In silico target prediction for protein-protein small molecule inhibitors using CGBVS

○Enzo Kawasaki[1]
kawasaki@intage.com

Chisato Kanai[1]
kanai@intage.com

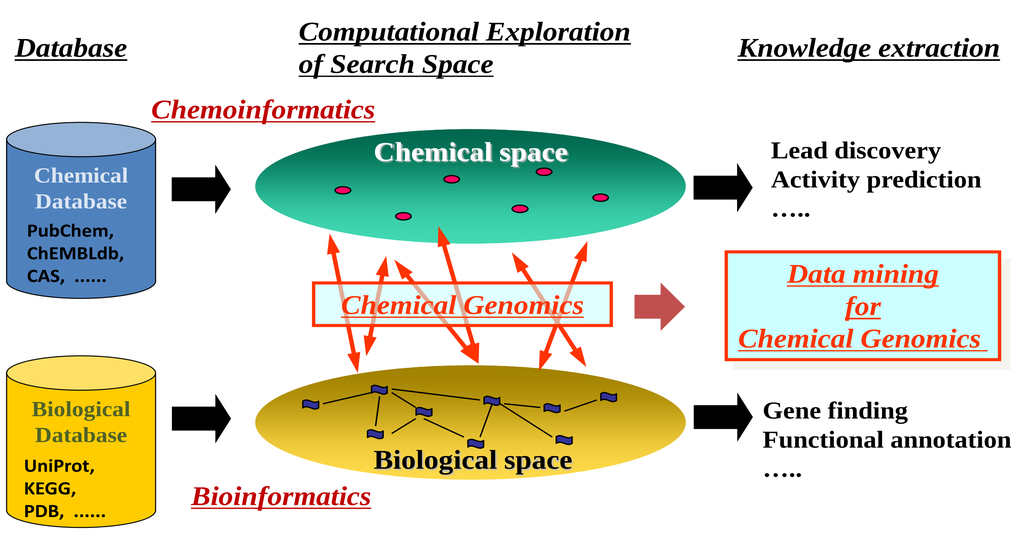
Atsushi Yoshimori[2]
yoshimori@itmol.com

[1] INTAGE Healthcare, Inc.
Kankoboko-cho, Shimogyo-ku, Kyoto, 600-8009.

[2] Institute for Theoretical Medicine, Inc.
26-1 Muraoka-Higashi 2-chome, Fujisawa, Kanagawa 251-0012.

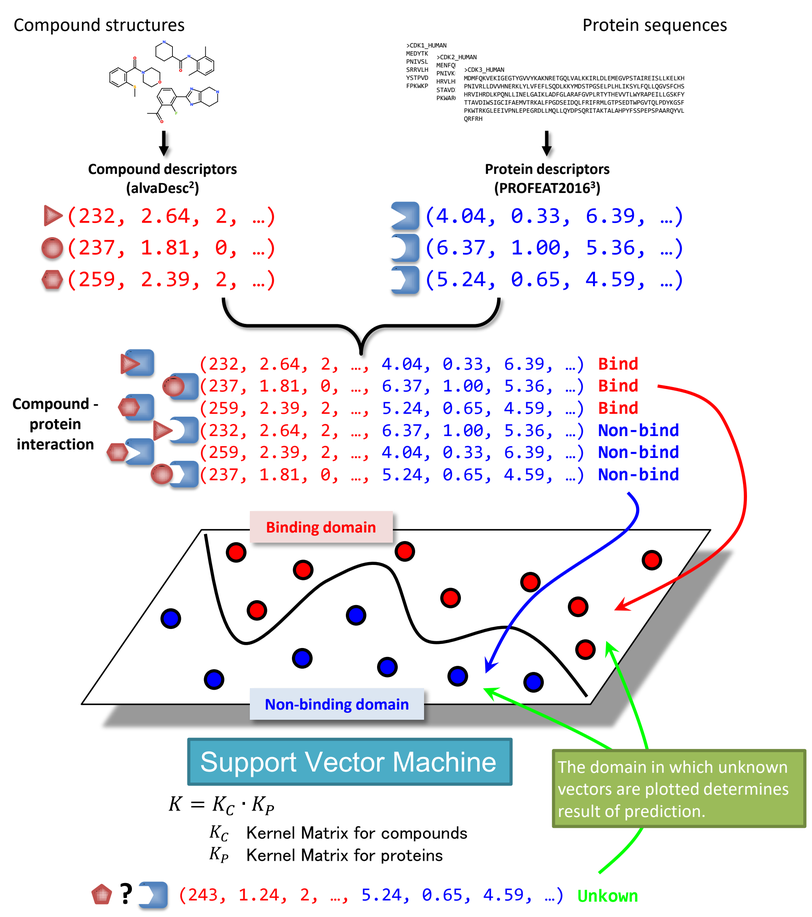
Chemical Genomics-Based Virtual Screening (CGBVS)

