

AmlActive (AIA): A Large-scale QSAR Based Target Fishing and Polypharmacology Predictive Web Tool

Luis Felipe de Moraes Melo

Federal University of Paraiba, Brazil

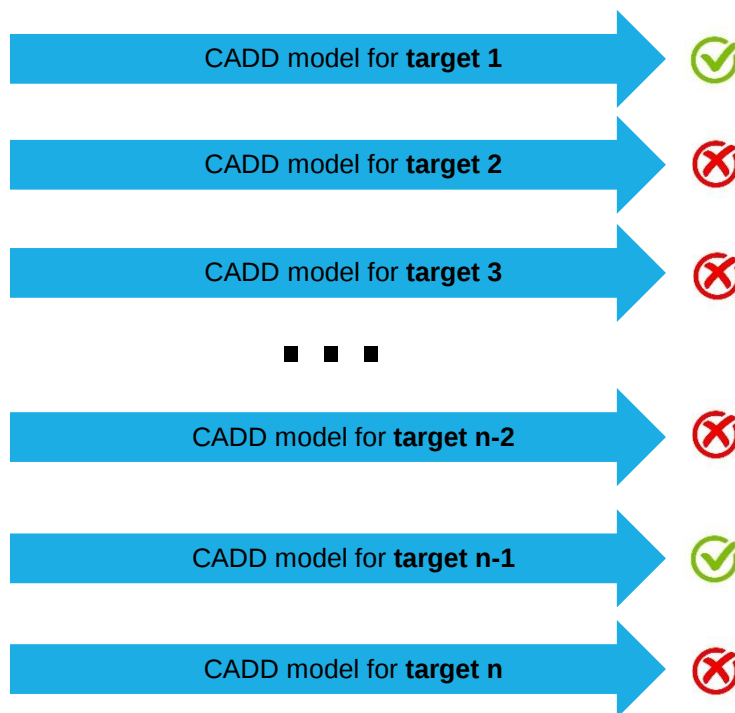
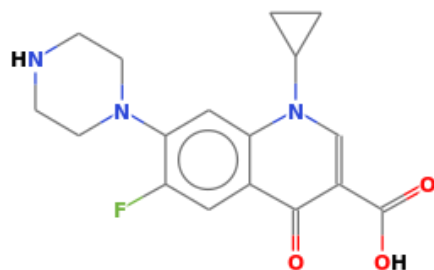


CHEMI**N**FORMATICS
LAB



Introduction

- Reverse virtual screening (target fishing)





