EDSP Prioritization: Collaborative Estrogen Receptor Activity Prediction Project (CERAPP)

Kamel Mansouri[1,2], Jayaram Kancherla[1,2], Ann Richard[2] and Richard Judson[2]
[1]Oak Ridge Institute for Science and Education, Oak Ridge, TN, USA
[2]National Center for Computational Toxicology, Office of Research and Development, U.S. Environmental Protection Agency, RTP, NC, USA

Kamel Mansouri | mansouri.kamel@epa.gov | 919-541-0545

www.epa.gov

Abstract

Humans are potentially exposed to tens of thousands of man-made chemicals in the environment. It is well known that some environmental chemicals mimic natural hormones and thus have the potential to be endocrine disruptors. Most of these environmental chemicals have never been tested for their ability to disrupt the endocrine system, in particular, their ability to interact with the estrogen receptor. EPA needs tools to prioritize thousands of chemicals, for instance in the Endocrine Disruptor Screening Program (EDSP). This project was intended to be a demonstration of the use of predictive computational models on HTS data including ToxCast and Tox21 assays to prioritize a large chemical universe of 32464 unique structures for one specific molecular target – the estrogen receptor. CERAPP combined multiple computational models for prediction of estrogen receptor activity, and used the predicted results to build a unique consensus model. Models were developed in collaboration between 17 groups in the U.S. and Europe and applied to predict the common set of chemicals. Structure-based techniques such as docking and several QSAR modeling approaches were employed, mostly using a common training set of 1677 compounds provided by U.S. EPA, to build a total of 42 classification models and 8 regression models for binding, agonist and antagonist activity. All predictions were evaluated on ToxCast data and on an external validation set collected from the literature. In order to overcome the limitations of single models, a consensus was built weighting models based on their prediction accuracy scores (including sensitivity and specificity against training and external sets). Individual model scores ranged from 0.69 to 0.85, showing high prediction reliabilities. The final consensus predicted 4001 chemicals as actives to be considered as high priority for further testing and 6742 as suspicious chemicals.

Project planning

Project goal:

The aim of CERAPP was to combine multiple computational models for prediction of estrogen receptor activity into a unique consensus model in order to prioritize a large chemical of 32464 structures.

Steps	Tasks
1: Structures curation	- Extract chemical structures from regulatory sources - Design and document a workflow for structure cleaning - Deliver the QSAR-ready training set and prediction set to all participants
2: Experimental data preparation	- Collect and clean experimental data for the evaluation set - Define the strategy used to evaluate models predictions
3: Modeling & predictions	- Trainor refine the models using the training set - Compiling of predictions and applicability domains for evaluation
4: Model evaluation	- Analyze the training and evaluation sets for consistency - Evaluate the predictions of each model separately
5: Consensus strategy	- Define and calculate scores for each model based on the evaluation step - Define a weighting scheme from the scores
6: Consensus modeling & validation	- Combine the individual predictions based on the weighting scheme and create a consensus prediction - Validate the consensus model using a new external dataset

Participants:

DTU/food: Technical University of Denmark/ National Food Institute

EPA/NCCT: U.S. Environmental Protection Agency / National Center for Computational Toxicology

FDA/NCTR/DBB: U.S. Food and Drug Administration/ National Center for Toxicological Research/Division of Bioinformatics and Biostatistics

FDA/NCTR/DSB: U.S. Food and Drug Administration/ National Center for Toxicological Research/Division of Systems Biology

Helmholtz/ISB: Helmholtz Zentrum Muenchen/ Institute of Structural Biology

ILS&EPA/NCCT: ILS Inc & EPA/NCCT

IRCSS: Istituto di Ricerche Farmacologiche "Mario Negri"

JRC_Ispra: Joint Research Centre of the European Commission, Ispra.

LockheedMartin&EPA: Lockheed Martin IS&GS/ High Performance Computing

NIH/NCATS: National Institutes of Health/ National Center for Advancing Translational Sciences

NIH/NCI: National Institutes of Health/ National Cancer Institute

RIFM: Research Institute for Fragrance Materials, Inc

UMEA/Chemistry: University of UMEA/ Chemistry department

UNC/MML: University of North Carolina/ Laboratory for Molecular Modeling

UniBA/Pharma: University of Bari/ Department of Pharmacy

UNIMIB/Michem: University of Milano-Bicocca/ Milano Chemometrics and QSAR Research Group

UNISTRA/Infochim: University of Strasbourg/ ChemoInformatique

U.S. Environmental Protection Agency
Office of Research and Development

Data preparation

Sources of chemicals:

- EDSP Universe (10K)
- Chemicals with known use (40K) (CPCat & ACToR)
- Canadian Domestic Substances List (DSL) (23K)
- EPA DSSTox – structures of EPA/FDA interest (15K)
- ToxCast and Tox21 (In vitro ER data) (8K)

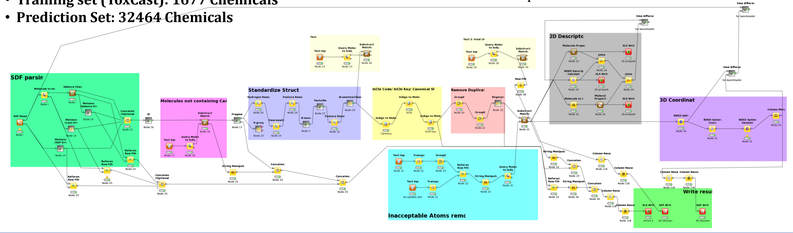
Sets of unique chemical structures:

- **Training set (ToxCast): 1677 Chemicals**
- **Prediction Set: 32464 Chemicals**

Chemical structure curation:

The main steps of this KNIME workflow were:

- Check the validity of the molecular file format
- Retrieve any missing structures from web-services
- Remove the inorganic and metallo-organic structures
- Remove salts and counter ions and fulfill valence
- Standardize stereo-isomers and tautomers
- Remove duplicates

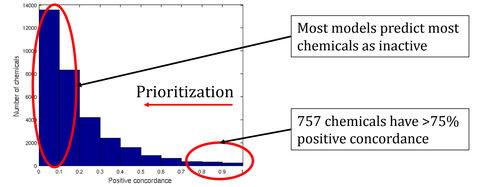


Models and evaluation

Categorical models:
- Binding: **22 models**
- Agonists: **11 models**
- Antagonists: **9 models**

Continuous models:
- Binding: **3 models**
- Agonists: **3 models**
- Antagonists: **2 models**

Concordance of binding models on the active compounds of the prediction set.