CGBVS[1] is a method for predicting the activity of a compound based on the binding pattern obtained from the interaction information (chemical genomics information) between the protein (biological space) and the compound (chemical space).



Huge amounts of chemogenomics information can be obtained from currently existing public databases.

[1] Analysis of multiple compound-protein interactions reveals novel bioactive molecules. *Mol. Syst. Biol. 7, 472, 2011*



[2] alvaDesc 1.0 (https://chm.kode-solutions.net/products_alvadesc.php)
[3] PROFEAT2016 (http://bidd2.nus.edu.sg/cgi-bin/profeat2016/main.cgi)

Protein-Protein Interaction Model for CGBVS

CGBVS Prediction Models:

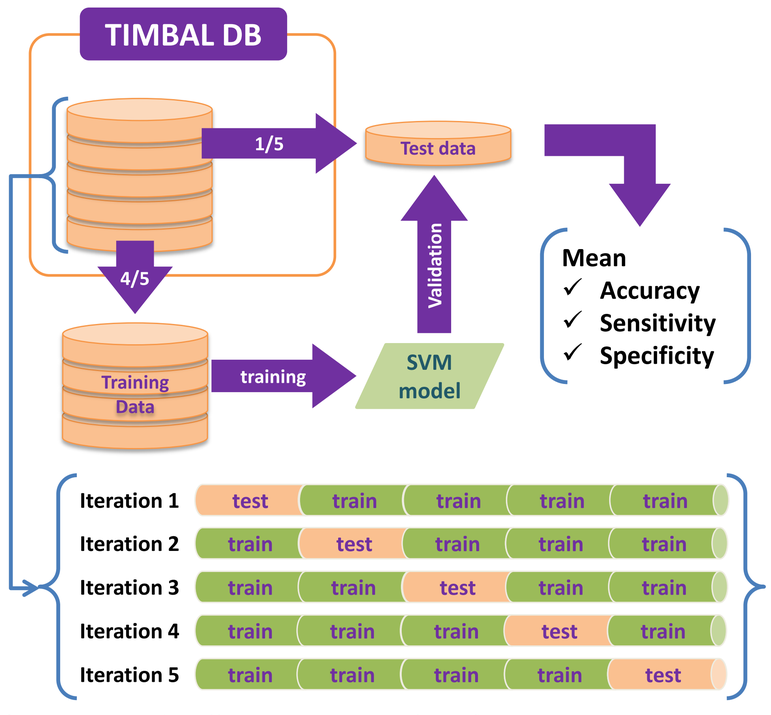
Currently, there are 6 available prediction models that can screen a total of 1,246 targets corresponding to the following protein groups: GPCR, Kinase, Ion channel, Nuclear receptor, Protease and Transporter. Protein-compound interaction data obtained from the ChEMBL25 database was used to train these prediction models.