Introduction

- Web tools for reverse virtual screening

PASS



SwissTargetPrediction
SwissDrugDesign

PLATO r35



Introduction

- Web tools for reverse virtual screening

PASS



SwissTargetPrediction
SwissDrugDesign

PLATO r35

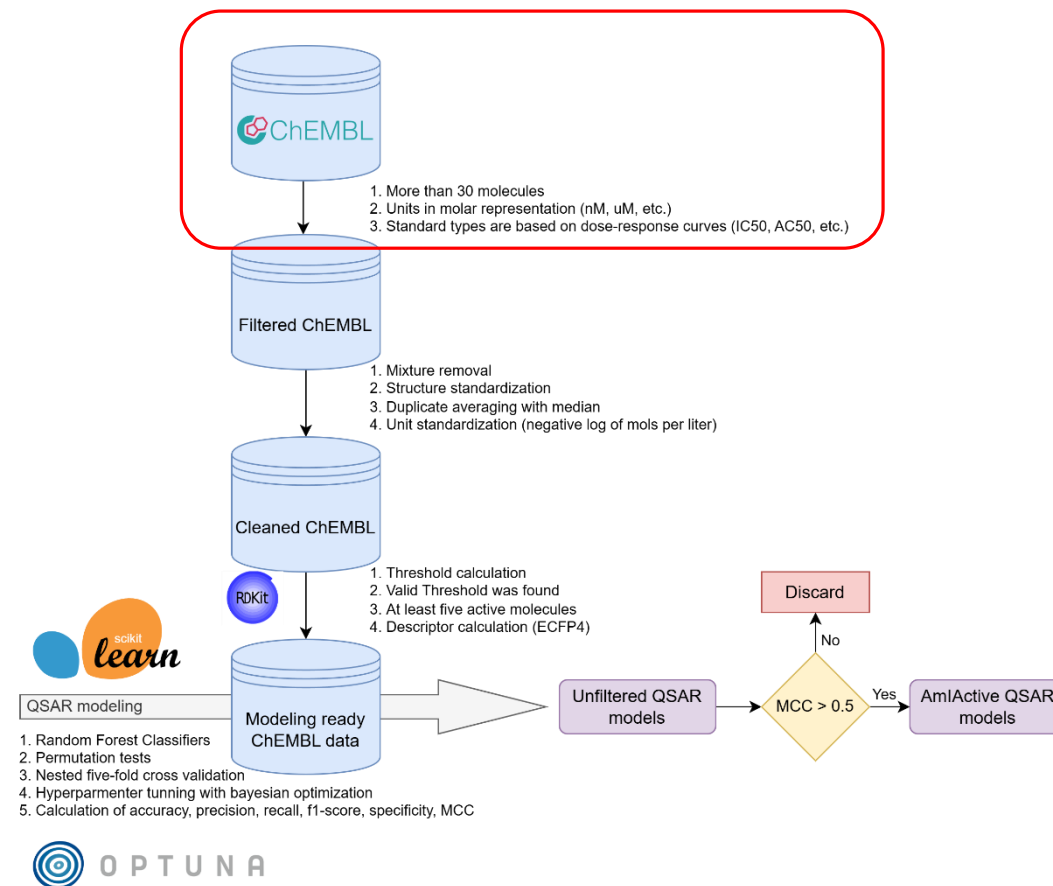


AmlActive

Materials and methods



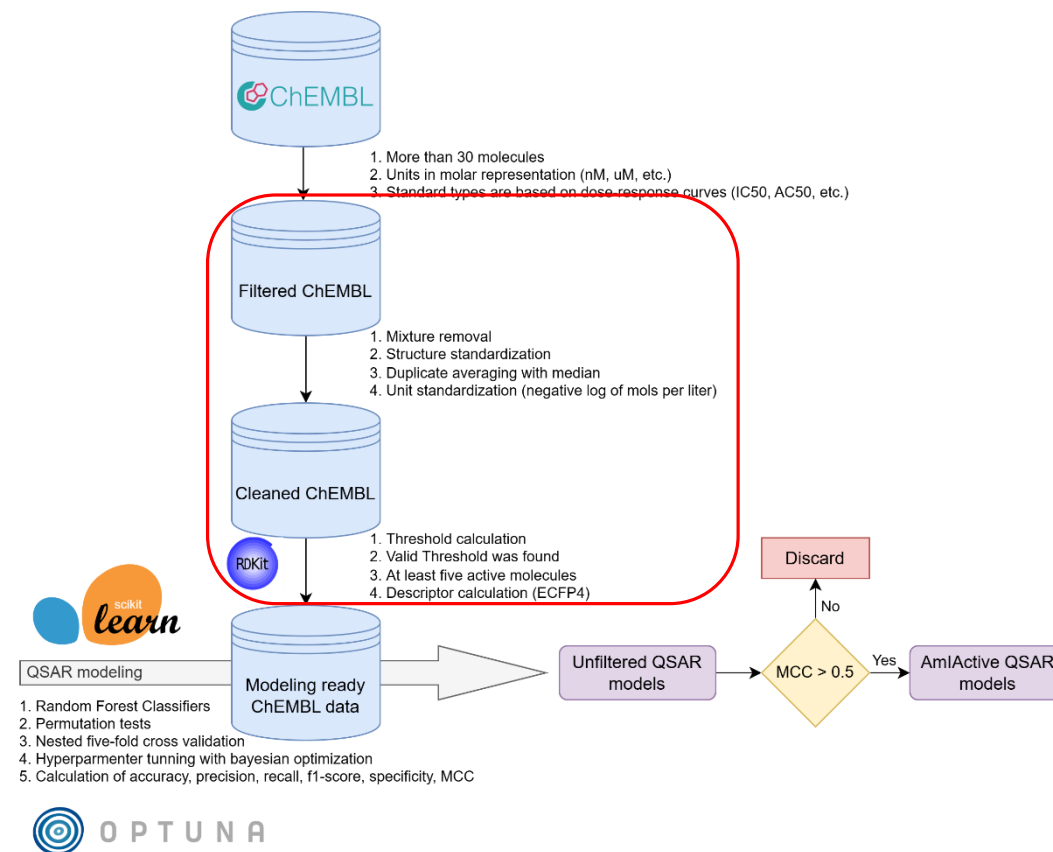
- Data was extracted from ChEMBL v30
- Targets where either:
 - Single proteins
 - Protein Families
 - Protein Complexes
 - Organisms
 - Cell-lines
 - Tissues
- Assay units where in some form of molar representation
- Assay metric was based on dose-response curves
- All filters considered, the final dataset had at least 30 molecules





