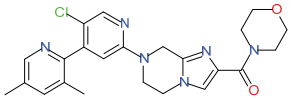
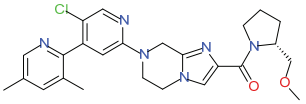
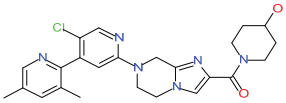
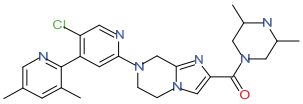
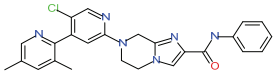
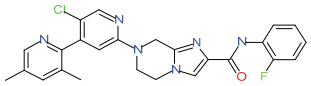
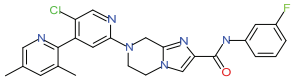
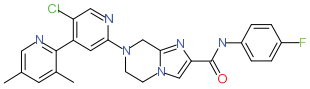
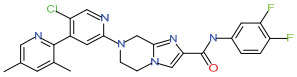
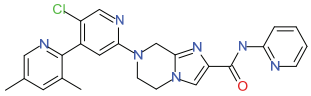
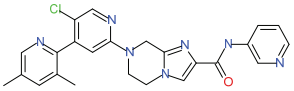
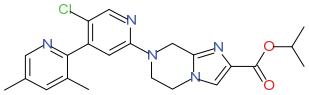
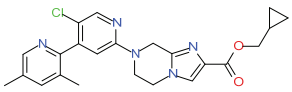
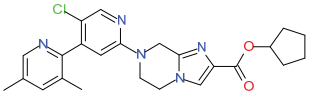
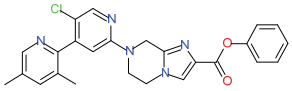
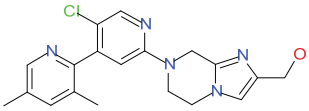
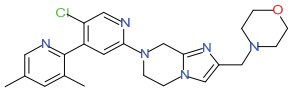
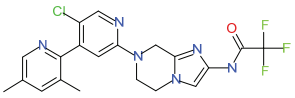
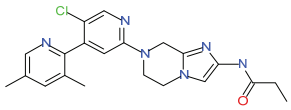
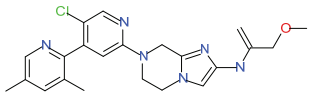
The ligands preparation for *in silico* studies was realized using LigPrep (Schrödinger),[1] by following the steps: optimization of the structures with the OPLS_2005 force field, ionization with Epik at pH 7.0, and generation of stereoisomers for the structures with unspecified chiralities.

ConfGen (Schrödinger)[2] was engaged in the generation of multiple conformers for each compound using default settings. The compounds were considered active if the pIC_{50} value is >6.3 and inactive if the pIC_{50} value is <5.8. The maximum number of pharmacophore sites was set to four and all the active compounds matched the common pharmacophore hypotheses further obtained.

TABLE S-I. The structure of the compounds and their experimental SMO inhibitory activity expressed in logarithmic units (pIC_{50}); the compounds structures were drawn with MarvinSketch (ChemAxon) (http://www.chemaxon.com)

No.	Structure	pIC_{50}	No	Structure	pIC_{50}
1		6.161	2[b]		5.568
3		5.886	4		5.283

*,** Corresponding authors. E-mail: (*)ana_borota@acad-icht.tm.edu.ro; (**)lumi_crisan@acad-icht.tm.edu.ro

5[a,b]		5.387	6[b]		6.568
7		6.102	8		6.259
9		5.920	10		5.823
11[a,b]		6.522	12[b]		6.455
13[a]		7.060	14[b]		6.356
15[a,b]		6.376	16[b]		6.568
17[a,b]		6.698	18[b]		6.420
19[a]		6.568	20		6.130
21		6.017	22[b]		5.387
23[b]		6.040	24[b]		5.744
25		6.050	26[c]		6.187

[a]The compounds used for decoys generation; [b]the test set compounds of the 3D-QSAR model

Test set selection

In order to split the dataset in training and test sets, we followed the recommendation of Golbraikh papers:[3–5]

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.

