OTAVA chemicals

Compounds for HTS
Chemical building blocks
Fragment libraries
Targeted libraries
Drug discovery services

ML (MACHINE LEARNING) SARS TARGETED LIBRARY

OTAVA chemicals

OTAVA offers ML (Machine Learning) SARS Targeted Library. This library provides an excellent basis for the development of specific antiviral therapeutics toward SARS-CoV-2 (COVID-19).

The library consists of **1,577 compounds***.

All compounds are:

- **in stock**; available amounts: 1 – 50 mg
- **Drug-like only**; reactive, pan-assay interference (PAINS), redox-active and aggregator compounds were removed from the library.

QA/QC passed:

- minimal purity of compounds is **90%**;
- by **NMR** and/or **GC/LC/MS**
- **NMR spectra are available** upon request

Frendly packing services:

- **Cherry-picking is available**
- Supplied as dry powder or DMSO solution**
- Packaging in deep-well plates or barcoded vials***
- **Weighing out is free**

*Please note that the library does not contain known inhibitors. The compounds were selected with computational approach and are intended for screening projects

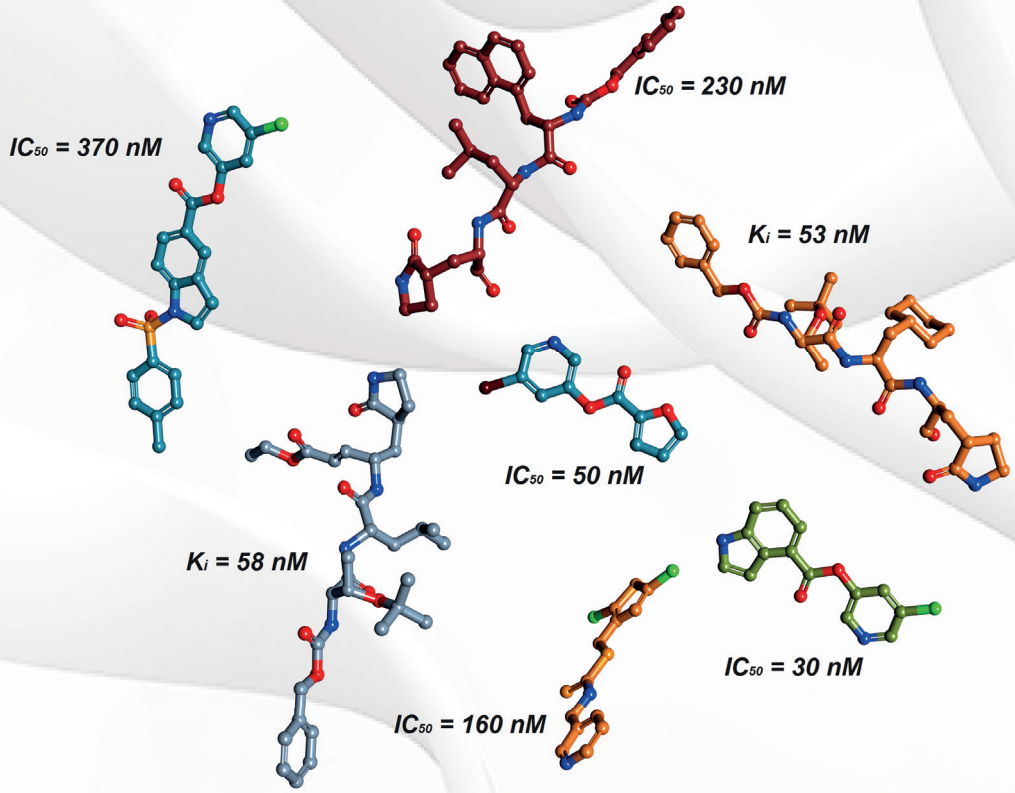
**there is additional fee for preparation of the solution

***4 ml amber glass vials or Deep-well plates: Matrix cat# 4247 (1.4 mL, Blank, Polypropylene, Round Bottom Tubes) w/CapMats. Or plates and vials provided by customer.

