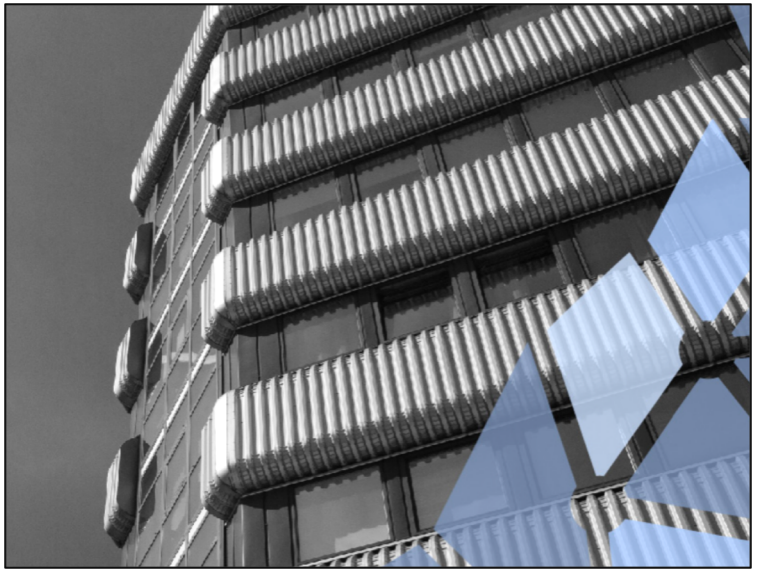
FAME 3: Predicting the Sites of Metabolism in Small Molecules for Phase 1 and Phase 2 Metabolic Enzymes

Assoc. Prof. Dr. Johannes Kirchmair
University of Bergen
Department of Chemistry and
Computational Unit (CBU)
Bergen, Norway

and

University of Hamburg
Centre for Bioinformatics
Hamburg, Germany

cbu.uib.no
johannes.kirchmair@uib.no



Acknowledgements

- University of Lausanne
  - Prof. Bernard Testa
- University of Milan
  - Prof. Alessandro Pedretti
  - Prof. Giulio Vistoli
  - Angelica Mazzolari
- University of Chemistry and Technology, Prague
  - **Daniel Svozil**
  - **Martin Šícho**

FUJITSU

Beiersdorf

TROND MOHN STIFTELSE

Erasmus+

DFG Deutsche Forschungsgemeinschaft

CHINA SCHOLARSHIP COUNCIL

- Universities of Bergen & Hamburg
  - Christoph Bauer
  - Ya Chen
  - Christina de Bruyn Kops
  - Ningning Fan
  - Nora Constanze Fohrmann
  - Nils-Ole Friedrich
  - Nina Nelson
  - Neann Mathai
  - Wojtek Plonka
  - **Conrad Stork**
  - Anke Wilm

Understanding xenobiotic metabolism is key to the design of safe and efficacious small molecules

- Metabolism is the main clearance pathway of 75 to 90% of all drugs
- Xenobiotics on average have six metabolites[1]
- Only 3% of all drug metabolites are confirmed to maintain their pharmacological activity[1]
- At least 7% of all metabolites are known to be toxic or reactive[1]

