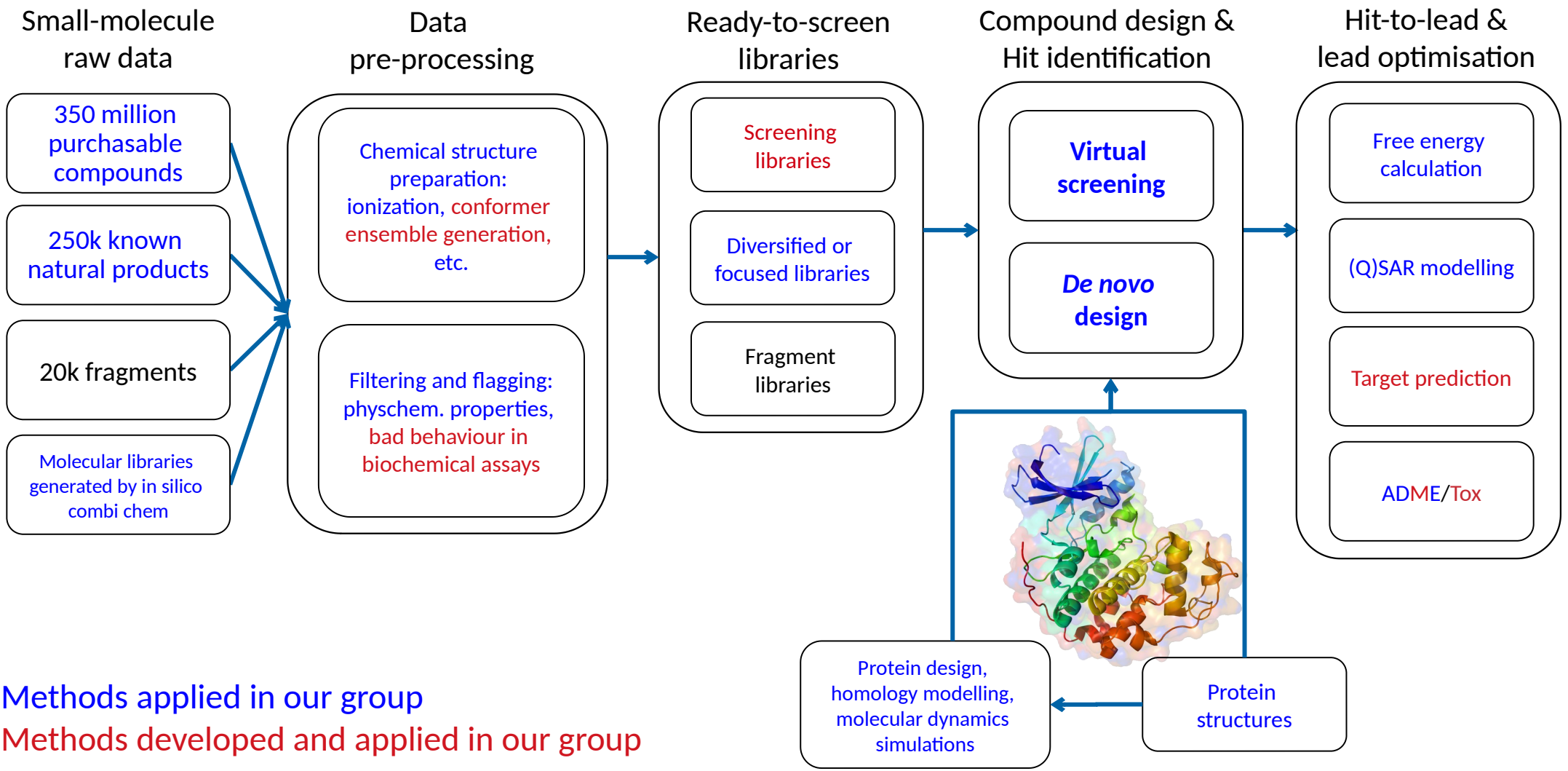
universität wien

Computational prediction of xenobiotic metabolism

Johannes Kirchmair

01/26/2021

We develop and apply a wide range of computational methods that can provide guidance to early drug discovery



Methods applied in our group

Methods developed and applied in our group

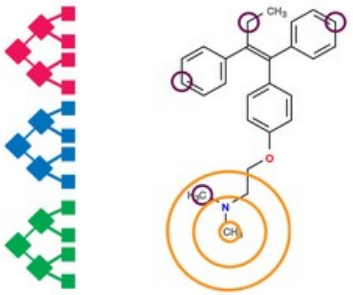
universität wien

Many of our models are accessible via a free web service at nerdd.zbh.uni-hamburg.de

