A STEP TOWARDS GPCRs POLYPHARMACOLOGY: AN ALGORITHM TO 5-HT$_7$/5-HT$_{1A}$ SELECTIVITY PREDICTION

Rafał Kurczab[1], Vittorio Canale[2], Paweł Zajdel[2], Andrzej J. Bojarski[1]

[1]Department of Medicinal Chemistry, Institute of Pharmacology Polish Academy of Sciences, 12 Smętna Street, 31-343 Kraków, Poland
[2]Department of Medicinal Chemistry, Jagiellonian University Medical College, 9 Medyczna Street, 30-688 Kraków, Poland

Methodology

- **Sets**
  - source: ChEMBL,
  - compounds with activity data to both 5-HT$_{1A}$R and 5-HT$_7$R,

ChEMBL

	Set Name	Description/Conditions used to generate the set	Amount
●	Selective	Selective ligands to 5-HT$_7$R: K_i(5-HT$_7$) < 100 nM & K_i(5-HT$_{1A}$)/K_i(5-HT$_7$) > 5	69
■	DUD	Decoy set generated for selective set using e-DUD service	5198
▲	Revsel	Reverse selective ligands (ligands selective to 5-HT$_{1A}$R): K_i(5-HT$_{1A}$) < 100 nM & K_i(5-HT$_7$)/K_i(5-HT$_{1A}$) > 5	124
★	Notsel	A set of ligands that do not meet the condition of selectivity: 0.20 < K_i(5-HT$_{1A}$)/K_i(5-HT$_7$) < 5	440
◆	Nselbact	The set of active ligands to both receptors simultaneously: K_i(5-HT$_7$) & K_i(5-HT$_{1A}$) < 100 nM	89

ML set preparation

The strategy of testing and training sets generation

Set ID	Class {+1}	Class {-1}
Dud	●	■
Revsel	●	▲
Notsel	●	★
Nselbact	●	◆

- **ML models**
  - two approaches (*structure-based* and *ligand-based*),
  - optimization of SVM RBF kernel parameters,
  - the 10 random training and testing sets were generated
  - diffrent size of training and testig sets,

structure-based approach

Sets

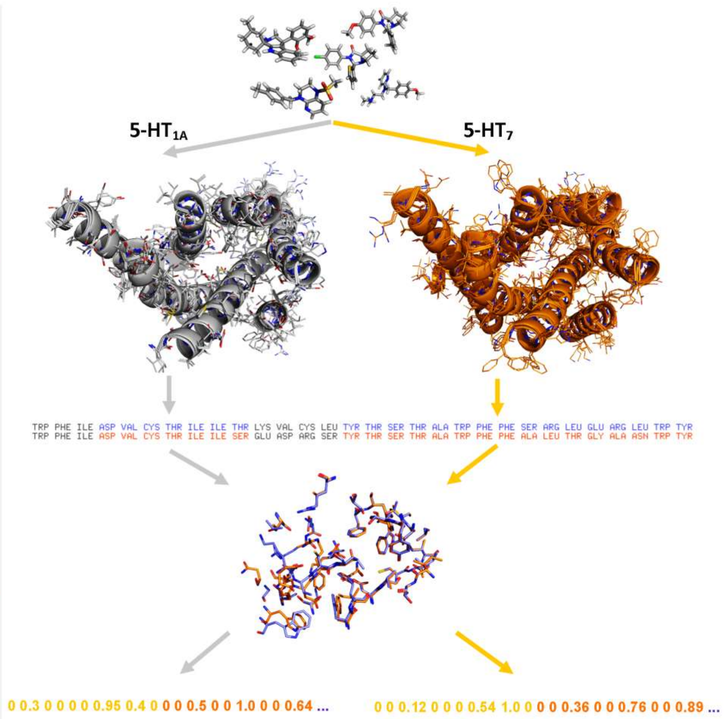
Docking (Glide SP)
- two templates: 5-HT$_{1B}$ and D$_3$ (with and without loops),
- three conformations per template,

Interaction analysis
- SIFt statistics

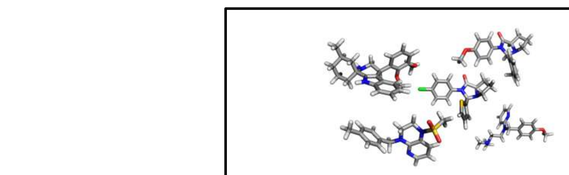
AA base setup
- selection the common set of amino acids,