Opportunities

Detoxification

Targeted (de-) activation

- Organisms, tissues, cells

Challenges and Risks

(De-) activation

Toxification

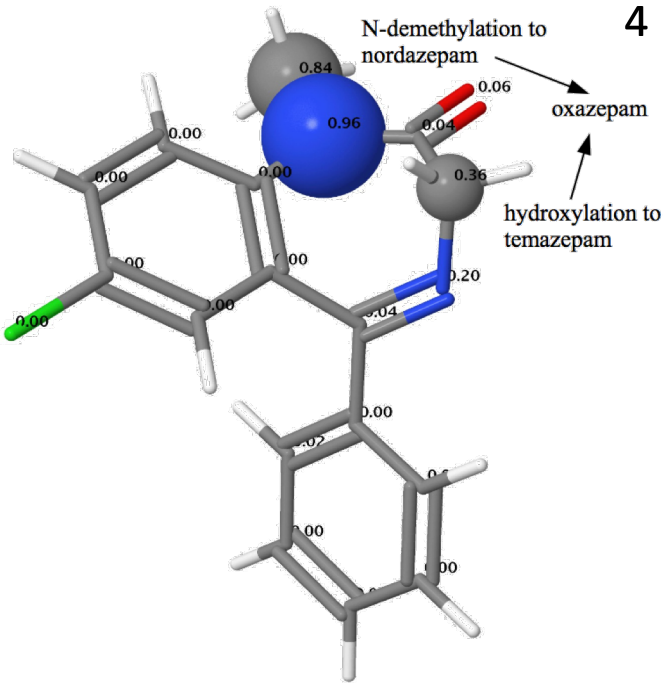
Changes in distribution

Drug-drug interaction

Drug resistance

[1]Testa et al, Drug Discov Today 2012, 17, 549-560. doi: 10.1016/j.drudis.2012.01.017

What atoms of my small-molecule are susceptible to metabolism?



- Knowing the SoMs in a molecule can aid the derivation of likely metabolites and hence, optimisation strategies
- Models based on diverse approaches

\+ Several good models available for CYPs, few for other metabolizing enzymes

\+ Some models cover different mammalian species

\+ Accuracy: At least one known SoM among the top-2 ranked atom positions in a molecule in >85% of all cases

\+ Large applicability domain

– Most models limited to CYPs

– Most models lack definition of applicability domain and error estimation

– Models able to discriminate major and minor metabolites at most

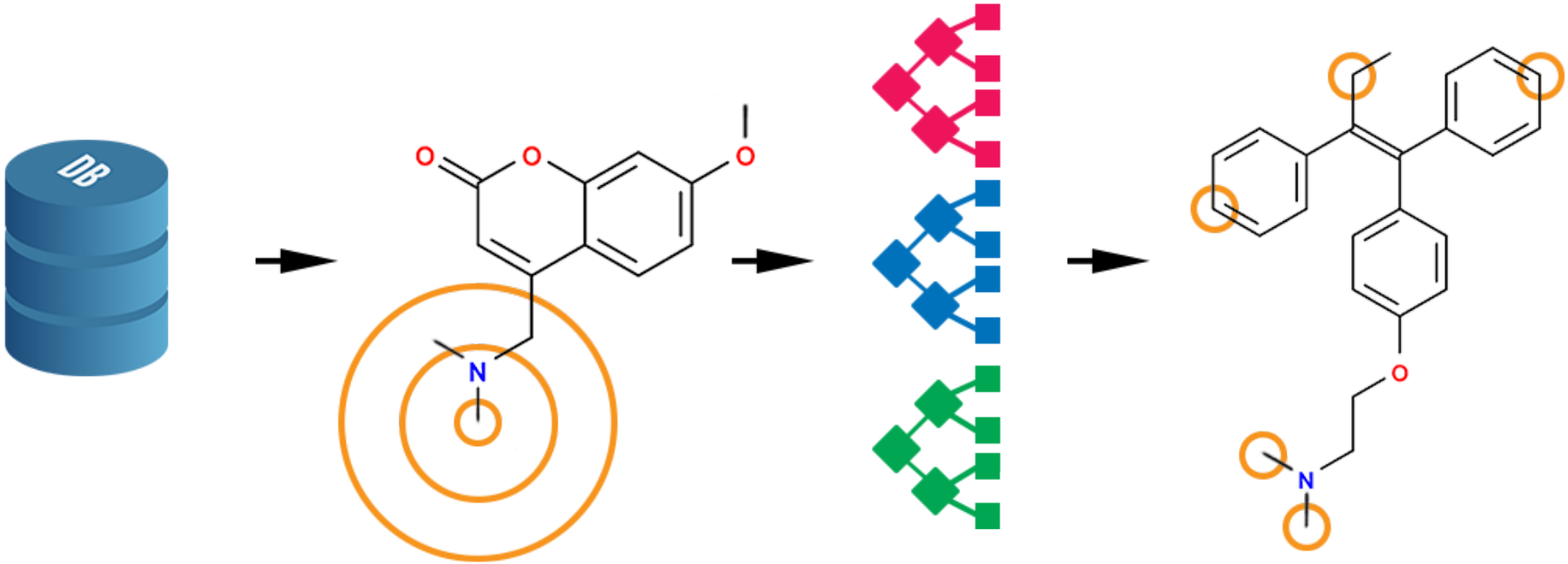
Development of FAst Metabolizer - FAME 3