NERDD

New E-Resource for Drug Discovery

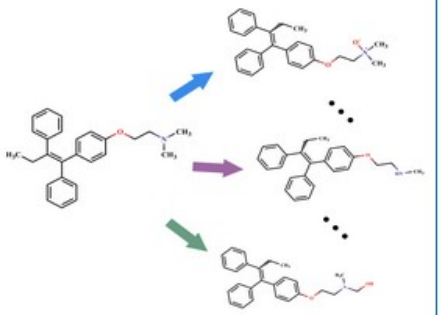
Sites of Metabolism



FAME 3

Regioselectivity prediction for phase 1 and phase 2 metabolism

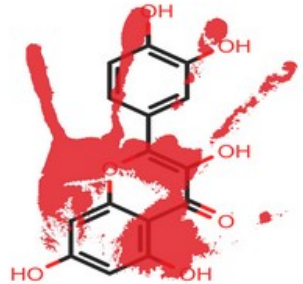
Metabolite Structures



GLORY

Metabolite structure prediction for cytochrome P450 metabolism

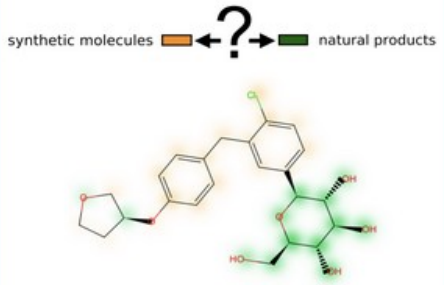
Frequent Hitters



Hit Dexter 2.0

Prediction of frequent hitters

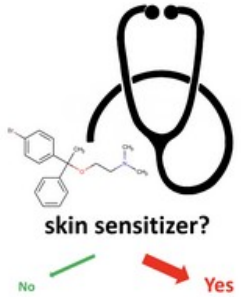
Natural Product-Likeness



NP-Scout

Identification and visualization of natural product-likeness

Skin Sensitization



Skin Doctor

Prediction of skin sensitization potential

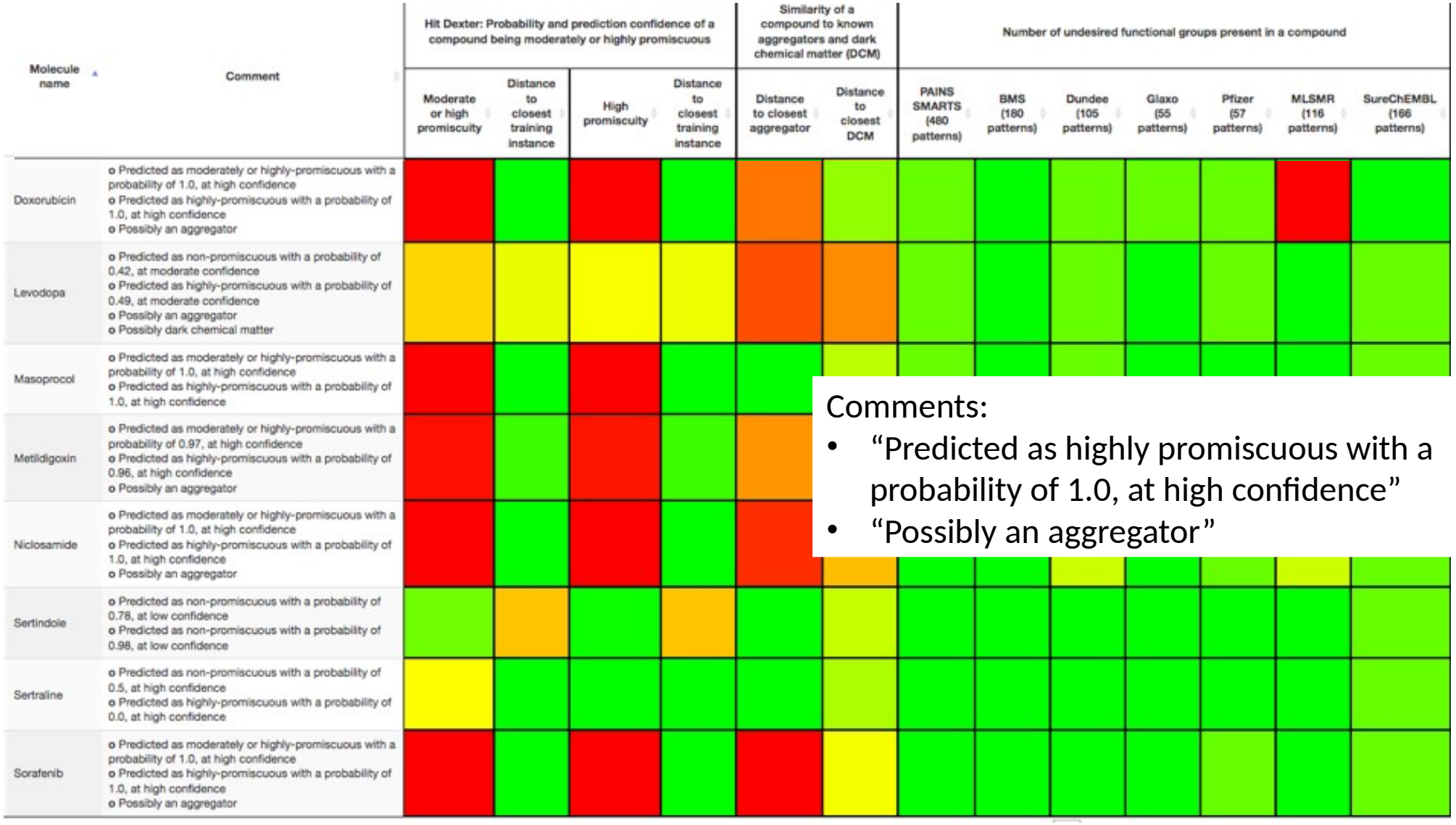
Stork et al., *Bioinformatics* 2019, 36, 1291–1292.