Design speciality:

The ML SARS Targeted Library has been carefully designed using machine learning (artificial neural networks and Bayesian statistics). (For more information see Appendix 1)

For machine learning 306 compounds with anti-SARS activity (ChEMBL Targets ID: CHEMBL3927, CHEMBL5118, CHEMBL612575) were randomly divided into two equal groups. Each group had the same amounts of active and inactive compounds (cutoffes were 10 and 15 μM for two different independent researches with models developing and virtual screenings) and was used as training and test set, then vice versa. The training sets were used for development of Bayesian and artificial neural networks models. Both methods were based on a number of different molecular descriptors – fingerprints (FCFP4, FCFP6, ECFP4 or ECFP6), molecular weight, number of hydrogen acceptors and donors, number of rings, number of rotatable bonds, LogP, PSA, topological descriptors and other. To construct each model algorithm automatically selected different fingerprints, since they carry structural information about the compounds, and also about 20 descriptors from 131 based on Student's correlation. The test sets were used for validation of all models.

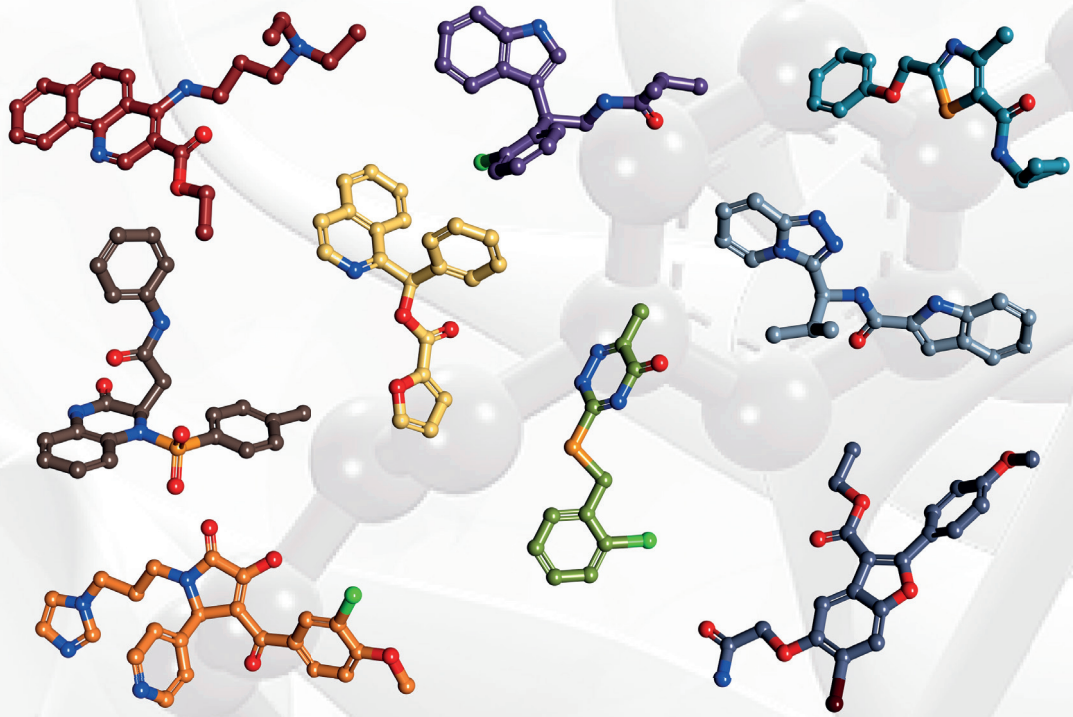


Example of compounds from training set for machine learning and their anti-SARS activities:

16 2-class neural network models, 16 4-class neural network models and 16 Bayesian models were built.

Based on ROC-score, MCC and other statistics, 21 best models for screening were selected - 6 2-class neural network models, 8 4-class neural network models and 7 Bayesian models.