Martin Sicho

UNIL | Université de Lausanne

UNIVERSITÀ DEGLI STUDI DI MILANO

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



Sicho et al., J Chem Inf Model 2019, 59, 3400-3412. doi: 10.1021/acs.jcim.9b00376

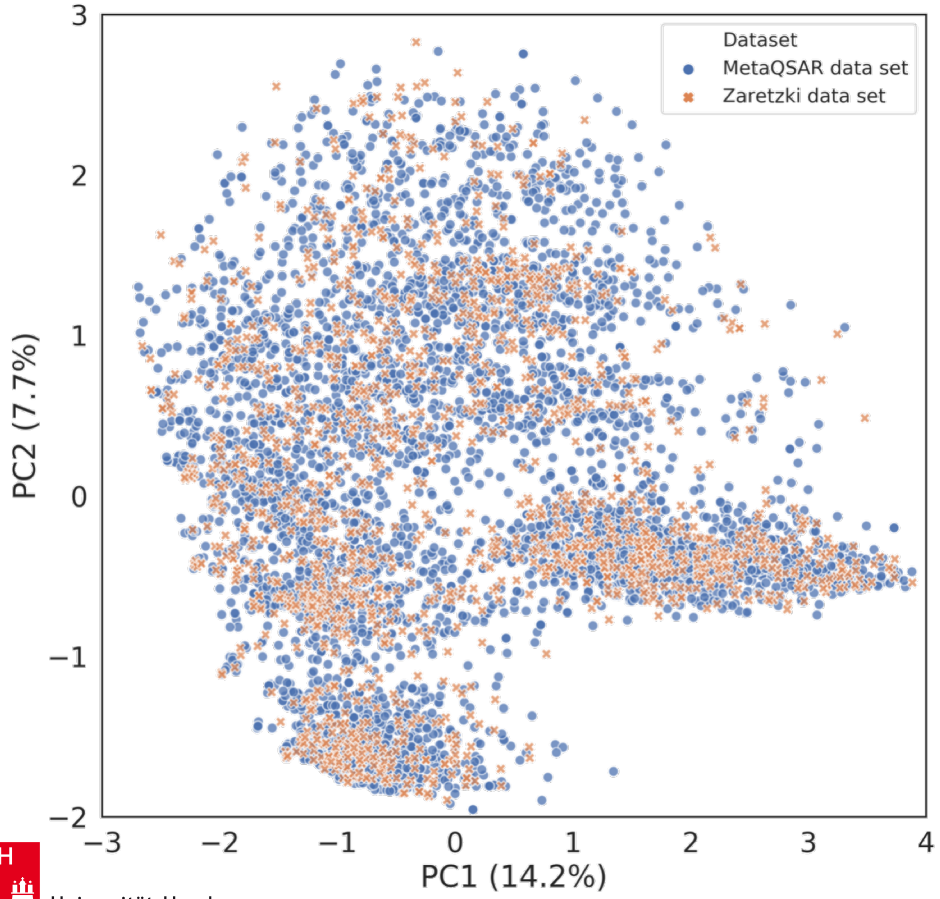
FAst Metabolizer (FAME)

