**Supporting information:**

**Benchmarks for interpretation of QSAR models**

Mariia Matveieva, Pavel Polishchuk\*

Institute of Molecular and Translational Medicine, Faculty of Medicine and Dentistry, Palacky University and University Hospital in Olomouc, Hnevotinska 5, 77900 Olomouc, Czech Republic

pavlo.polishchuk@upol.cz

Table S1. Correlations between count of patterns of interest for molecules of each regression data set and counts of the most common chemical elements.

|  |  |  |
| --- | --- | --- |
| Dataset | Pattern 1 (SMARTS) | Pattern 2 (SMARTS) |
| [C,c] | [O,o] | [S,s] | [N,n] | Cl | Br |
| N | [N,n] | 0.09 | -0.13 | 0.07 | 1 | -0.02 | -0.04 |
| N-O | [N,n] | 0.09 | 0.02 | 0.02 | 1 | -0.04 | -0.07 |
| [O,o] | 0.06 | 1 | 0.07 | 0.02 | -0.1 | -0.04 |
| N+O | [N,n]  | 0.23 | 1 | 0.1 | 1 | -0.11 | 0.0 |
| Amide | NC=O | 0.09 | 0.3 | 0.11 | 0.25 | 0.04 | 0.02 |



Figure S1. Class-wise distributions of hydrogen bond donors and acceptors for the pharmacophore data set



Figure S2. Architecture of Graph convolutional network

Figure S3. Distribution of predicted class probabilities by GC model for the pharmacophore data set