Evaluation set for binary categorical models

	Active	Inactive	Total
Binding	1982	5301	7283
Agonist	350	5969	6319
Antagonist	284	6255	6539
Total	2617	7024	7522

Evaluation set for continuous models

	Inactive	V. Weak	Weak	Moderate	Strong	Total
Binding	5042	685	894	72	77	6770
Agonist	5892	19	179	31	42	6163
Antagonist	6221	76	188	10	10	6505
Total	6892	702	9016	81	93	7253

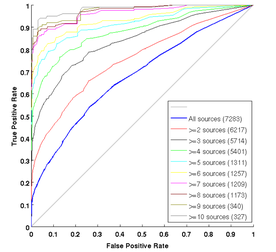
$$score\_1 = {}^1/_3\left(\frac{NER_{ToxCast} * N\_pred_{ToxCast}}{N\_tot_{ToxCast}} + \frac{N\_pred}{N\_tot} + {}^1/_{N_{filter}} \sum_{i=1}^{N_{filter}} \frac{NER_i * N\_pred_i}{N\_tot_i}\right)$$

$$score\_2 = {}^1/_2\left(NER_{ToxCast} + NER_{all\_filters}\right)$$

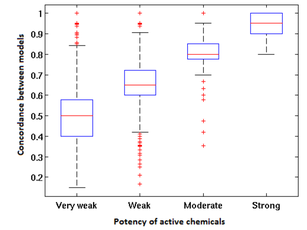
	DTU	EPA_NCCT	FDA_NCTR_DBB	FDA_NCTR_DSB	OCHEM	ILS_EPA	IRCCS_CART	IRCCS_Ruleset	IRC_Ispra	Lockheed&Martin_EPA	NIH_NCATS	NIH_NCI_GUSAR	NIH_NCI_PASS	RIFM	UMEA	UNC_MML	UNIBA	UNIMIB_Michem	UNISTRA
BA	0.78	0.69	0.68	0.66	0.72	0.75	0.75	0.62	0.67	0.66	0.65	0.69	0.66	0.65	0.70	0.65	0.73	0.71	0.60
% pred	49.48	100	100	0.62	96.47	99.9	89.20	94.88	100	100	99.97	94.87	97.4	100	99.89	100	46.75	36.44	100
Score_1	0.43	0.82	0.87	-	0.83	0.82	0.78	0.75	0.77	0.75	0.77	0.88	0.78	0.78	0.82	0.80	0.40	0.32	0.80
Score_2	0.80	0.78	0.84	0.69	0.80	0.79	0.77	0.77	0.74	0.75	0.67	0.84	0.76	0.69	0.76	0.73	0.80	0.85	0.73

BA: balanced accuracy inside the applicability domain for the unambiguous compounds. % pred: percentage of the predicted within 32464 chemicals.

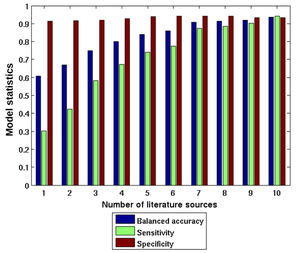
Consensus model



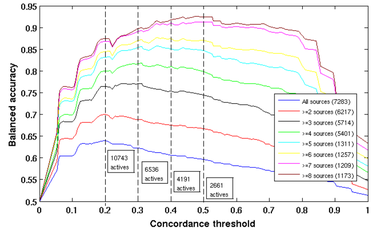
ROC curve of the consensus predictions evaluated using variable number of literature sources



Box plot of the correlation between the positive concordance of the categorical models and the potency level predicted by the continuous models, positive classes predicted by the consensus model



Plot of the consensus accuracy showing the importance of using multiple literature sources



Variation of the balanced accuracy with positive concordance thresholds at different numbers of literature sources and corresponding number of actives

Statistics of the consensus model

Observed\Predicted	ToxCast data: Actives	ToxCast data: Inactives	Literature data (All: 7283): Actives	Literature data (All: 7283): Inactives
Actives	83	6	597	1385
Inactives	40	1400	463	4838

	ToxCast data	Literature data (All: 7283)	Literature data (>6 sources: 1209)
Sensitivity	0.93	0.30	0.87
Specificity	0.97	0.91	0.94
Balanced accuracy	0.95	0.61	0.91

Conclusions

- Successful collaboration of 17 international research groups resulted in consensus predictions.
- Most individual models performed very well on both ToxCast and literature data.
- Model accuracy differences between ToxCast and literature may be due to noise in the literature data.
- A total number of 4001 out of 32464 chemicals are predicted by the consensus as active binders. For prioritization purposes, 6742 additional chemicals (down to 0.2 positive concordance) could be considered as suspicious.
- The consensus model predictions correlate better with literature data from multiple sources, likely due to greater noise in data with unique or low number of sources.
- Consensus prediction results are being used in the EDSP program. See http://actor.epa.gov/edsp21.