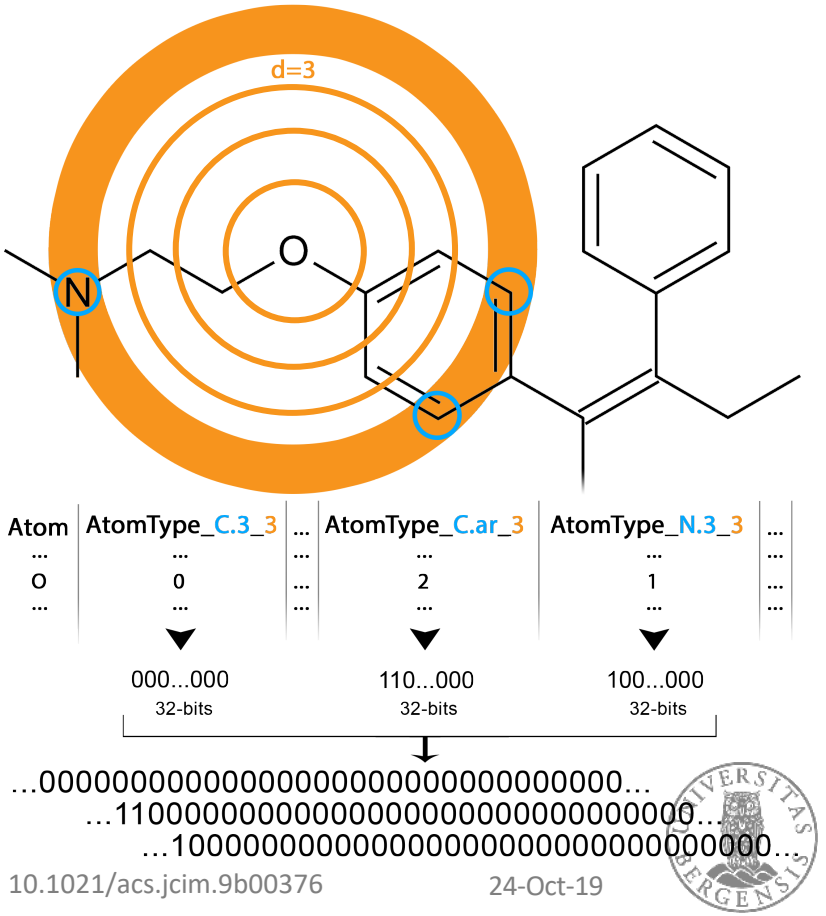
	FAME 1 (2013)	FAME 2 (2017)	FAME 3 (2019)
Training set source	Metabolite DB (proprietary, discontinued)	Zaretzki Dataset	MetaQSAR DB
Training set size	Up to ~21,000 substrates	Up to ~540 substrates	Up to ~2150 substrates
CYP P450 enzymes	Yes	Yes	Yes
Phase 1 metabolism	Yes	CYPs only	Yes
Phase 2 metabolism	Yes	No	Yes
SoM quality	Automated assignment based on substructure matching	Expert-curated but some quality issues	**Expert-curated**
Machine learning approach	Random forest	Extremely randomized trees	
Descriptors	15 2D-descriptors including Sybyl atom types	Circular fingerprints encoding Sybyl atom types plus 15 2D-descriptors	
Applicability domain definition and error estimation	No	No	**Yes**
Prediction accuracy	Mediocre	High	High
Availability	Discontinued	Software package	Software package and web service

Sicho et al., J Chem Inf Model 2019, 59, 3400-3412. doi: 10.1021/acs.jcim.9b00376

Characterization of the MetaQSAR database

Content	No. of Substrates	No. of reactions
Total	2167	6307
CYP substrates	1106	3517
Phase 1 (excl. CYP)	622	1551
Phase 2	784	1239
Zaretzki dataset	678	1672



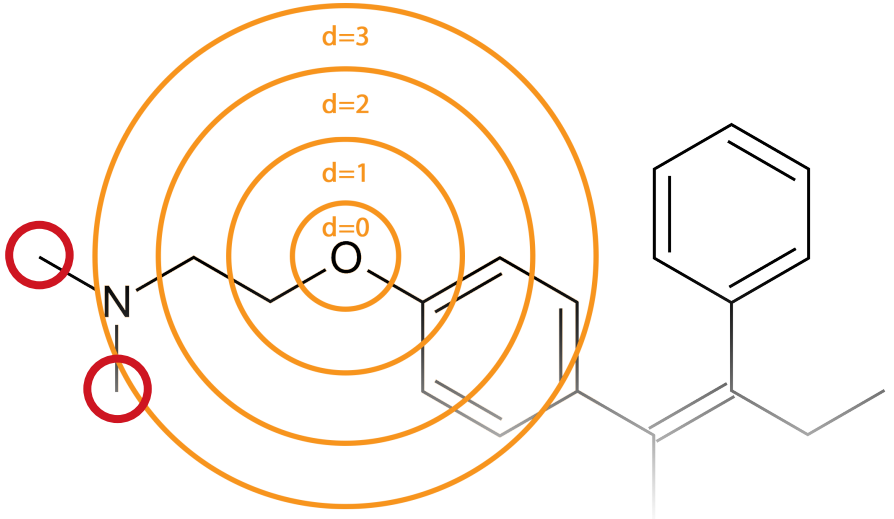


Sicho et al., J Chem Inf Model 2019, 59, 3400-3412. doi: 10.1021/acs.jcim.9b00376