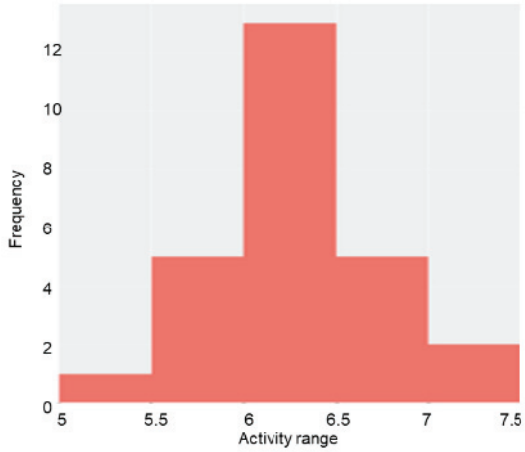


Fig. S-1. Histogram of dataset pIC50 distribution.

Peter Willett[6] 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.[7] The test–training pairwise similarity values display a distribution shifted toward high values (0.85–1, Fig. S-2).

In addition, we have 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 (Fig. S-3).

Furthermore, the median values for the most important 2D properties (FILTER (OpenEye))[8] 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 S-II).

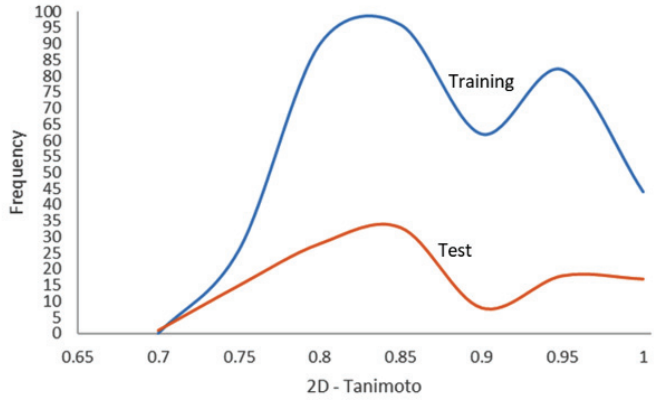


Fig. S-2. The distribution of 2D Tanimoto coefficients values for training and test sets.

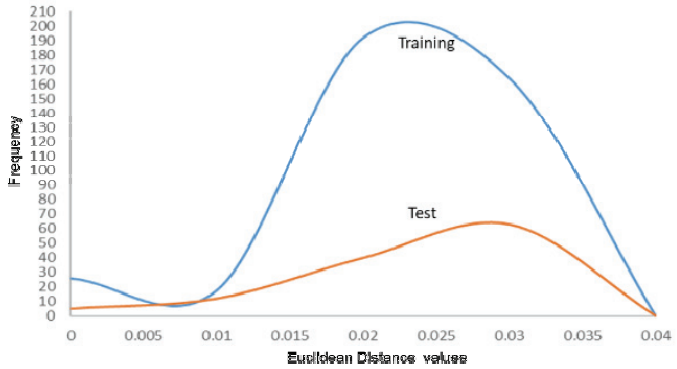


Fig. S-3. The distribution of Euclidean Distance values for training and test sets

TABLE S-II. Characteristics (min, max and median values) of the training and test set molecules; MW – molecular weight; NRS – number of ring systems; RBN – rotatable bonds; RB – rigid bonds

Set	Name	MW	NRS	RBN	RB	XLogP	2d PSA
Training	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**