Case study on the 15 most noisiest, approved drugs identified by GSK[1]

Molecule name	Comment	Hit Dexter: Probability and prediction confidence of a compound being moderately or highly promiscuous				Similarity of a compound to known aggregators and dark chemical matter (DCM)		Number of undesired functional groups present in a compound						
		Moderate or high promiscuity	Distance to closest training instance	High promiscuity	Distance to closest training instance	Distance to closest aggregator	Distance to closest DCM	PAINS SMARTS (480 patterns)	BMS (180 patterns)	Dundee (105 patterns)	Glaxo (55 patterns)	Pfizer (57 patterns)	MLSMR (116 patterns)	SureChEMBL (166 patterns)
Apomorphine	o Predicted as moderately or highly-promiscuous with a probability of 1.0, at high confidence o Predicted as highly-promiscuous with a probability of 1.0, at high confidence													
Bithionol	o Predicted as moderately or highly-promiscuous with a probability of 1.0, at high confidence o Predicted as highly-promiscuous with a probability of 1.0, at high confidence													
Broxyquinolone	o Predicted as moderately or highly-promiscuous with a probability of 1.0, at high confidence o Predicted as highly-promiscuous with a probability of 1.0, at high confidence													
Clomipramine	o Predicted as non-promiscuous with a probability of 0.0, at high confidence o Predicted as highly-promiscuous with a probability of 0.13, at low confidence													
Daunorubicin	o Predicted as moderately or highly-promiscuous with a probability of 1.0, at high confidence o Predicted as highly-promiscuous with a probability of 1.0, at high confidence o Possibly an aggregator													
Deslanoside	o Predicted as moderately or highly-promiscuous with a probability of 0.62, at high confidence o Predicted as highly-promiscuous with a probability of 0.58, at moderate confidence o Possibly an aggregator													
Digitoxin	o Predicted as moderately or highly-promiscuous with a probability of 1.0, at high confidence o Predicted as highly-promiscuous with a probability of 1.0, at high confidence o Possibly an aggregator													

 [1] Chakravorty et al., SLAS Discovery 2018, 23, 532–544. DOI: 10.1177/2472555218768497

Case study on the 15 most noisiest, approved drugs identified by GSK[1]