FAME 3: Atom descriptors

- Four sets of descriptors have been explored
- Combination of ATFs with circCDK descriptors identified as most suitable descriptors set

Acronym	Description
ATF	Circular fingerprint based on Sybyl atom types
CDK	15 Basic 2D descriptors implemented in CDK
circCDK	Circular descriptors derived from the CDK descriptor set
QC	10 AM1-based descriptors calculated with MOPAC



Atom	AtomType_O.3_0	...	AtomType_C.ar_3	AtomType_N.3_3	...
...	...	...	...	...	...
O	1	...	2	1	...
...	...	...	...	...	...

fingerprints

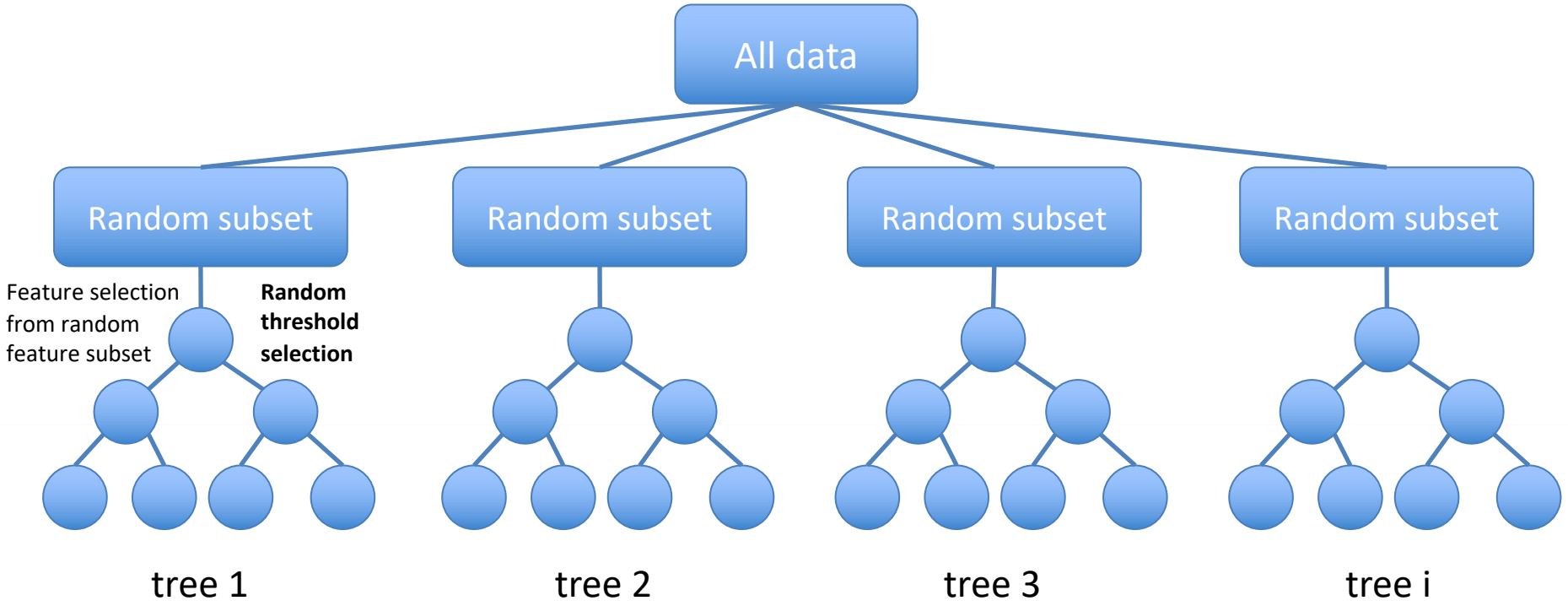
Atom	sigmaEng_O.3_0	...	sigmaEng_C.ar_3	sigmaEng_N.3_3	...
...	...	...	...	...	...
O	9.91	...	8.31	8.35	...
...	...	...	...	...	...

