PPI Focused Libraries
Ligand-based approach

Protein-protein interactions (PPIs) are involved in important biological processes in life, and their regulation can be crucial for the treatment of numerous diseases. Low molecular weight PPI inhibitors able to selectively and potently modulate protein–protein interactions have recently reached clinical trials.

Life Chemicals has prepared following Libraries of potential PPI inhibitors using various features within ligand-based approach:

- PPI Focused Library Machine Learning (Decision Tree) Method
- PPI Focused Library 2D Similarity Search to Timbal DB
- PPI Focused Library 2D Similarity Search to Binding DB and Pubmed DB
- PPI Focused Library "Rule of four"

PPI Focused Library - Machine Learning (Decision Tree) Method

To predict which compounds could affect PPI a machine learning method (decision tree, DT) was used [1]. This method is known to be a useful tool to identify a PPI inhibitor profile. DT method is based on a cross-validation protocol to provide the balance between enrichment, sensitivity and specificity *on the learning data set*. With the comparison of unique physicochemical features of PPI and non-PPI inhibitors, several descriptors showing a correlation for PPI inhibitors in specified range of values were found:

- RDF 070m (≤ 3.31) is the shape-based descriptor which defines a radial distribution function of an ensemble of atoms in a spherical volume with radius 7Å.
- Ui (> 4.13) - an unsaturation index, directly linked to the number of multiple bonds, which contains double, triple and aromatic bonds
- SHP2 (≤ 0.30) - average shape profile index of order.2, deduced from the distance distribution of the geometry matrix
- Mor11m (> -0.1) - a descriptor which reflects a sum of atom weights with different angular scattering function

Filters applied to the entire Life Chemicals in-stock collection [1]:

- ClLogP = 1.5–4.5
- TPSA=75- 120
- MW =≤ 475
- HBD = 0 – 4
- HBA = 4 – 9
- PAINS filters

The resulted compounds were included in Life Chemicals PPI Machine Learning Method Library which finally comprised about **2,600** compounds (Fig. 1).

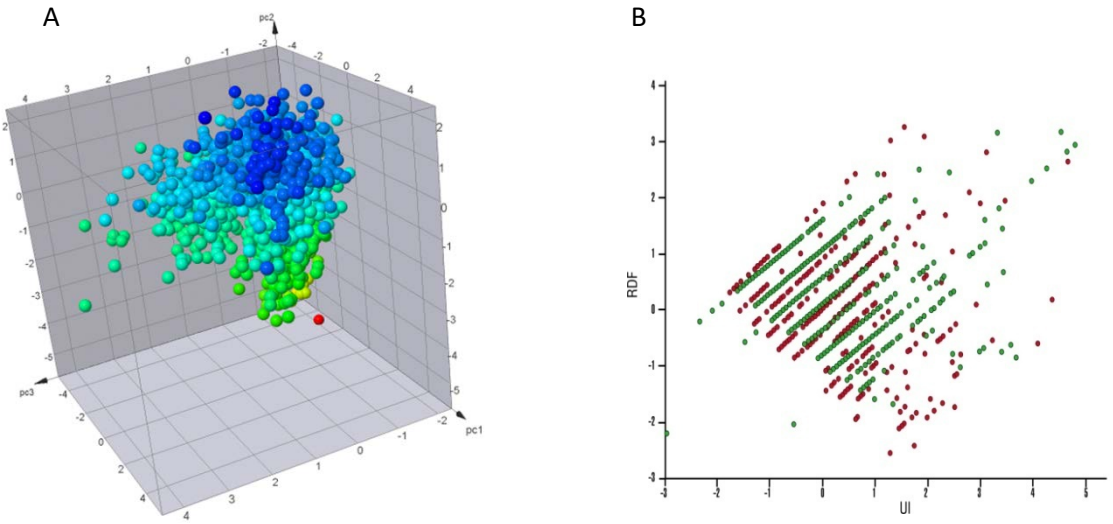

Fig. 1. A. Principal component analysis (PCA) showing accumulation of compounds with best matching to our parameters (see the list of parameters in the text). **B**. Distribution of compounds within determined values of descriptors. The plot was built to validate the method: red points correspond to the compounds obtained from Timbal database with molecular weight lower than 450, and green –to the Life Chemicals PPI Inhibitors Machine Learning Method Library. All descriptors were calculated with PyChem.

PPI Focused Library 2D Similarity Search to Timbal DB

About **1,500** compounds were extracted from Life Chemicals' stock HTS collection by 2D fingerprint similarity search towards Timbal DB [2] with Tanimoto 85% threshold.