Building of the external validation dataset

The external validation dataset was assembled by extracting data from the ChEMBL database.[9] Firstly, 818 compounds having inhibitory activity against SMO receptor were extracted. Then, the compounds with unspecified bioactivity IC_{50} against SMO receptor, as well as those having IC_{50} values expressed with a qualifier of type "<" and ">" were removed. Finally, the duplicates were discarded using Instant JChem software (Instant JChem v. 17.17.0, https://chemaxon.com/) and 689 active compounds were retained. Subsequently, drug-like filtering criteria, *e.g.*, HBA (hydrogen bond acceptors) = 3–4, HBD (hydrogen bond donors) = 0–3, MW (molecular weight) = 371–496, RBN (number of rotatable bonds) = 3–6, *X*Log*P* = −0.4–3.7, 2dPSA (2d polar surface area) = 60–90 were applied, resulting 179 compounds which have been designated as actives in the evaluation procedure.[9–11] In order to select the decoys, a free on-line automated tool from the Directory of Useful Decoys, enhanced (DUD-E, http://decoys.docking.org), was used.[12] The decoys set was compiled based on the similarities with the compounds used for pharmacophore generation (Table I), employed as queries. Thus, 50 decoys were selected for each query, with the exception of compounds **2** and **22** (for which 100 decoys were identified for each of them). The extracted decoys have similar physicochemical properties (MW, HBA, HBD, log*P* and RBN) with queries but dissimilar 2-D topology. Therefore, 179 compounds designated as actives and 1350 decoys entitled as inactive were used for virtual screening (VS) experiments in order to find matches over the obtained pharmacophore hypotheses. The fitness scores were used for the ranking of the compounds over the best pharmacophore hypothesis.

REFERENCES

1. Schrödinger Release 2016-1: LigPrep v3.1 (2016) Schrödinger, LLC, New York, NY
2. K. S.Watts, P. Dalal, R. B. Murphy, W. Sherman, R. A. Friesner, J. C. Shelley, *J. Chem. Inf. Model*. **50** (2010) 534 (https://doi.org/10.1021/ci100015j)
3. A. Golbraikh, A. Tropsha, *J. Mol. Graph. Model.* **20** (2002) 269 (https://doi.org/10.1016/S1093-3263(01)00123-1)
4. A. Golbraikh, A. Tropsha, *J. Comput.-Aided Mol. Des.* **16** (2002) 357 (https://doi.org/10.1023/A:1020869118689)
5. A. Golbraikh, M. Shen, Z. Xiao, Y.-D. Xiao, K.-H. Lee, A. Tropsha, *J. Comput.-Aided Mol. Des.* **17** (2003) 241 (https://doi.org/10.1023/A:1025386326946)
6. P. Willett, *Drug. Discov. Today* **11** (2006) 1046 10.1016/j.drudis.2006.10.005
7. S. Beisken, T. Meinl, B. Wiswedel, L. F. de Figueiredo, M. Berthold, C. Steinbeck, *BMC Bioinformatics* **14** (2013) 257 https://doi.org/10.1186/1471-2105-14-257
8. FILTER (2009), version 2.0.2 OpenEye Scientific Software, Inc. Santa Fe, NM, www.eyesopen.com
9. A. P. Bento, A. Gaulton, A. Hersey, L. J. Bellis, J. Chambers, M. Davies, et al., *Nucleic Acids Res.* **42** (2014) 1083 (http://dx.doi.org/10.1093/nar/gkt1031)
10. D. F. Veber, S. R. Johnson, H. Y. Cheng, B. R. Smith, K. W. Ward, K. D. Kipple, *J. Med. Chem.* **45** (2002) 2615 (http://dx.doi.org/10.1021/jm020017n)

11. Y. C. Martin, *J. Med. Chem.* **48** (2005) 3164 (http://dx.doi.org/10.1021/jm0492002)
12. M. M. Mysinger, M. Carchia, J. J. Irwin, B. K. Shoichet, *J. Med. Chem.* **55** (2012) 6582 (http://dx.doi.org/10.1021/jm300687e).