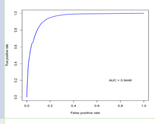
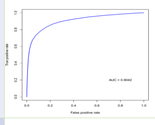
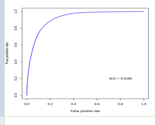
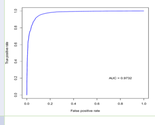
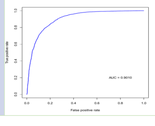
Table 2. Currently available CGBVS Prediction Models

Model	Target proteins	Training data (# of interactions)	Protein Family/Group
GPCR	237	145,632	Class Aα, Class Aβ, Class Aδ, Class Aγ, Class B, Class C
Kinase	415	135,896	AGC, CAMK, CMGC, STE, TK, TKL, etc.
Ion Channel	207	58,042	Voltage-gated, Ligand-gated, etc.
Transporter	123	43,654	Electrochemical, ATPase, ATP-binding cassette, etc.
Nuclear receptor	41	39753	NR1, NR2, NR3, NR4, NR5
Protease	223	117,346	Endopeptidase, Exopeptidase

Creation and Validation of the Protein-Protein Interaction (PPI) Model:

PPI data from TIMBAL DB[4] was used to create the PPI model. 5-fold cross-validation was performed to evaluate the PPI CGBVS model. Mean values of statistical analysis results are shown below.



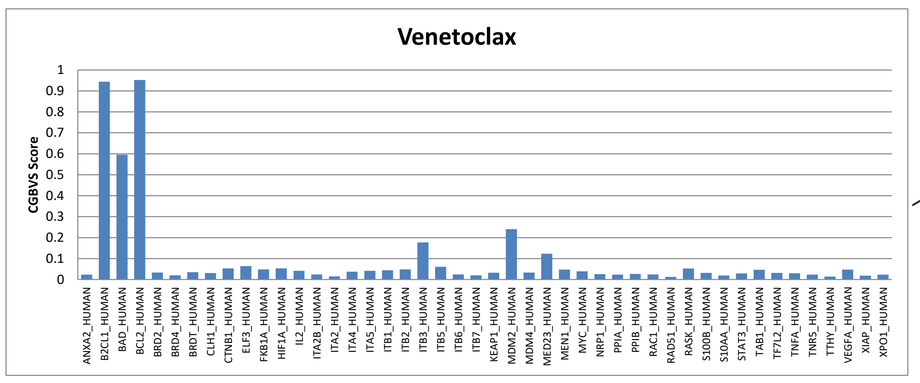
Model	Statistical Measures	ROC (AUC)
Ion Channel	Accuracy: 0.95, TPR: 0.96, TNR: 0.77, PPV: 0.98, NPV: 0.62	
Nuclear receptor	Accuracy: 0.83, TPR: 0.83, TNR: 0.84, PPV: 0.92, NPV: 0.68	
Protease	Accuracy: 0.91, TPR: 0.94, TNR: 0.72, PPV: 0.96, NPV: 0.66	
Transporter	Accuracy: 0.95, TPR: 0.96, TNR: 0.87, PPV: 0.98, NPV: 0.73	
PPI	Accuracy: 0.89, TPR: 0.94, TNR: 0.64, PPV: 0.93, NPV: 0.68	

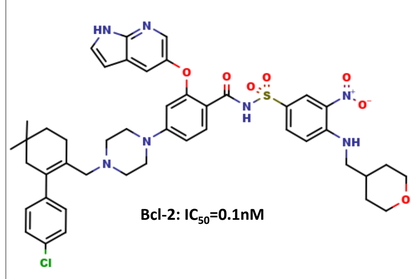
TPR: True Positive Rate, TNR: True Negatiive Rate
PPV: Positive Predictive Value, NPV: Negative Predictive Value