PPI Focused Library 2D Similarity Search to Binding DB and Pubmed DB

With use of Binding DB and Pubmed DB a reference set of small organic molecules with activity in 28 PPI assays towards following 7 targets was collected: toll-like receptor 4; Hepatitis C virus core protein (dimerization inhibition); Tyrosine-protein kinase TYRO11; runt-related transcription factor 1 isoform AML1c; core-binding factor beta subunit isoform 1; mitogen-activated protein kinase 2 (MAP2); mitogen-activated protein kinase 3 (MAP3) [3,4]. After filtering and merging their activity data, resulting 10,000 unique substances were obtained and were further used as a basis for compound selection. The MDL public keys and the Tanimoto similarity cut-off 90% were applied to Life Chemicals' stock collection that afforded almost **23,000** compounds Library.

PPI Focused Library "Rule of four"

This library was created on the basis of the study done by X.Morelli, R.Bourgeas et al [5]. The analysis of 2P2I dataset (http://2p2idb.cnrs-mrs.fr) determined a group of structural and chemical features which were collectively named "Rule of Four" (Fig. 2). It has been shown that specific value range for AlogP/ClogP (ALOGP/CLOGP > 4), molecular weight (MW > 400), number of hydrogen bond acceptors (HDA > 4) and number of rings (Ring > 4) define the properties of PPI inhibitor. The rule was used as a filter to accelerate the process of identification of potential PPI inhibitors and gave rise to the library containing **4,300** compounds (Fig. 3). All the compounds were passed through PAINS filters.

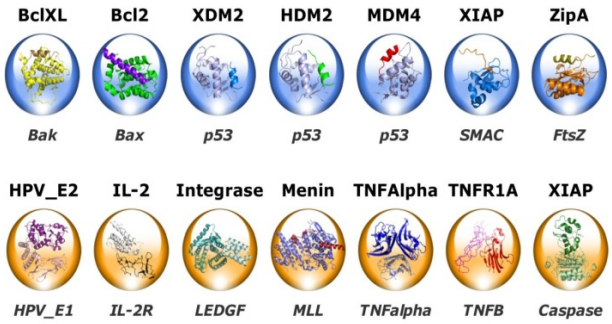

Fig. 2. A group of proteins presented in the study which revealed a concept of "Rule of Four".

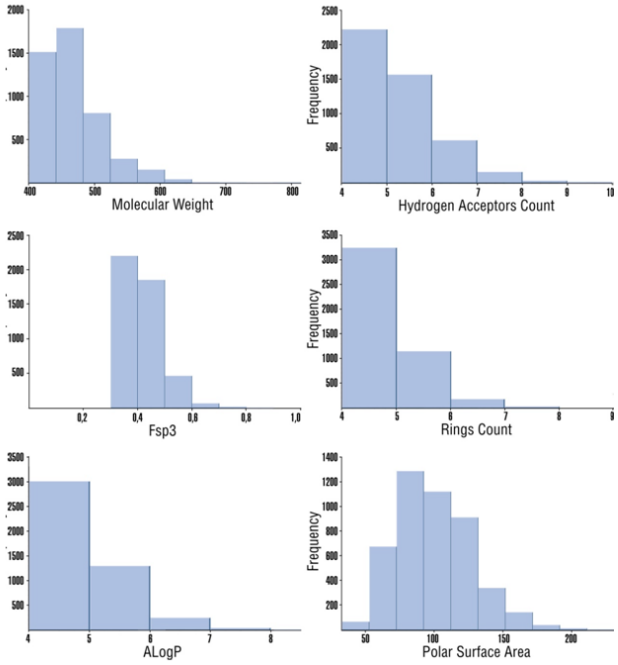

Fig. 3. Scatter plots prepared with SYBYL-X software which demonstrate distribution of potential PPI inhibitors according to the values of "Rule of Four" descriptors.

References

1. Designing focused chemical libraries enriched in protein-protein interaction inhibitors using machine learning methods. Reynès C, Host H, Camproux AC et al. PLoSComput Biol. 2010 Mar 5;6(3):e1000695. doi: 10.1371/journal.pcbi.1000695
2. http://www-cryst.bioc.cam.ac.uk/databases/timbal
3. http://www.bindingdb.org
4. http://www.ncbi.nlm.nih.gov/pccompound
5. Morelli X, Bourgeas R, Roche P: Chemical and structural lessons from recent successes in protein-protein interaction inhibition (2P2I). Curr Opin Chem Biol 2011,**15:**475-481

To download a file with compound structures for this library, please follow this link:

http://www.lifechemicals.com/downloads/Screening_Libs/13062/PPI_download