descriptors

Sicho et al., J Chem Inf Model 2017, 57, 1832–1846. doi: 10.1021/acs.jcim.7b00250
Sicho et al., J Chem Inf Model 2019, 59, 3400-3412. doi: 10.1021/acs.jcim.9b00376

FAME 3: Model development

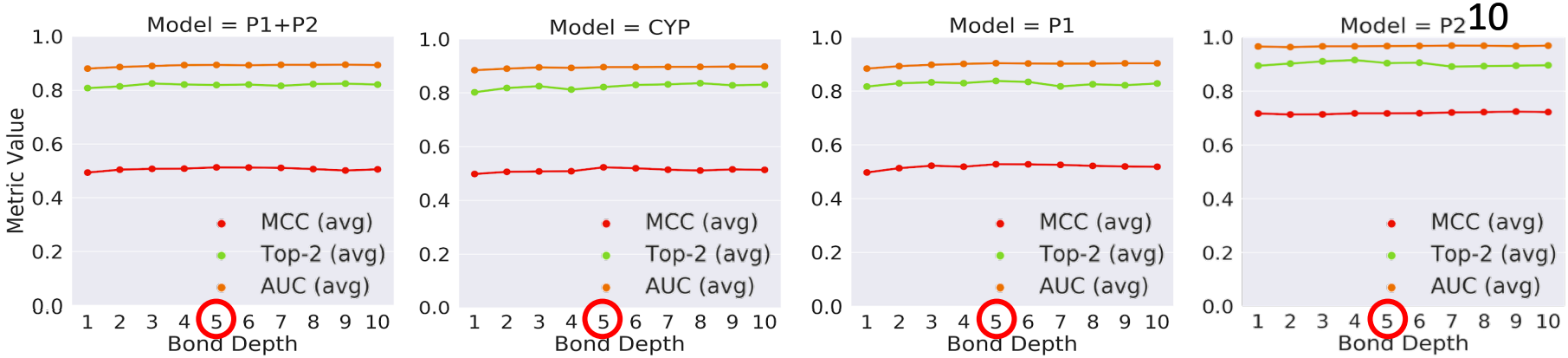
- MetaQSAR database split into training set (80%) and test set (20%)
- Four different sets of descriptors (ATF, CDK, circCDK and QC) explored
- Feature reduction down to max. of 400 by ANOVA F-Test
- Model generation: Extremely randomized trees
- Hyperparameters derived by grid search with 10-fold cross-validation



Sicho et al., J Chem Inf Model 2019, 59, 3400-3412. doi: 10.1021/acs.jcim.9b00376