Molecule name	Comment	Hit Dexter: Probability and prediction confidence of a compound being moderately or highly promiscuous				Similarity of a compound to known aggregators and dark chemical matter (DCM)		Number of undesired functional groups present in a compound						
		Moderate or high promiscuity	Distance to closest training instance	High promiscuity	Distance to closest training instance	Distance to closest aggregator	Distance to closest DCM	PAINS SMARTS (480 patterns)	BMS (180 patterns)	Dundee (105 patterns)	Glaxo (55 patterns)	Pfizer (57 patterns)	MLSMR (116 patterns)	SureChEMBL (166 patterns)
Doxorubicin	o Predicted as moderately or highly-promiscuous with a probability of 1.0, at high confidence o Predicted as highly-promiscuous with a probability of 1.0, at high confidence o Possibly an aggregator													
Levodopa	o Predicted as non-promiscuous with a probability of 0.42, at moderate confidence o Predicted as highly-promiscuous with a probability of 0.49, at moderate confidence o Possibly an aggregator o Possibly dark chemical matter													
Masoprocol	o Predicted as moderately or highly-promiscuous with a probability of 1.0, at high confidence o Predicted as highly-promiscuous with a probability of 1.0, at high confidence													
Metildigoxin	o Predicted as moderately or highly-promiscuous with a probability of 0.97, at high confidence o Predicted as highly-promiscuous with a probability of 0.96, at high confidence o Possibly an aggregator													
Niclosamide	o Predicted as moderately or highly-promiscuous with a probability of 1.0, at high confidence o Predicted as highly-promiscuous with a probability of 1.0, at high confidence o Possibly an aggregator													
Sertindole	o Predicted as non-promiscuous with a probability of 0.78, at low confidence o Predicted as non-promiscuous with a probability of 0.98, at low confidence													
Sertraline	o Predicted as non-promiscuous with a probability of 0.5, at high confidence o Predicted as highly-promiscuous with a probability of 0.0, at high confidence													
Sorafenib	o Predicted as moderately or highly-promiscuous with a probability of 1.0, at high confidence o Predicted as highly-promiscuous with a probability of 1.0, at high confidence o Possibly an aggregator													

Comments:

- “Predicted as highly promiscuous with a probability of 1.0, at high confidence”
- “Possibly an aggregator”

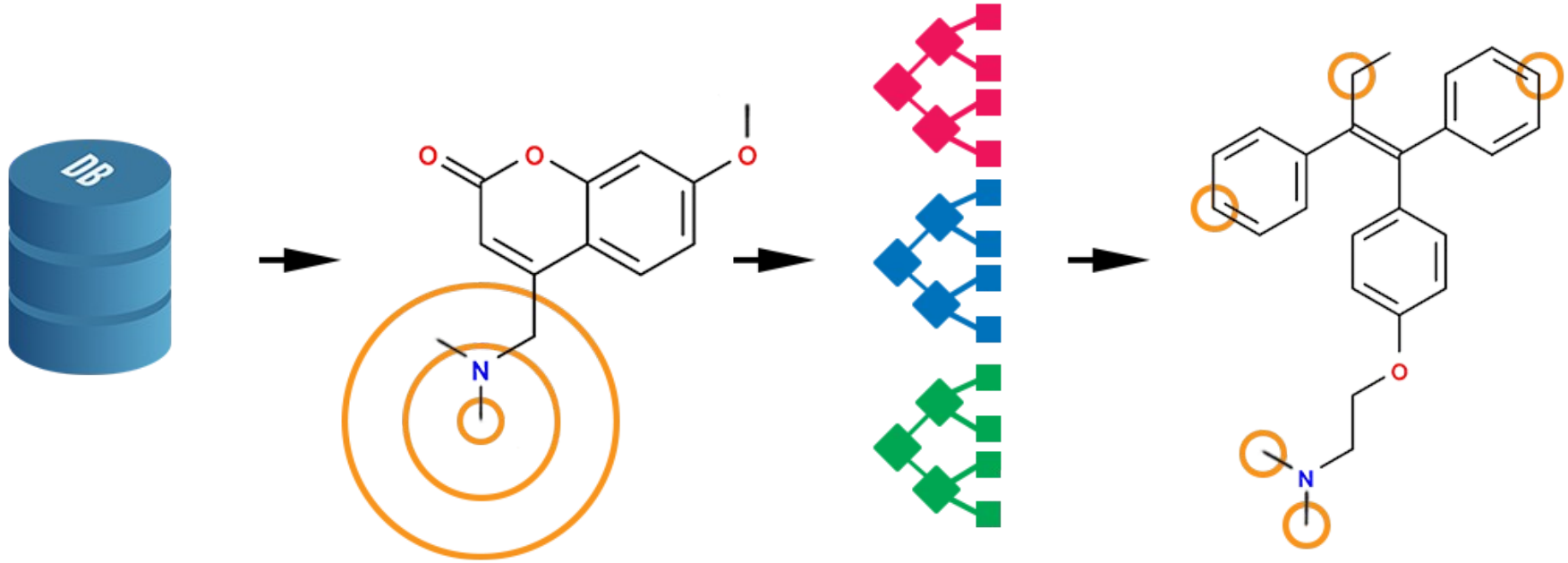
 [1] Chakravorty et al., SLAS Discovery 2018, 23, 532–544. DOI: 10.1177/2472555218768497