Virtual screenings of Drug-like Green Collection toward best Bayesian and artificial neural networks models were performed. Top-scored compounds of machine learning methods and visually analyzed. The application of number different artificial neural networks and models should allow increasing the number of active compounds identified during screenings.

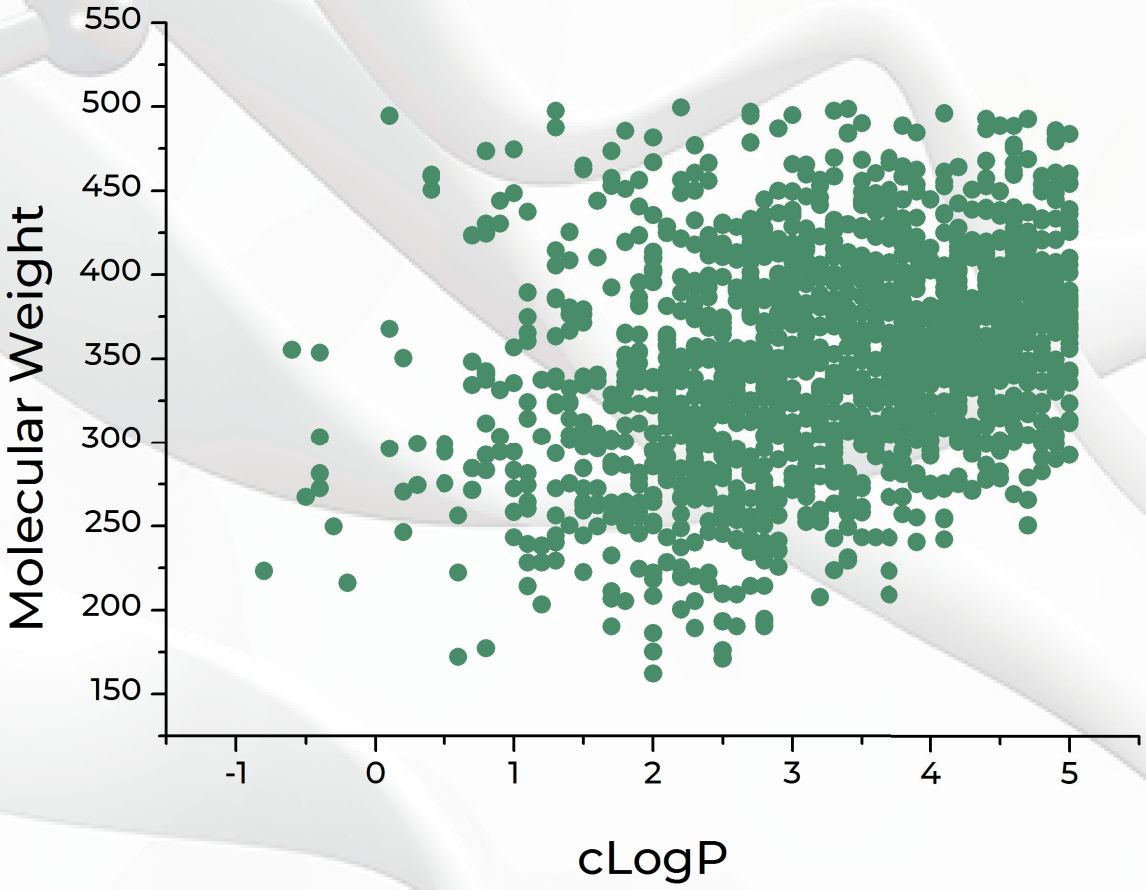


Example of compounds selected by machine learning

Distribution of physicochemical properties of compounds in the library:

100% Drug-like

56% Lead-like



The summary of the library characteristics:

	Minimum	Maximum	Average value
Molecular Weight	162,1	499,6	349
Number of Hydrogen Bond Donors	0	4	0,8
Number of Hydrogen Bond Aceptors	1	9	3,8
Number of Rotatable Bonds	0	10	3,8
CLogP	-0,8	5	3,3
Number of Rings	1	6	3,41
Polar Surface Area	17,1	169,3	75,7