FAME 3: Performance of “circCDK+ATF” models

10-fold cross-validation

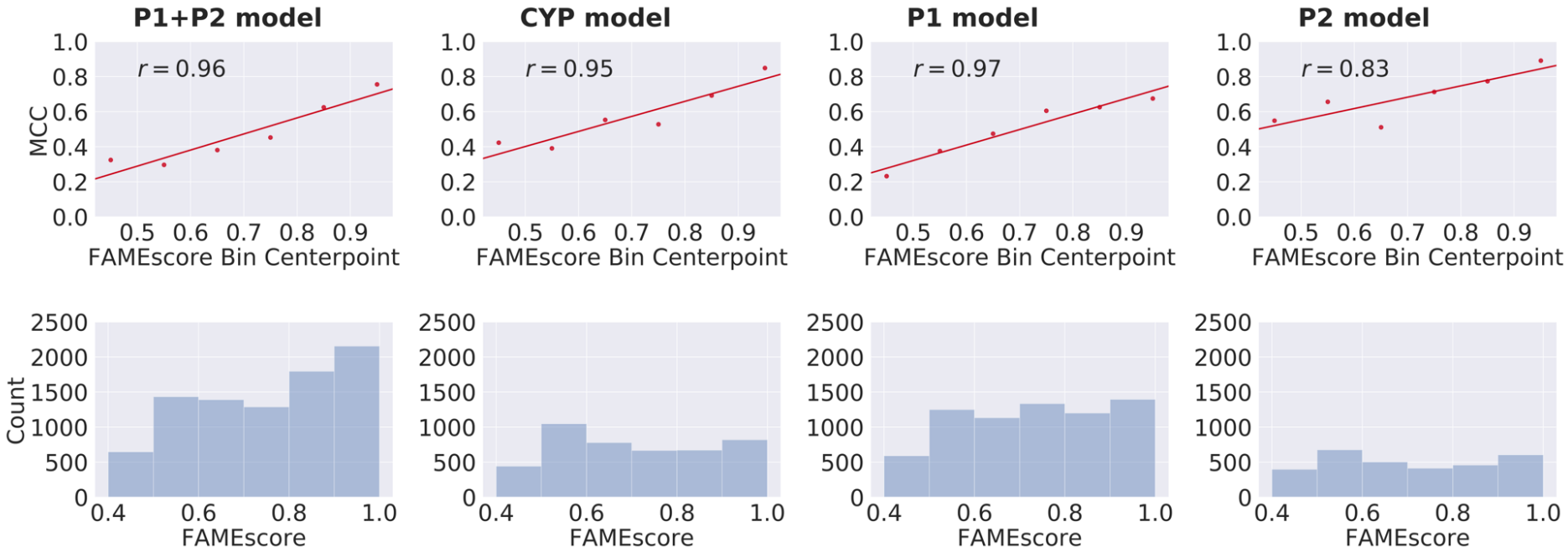


test on holdout data

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%

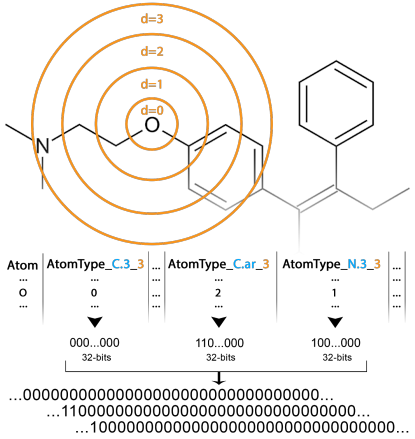
FAME 3: Performance of the final models on holdout data



$$FAMEscore = 1 - \frac{\sum_{i=1}^{k} d_i}{k}$$

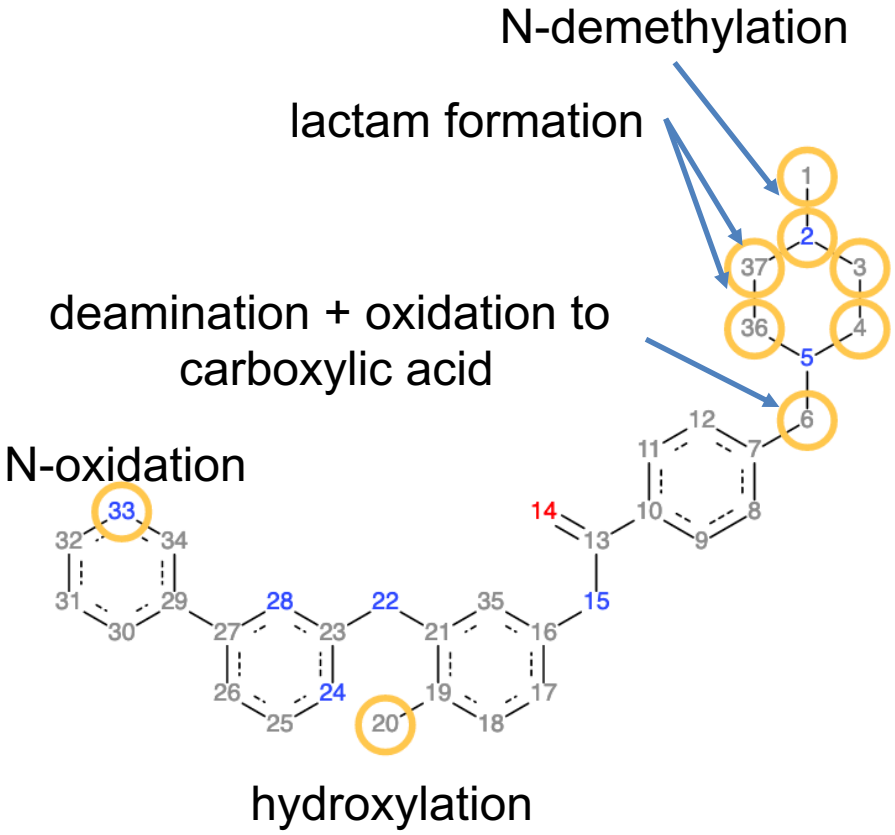
d… distance (Tanimoto coefficient)