Development of FAst Metabolizer - FAME 3

UNIL | Université de Lausanne

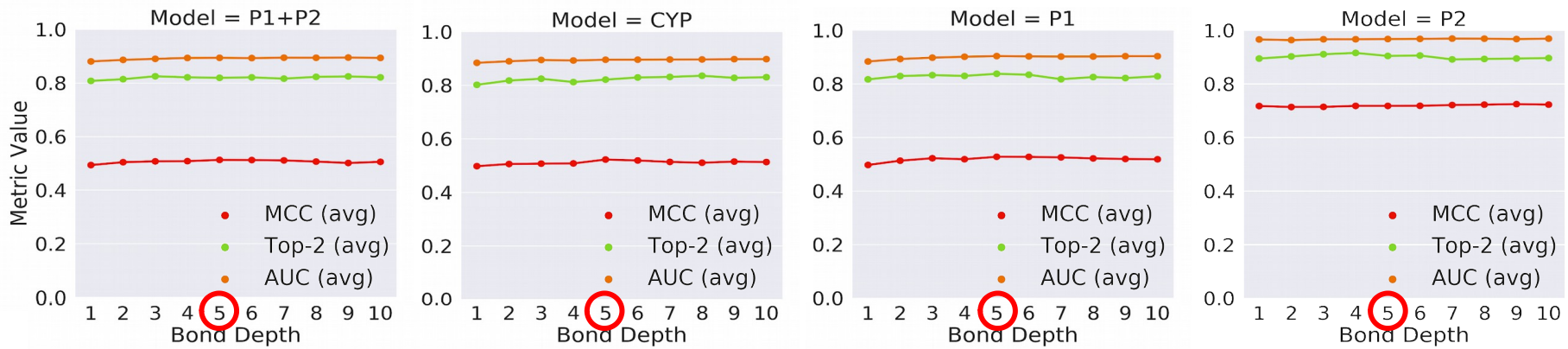
UNIVERSITÀ DEGLI STUDI DI MILANO

Database → Atom environment encoding → Machine learning → Site of metabolism prediction