Materials and methods

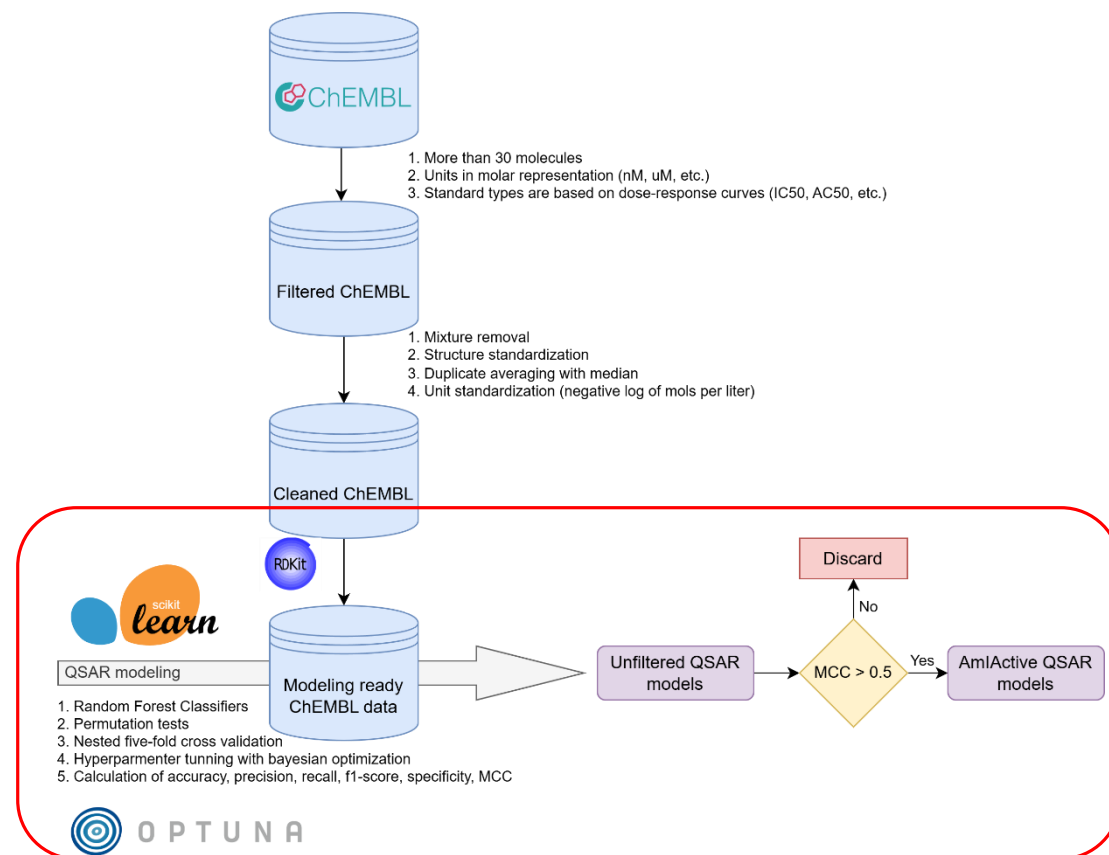
- Mixtures were removed
- Structures were standardized
- Duplicate activities were averaged (median)
- Activity values were converted to a negative logarithm of moles per liter form
- Classifier activity thresholds were established
- Extended connectivity fingerprints (ECFP4)





Materials and methods

- Random forest algorithm
- Hyperparameter tuning with bayesian optimization and nested five-fold cross-validation
- Calculated classifier metrics:
 - Accuracy
 - Precision
 - Recall
 - F1-Score
 - Specificity
 - Matthews Correlation Coefficient (MCC)
- Models that had a $MCC > 0.5$ were selected as the models of the system





Results and Discussion

- After all the filtering and cleaning procedures, we obtained **3239** QSAR models.
- Of these 3239 models, **2277** corresponded to distinct targets of interest.
- The target types distribution was:
 - Single proteins ($\approx 69\%$)
 - Cell lines ($\approx 15\%$)
 - Organisms ($\approx 6\%$)
 - Protein complexes ($\approx 4\%$)
 - Protein Families ($\approx 4\%$)
 - Tissues ($\approx 0.6\%$)



Results and Discussion

	Threshold	Accuracy	F1-score	MCC	Precision	Recall	Specificity
Cell-line	5.40	0.82	0.81	0.65	0.84	0.81	0.83
Organism	5.47	0.82	0.81	0.65	0.83	0.81	0.82
Protein Complex	6.64	0.82	0.82	0.65	0.82	0.83	0.80
Protein Family	6.25	0.82	0.82	0.65	0.83	0.83	0.81
Single Protein	6.43	0.82	0.82	0.65	0.83	0.83	0.81
Tissue	5.49	0.82	0.82	0.65	0.83	0.82	0.82

Results and Discussion



<https://amiactive.ccen.ufpb.br/>



AmlActive is a target fishing web tool that allows the scientific community to screen thousands of biological activities by clicking a single button.