SIFt profiles

SIFts Concatenation



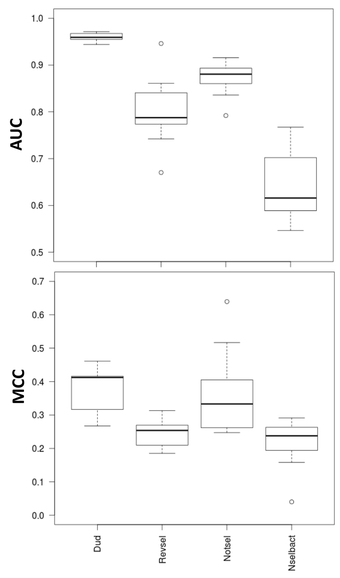
ligand-based approach



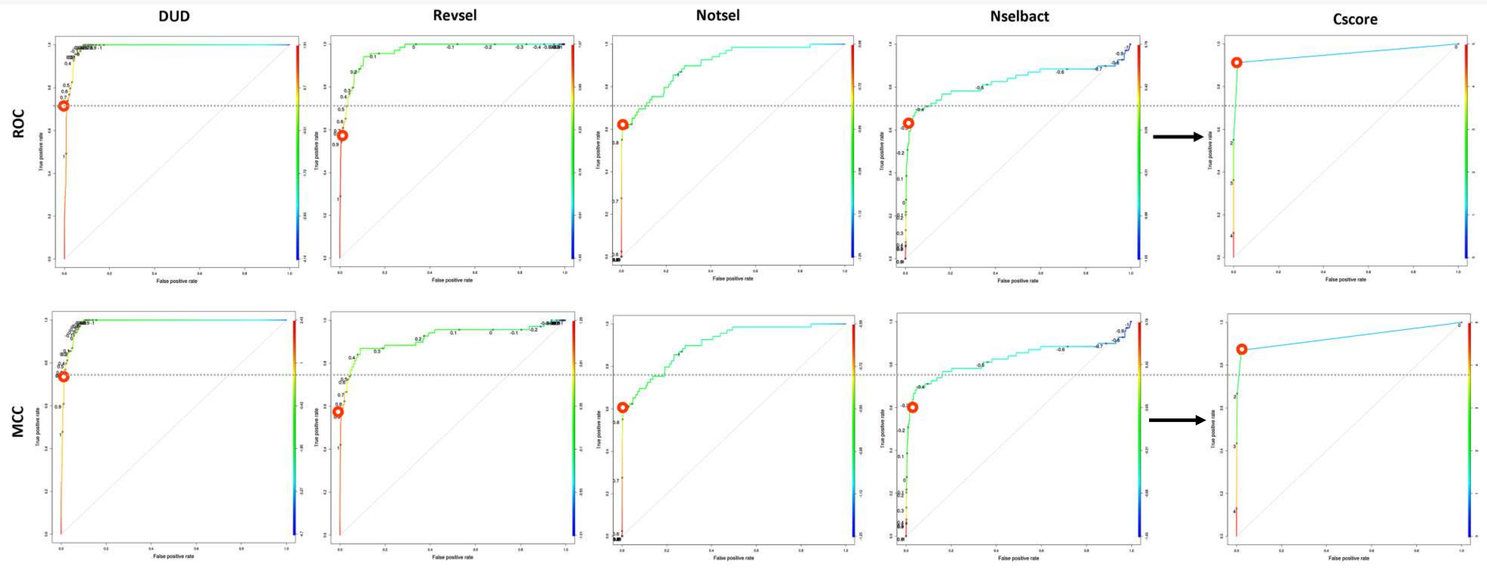
Molecular fingerprints

- MACCSFP 166 bits
- CDKFP 1024 bits
- Klekota-Roth FP 4068 bits

- **Cscore**
  - two optimization approaches (AUC and MCC),
  - only best models survived,
  - cutoff selection based on MCC maximization,
  - data fusion method (SUM rule) used to constuct final Cscore model,