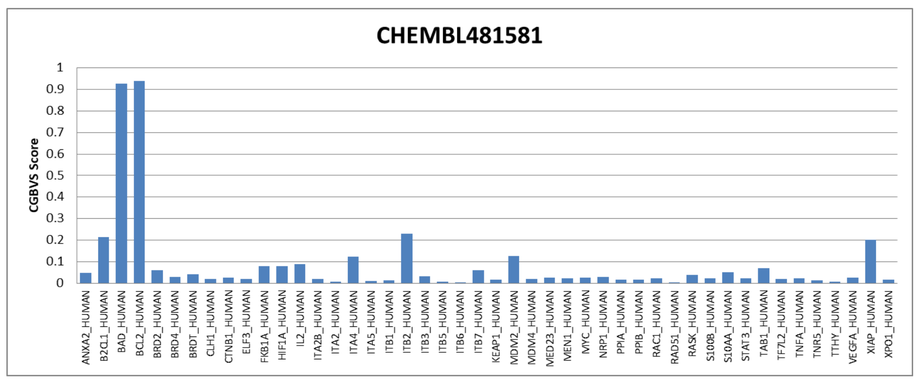
[4] TIMBAL database (http://mordred.bioc.cam.ac.uk/timbal/)

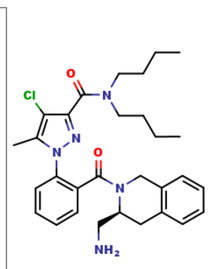
Examples of target prediction of PPI inhibitors

Screening of compounds existing in ChEMBL25 but not in TIMBAL DB was performed. Selected Bcl-xL/Bcl-2 antagonists are shown below.

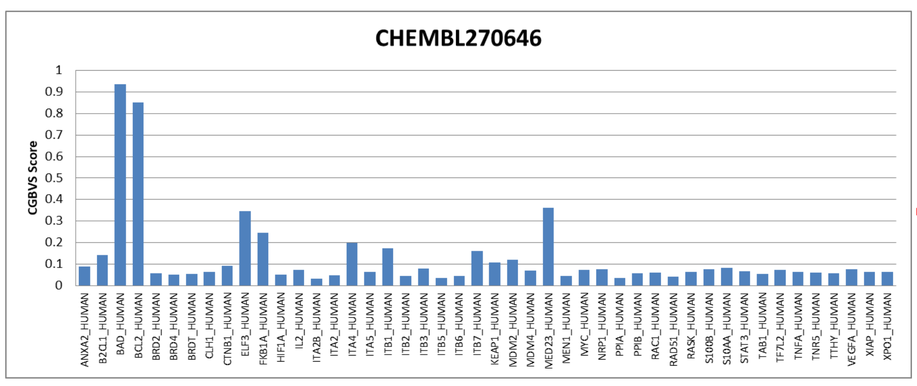


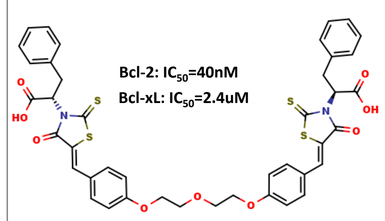


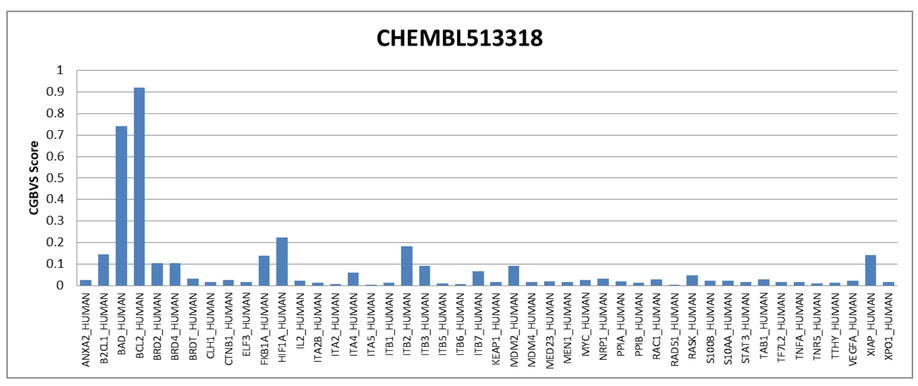


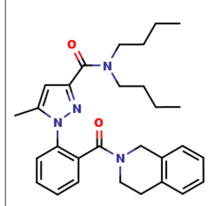


Bcl-2: IC_{50}=20nM
Bcl-xL: IC_{50}=250nM









Bcl-2: IC_{50}=230nM
Bcl-xL: IC_{50}=25.2uM