Appendix 1

Machine Learning:

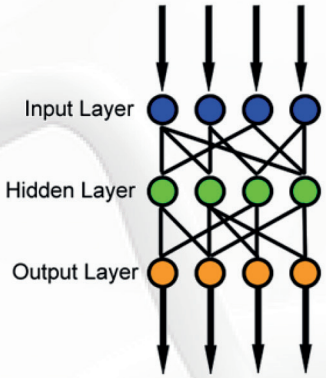
- Artificial Neural Networks (NNET)
- Bayesian modeling
- k-nearest neighbors algorithm (k-NN)

For the application of these methods are used training and test sets of compounds with known activity against molecular target of interest. The training sets are used for development and parameterization of Bayesian and artificial neural networks models (two-class models - active / inactive and 4-class - high / intermediate / weakly active / inactive). Both methods were based on different molecular descriptors - number of rings, number of hydrogen donors and acceptors, LogP, molecular weight, number of rotatable bonds, PSA, topological descriptors, fingerprints and other. Also the training sets are used as a template for compounds selection with k-NN based on different fingerprints. The test sets were used for validation of Bayesian models, neural networks and k-NN. Top-scored compounds obtained by application of different machine learning methods are crossed and visually analyzed. The combination of artificial neural networks, Bayesian statistics and k-NN should allow increasing the number of active compounds identified during screening.

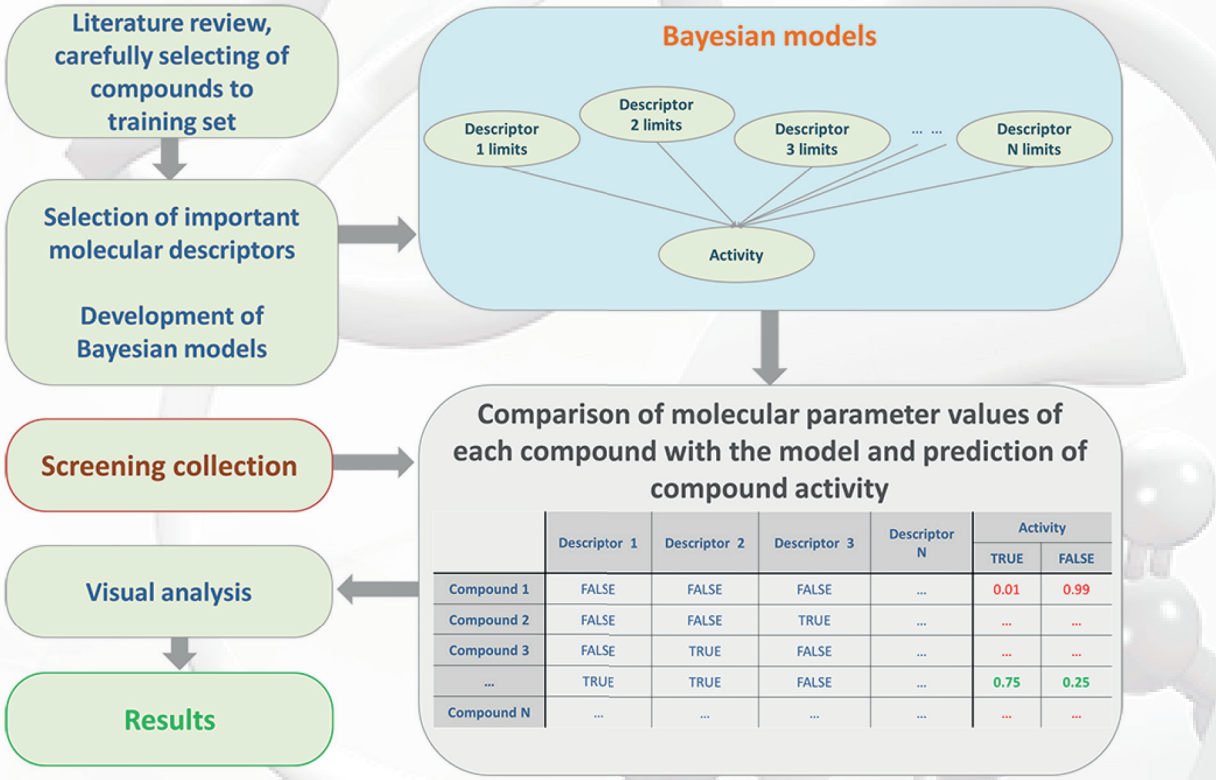
Feedforward Neural Network:

The **Feedforward Neural Network** with one hidden layer is the kind of neural network in which the signals propagate in one direction, starting from the input layer of the neurons, through the hidden layer to the output layer. The result of the processing of the signal is obtained on the output neurons. In networks of this kind there are no reciprocal links.

During training, the parameters of the neural network change, so that the given input to the network provides a suitable output. This learning process usually involves changing the weights and thresholds of the network variables.

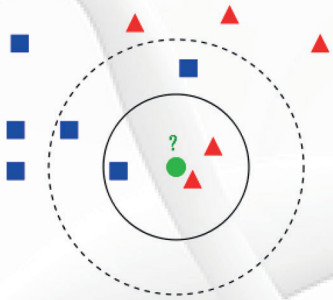


Bayesian modeling:



K-nearest neighbors algorithm (k-NN):

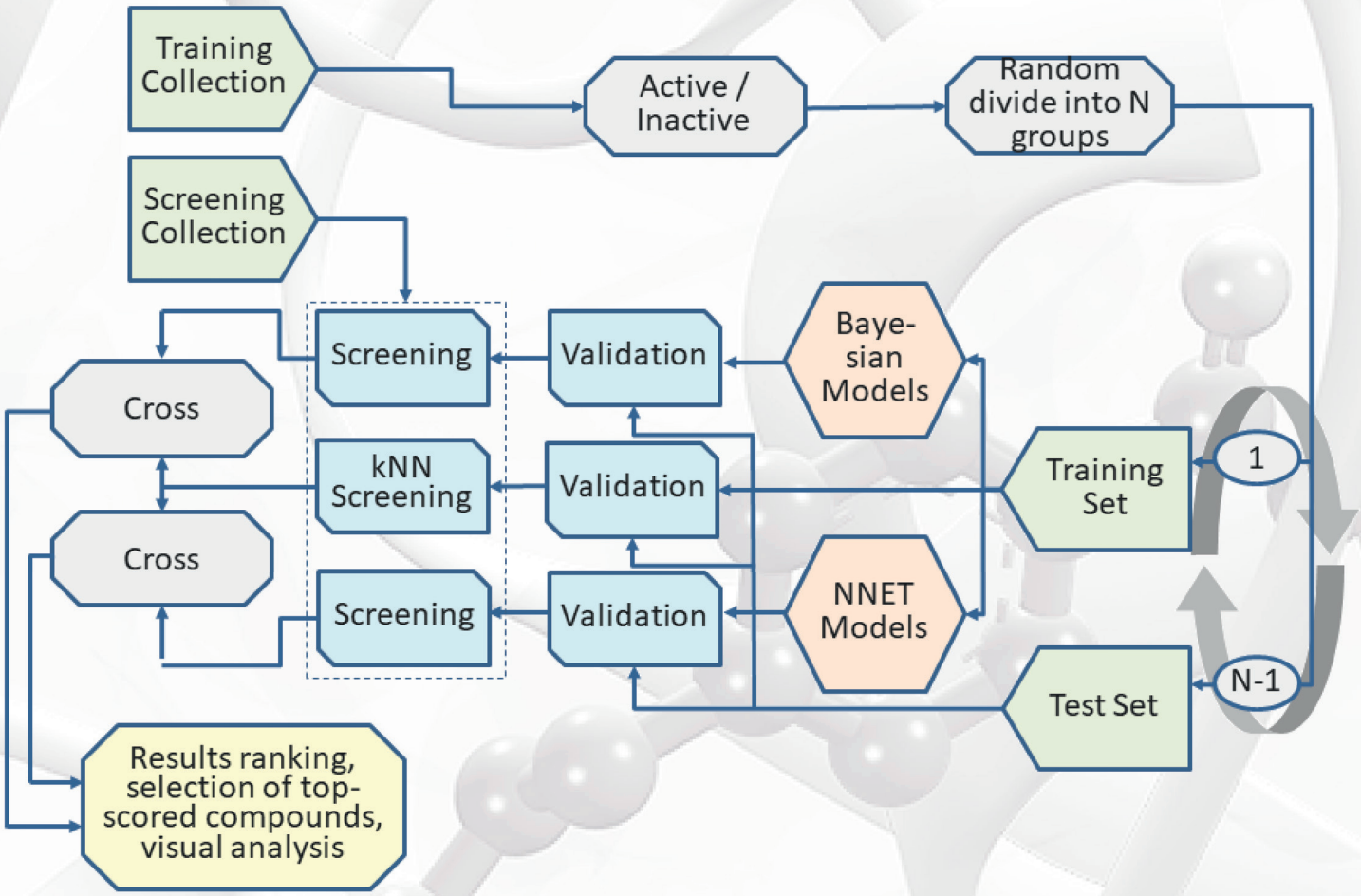
The k-nearest neighbors algorithm (k-NN) is a method used for classification. The input consists of the k closest training examples in the feature space. The output is a class membership. An object is classified by a majority vote of its neighbors, with the object being assigned to the class most common among its k nearest neighbors (k is a positive integer, typically small). If k = 1, then the object is simply assigned to the class of that single nearest neighbor.