FAME 3: Performance of “circCDK+ATF” models

10-fold cross-validation

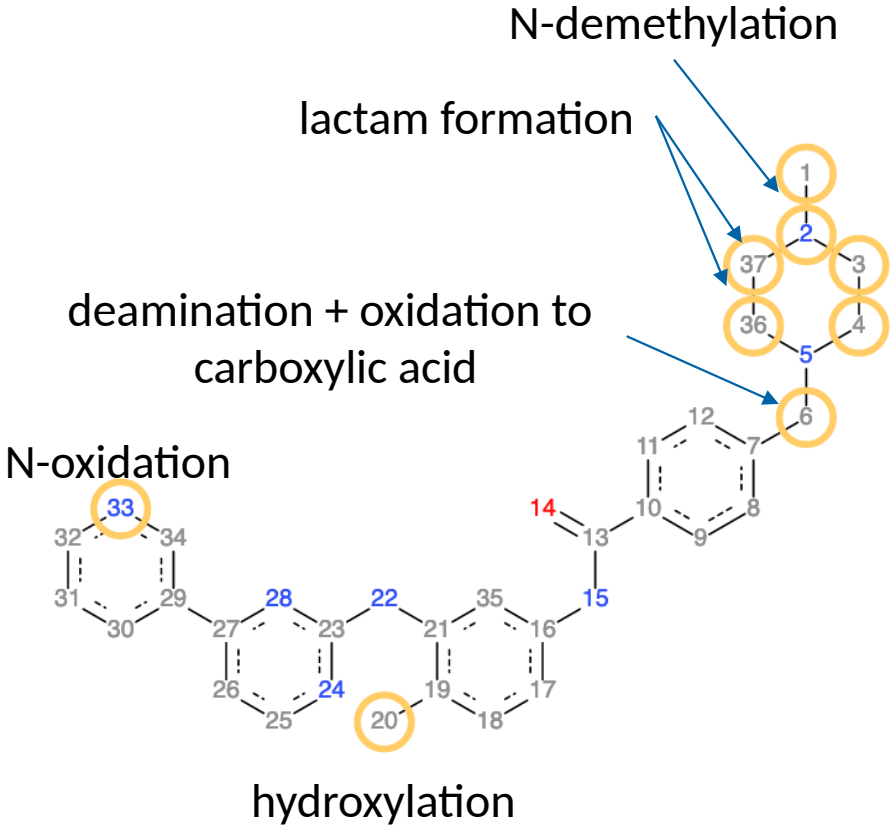


test on holdout data

bond depth=5

Model	MCC	AUC	Top-2
P1+P2	0.50	0.90	82%
P1+P2 100+	0.55	0.92	87%
CYP	0.57	0.92	90%
CYP 100+	0.63	0.94	86%
P1	0.53	0.88	83%
P1 100+	0.52	0.92	80%
P2	0.71	0.97	92%
P2 100+	0.75	0.97	91%

FAME 3: Prediction of the sites of metabolism of imatinib

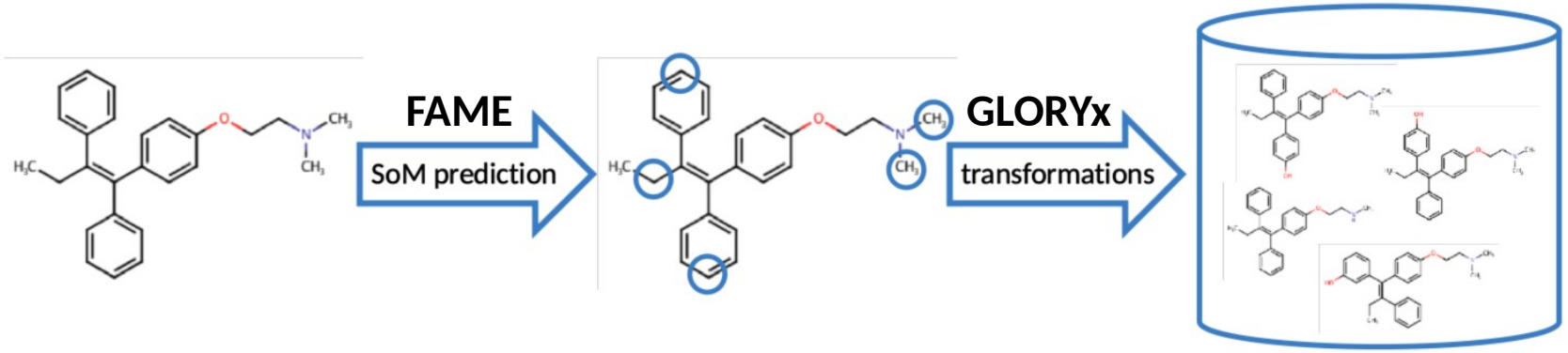


Model: P1+P2 (depth: 5)

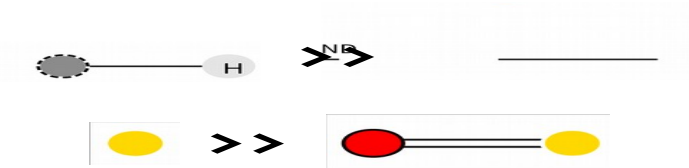
Molecule mol_1

Atom	Probability	FAMEscore
N.2	**0.888**	**0.785**
C.1	**0.884**	**0.809**
C.36	**0.684**	**0.944**
C.4	**0.684**	**0.944**
C.6	**0.668**	**0.808**
C.20	**0.66**	**0.826**
C.37	**0.652**	**0.939**
C.3	**0.652**	**0.939**
N.33	**0.644**	**0.912**
C.13	0.128	0.804
N.22	0.044	0.814
N.5	0.044	0.788

GLORYx: Predictor of likely metabolites



1. Extracted reaction types for phase I and phase II enzymes from the literature
2. Represented reaction types by SMIRKS:
   - e.g. "[c:1][H:2]>>[c:1][O][H:2]"
3. Applied transformations using AMBIT SMIRKS
   - Open-source Java library (IdeaConsult Ltd)
4. The transformations are only applied at those positions



GLORYx: Performance on the Manually Curated Test Set

	GLORYx	SyGMa
Recall	0.77	0.68
Precision	0.061	0.120
Total # predictions	1724	800
# true positives	105	93