Use the tool



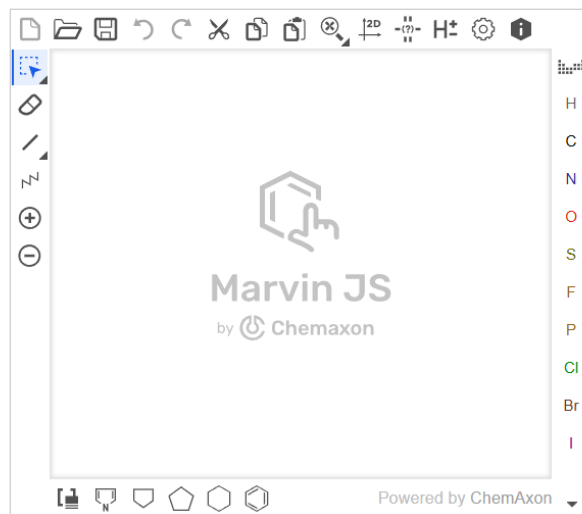
Results and Discussion



AmlActive

Help

Models info



Generate Smiles >>

Predict

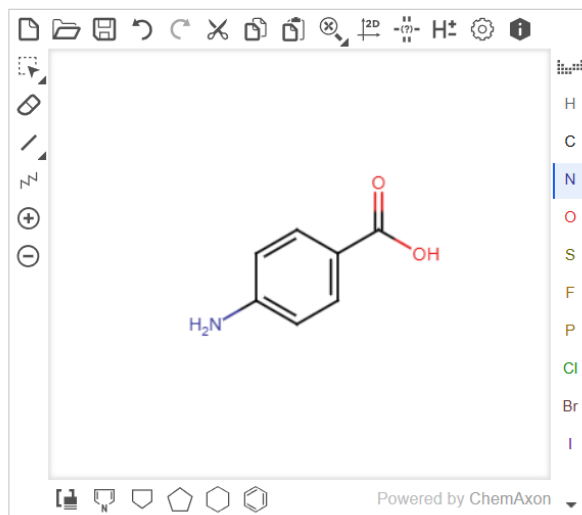
Results and Discussion



AmlActive

Help

Models info



NC1=CC=C(C=C1)C(O)=O

Generate Smiles >>

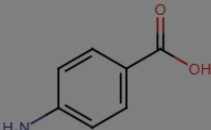
Predict

Results and Discussion



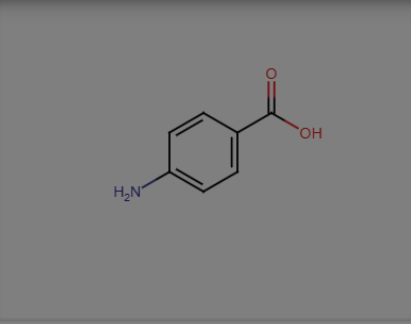
 **AmlActive** Help Models info

Processing Queue
Your predictions are currently being processed...
Abort Prediction Export to CSV New Prediction

NC(=O)c1ccc(N)cc1

H
C
N
O
S
F
P
Cl
Br
I

Generate Smiles >>



Powered by ChemAxon

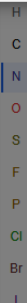
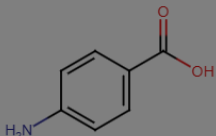
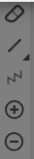
Predict

Results and Discussion



 **AmlActive** Help Models info

Processing Queue
Your predictions are finished, press the download button to get the results or start a new prediction
Abort Prediction Export to CSV New Prediction



Generate Smiles >>

Powered by ChemAxon

Predict



Conclusion

- AmlActive (AIA) is a brazilian QSAR based reverse screening webtool
- The AIA system has 3,239 models which correspond to 2,277 distinct targets
- The AIA system targets are either: Single proteins, Protein families, Protein complexes, organisms, cell-lines or tissues
- The AIA system can be accessed at: <https://amiactive.ccen.ufpb.br/>

Thank you!

Luis Felipe de Morais Melo
luisfelipe.melo@hotmail.com



CHEMI FORMATICS
LAB