Example of k-NN classification. The test sample (green circle) should be classified either to the first class of blue squares or to the second class of red triangles. If k = 3 (solid line circle) it is assigned to the second class because there are 2 triangles and only 1 square inside the inner circle. If k = 5 (dashed line circle) it is assigned to the first class (3 squares vs. 2 triangles inside the outer circle).

OTAVA chemicals

General scheme of combined machine learning application:



Validation of machine learning models:

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP) \times (TP + FN) \times (TN + FP) \times (TN + FN)}}$$

$$\text{Accuracy} = \frac{\text{TP} + TN}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \qquad \text{Recall} = \frac{TP}{TP + FN}$$

$$\text{Precision} = \frac{TP}{TP + FP} \qquad \text{F1} = 2\,\frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

$$\text{Kappa} = \frac{\text{Accuracy} + p_e}{p_e}$$

$$p_{false} = \frac{TN + FN}{TP + TN + FP + FN}\frac{TN + FP}{TP + TN + FP + FN} \qquad p_e = p_{true} + p_{false}$$

$$p_{true} = \frac{TP + FN}{TP + TN + FP + FN}\frac{TP + FP}{TP + TN + FP + FN}$$

If the model is multi-class, the average statistical values is calculated based on the following parameters for each class.

Our contacts:

OTAVA LTD.
400 Applewood Crescent, Unit 100
Vaughan, Ontario, L4K 0C3, CANADA
E-mail: north.america@otavachemicals.com
Tel.: +1-416-549-8030

OTAVAchemicals Europe Distributors
Meistrų g. 9
Vilniaus, 02189, LITHUANIA
E-mail: eurasia@otavachemicals.com
Tel.: +3-706-738-3544

web: https://otavachemicals.com/

Follow us:
http://fb.me/otavachemicals
http://linkedin.com/company/otavachemicals

OTAVA chemicals

Custom synthesis

Molecular modeling

Amyloids detection

Contract research

OTAVA Ltd.
65 Ellerslie Ave., Suite 560
Toronto, Ontario, M2N 1Y1
CANADA

OTAVACHEMICALS.COM