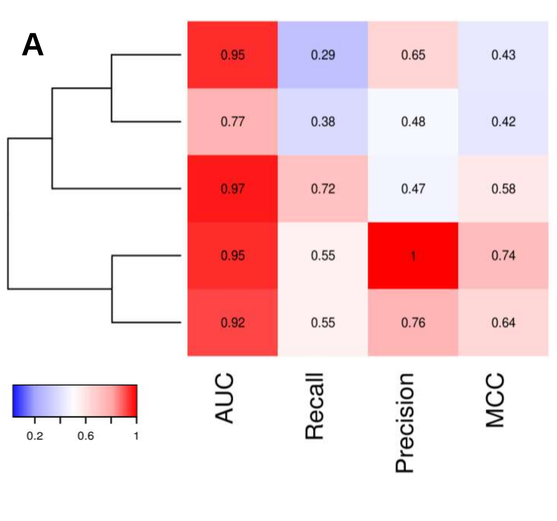
Selection of the best models



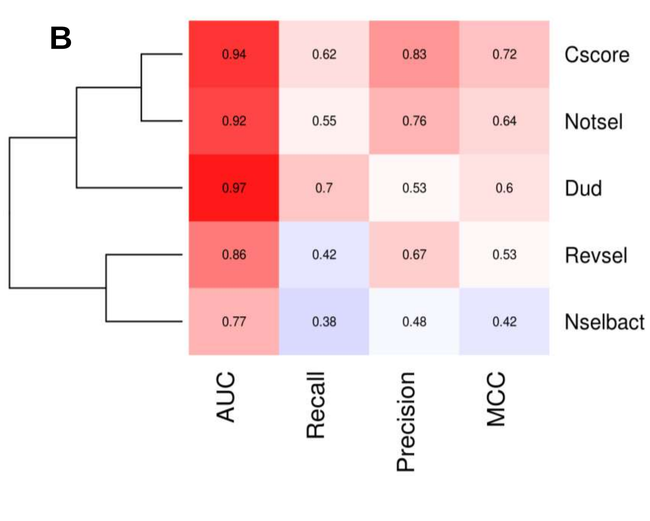
ROC curves of the best models selected to construct Cscore models.

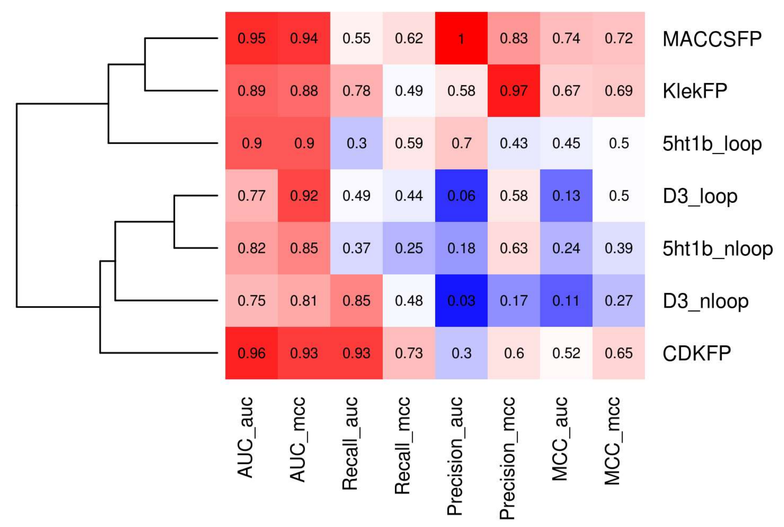
Results

- Cscore models performed globlally the best,
- ligand-based approach outperformed structure-based,
- larger training set do not significantly influenced on the models performance,



An example of performance for all models generated based on MACCSFP (*ligand-based* approach) with different distribution coefficients (A - 0.40, B – 0.60) of initial set of ligands.





The heat map comparing the performance of CScore models obtained for all studied cases (i.e. three molecular fingerprints and SIFt generated on four homology models).

Acknowledgments

The study was partially supported by the Polish-Norwegian Research Programme operated by the National Centre for Research and Development under the Norwegian Financial Mechanism 2009-2014 in the frame of Project PLATFORMex (Pol-Nor/198887/73/2013) and by the National Science Center Grant No DEC-2012/05/B/N27/03076.

NATIONAL SCIENCE CENTRE POLAND

eea grants · norway grants · Narodowe Centrum Badań i Rozwoju · Polish-Norwegian Research Programme