k… number of nearest neighbours (we use *k*=3)



Sicho et al., J Chem Inf Model 2019, 59, 3400-3412. doi: 10.1021/acs.jcim.9b00376

FAME 3: Prediction of the sites of metabolism of imatinib



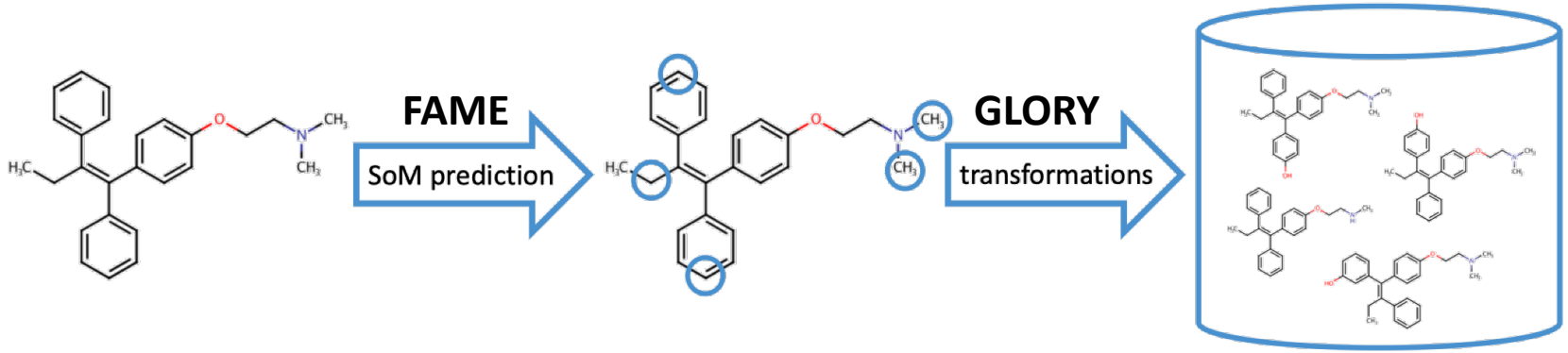
Model: P1+P2 (depth: 5)

Molecule mol_1

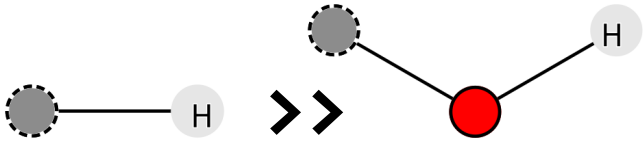
Atom	Probability	FAMEscore
N.2	**0.888**	**0.785**
C.1	**0.884**	**0.809**
C.36	**0.684**	**0.944**
C.4	**0.684**	**0.944**
C.6	**0.668**	**0.808**
C.20	**0.66**	**0.826**
C.37	**0.652**	**0.939**
C.3	**0.652**	**0.939**
N.33	**0.644**	**0.912**
C.13	0.128	0.804
N.22	0.044	0.814
N.5	0.044	0.788

GLORY: Predictor of likely metabolites

Christina
de Bruyn Kops



1. Extracted CYP-mediated reaction types 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. If sites of metabolism were predicted, the transformations are only applied at those positions



de Bruyn Kops et al., Front Chem 2019, 7, 402. doi: 10.3389/fchem.2019.00402

FAME 3: CDK descriptors

Name	Description
effectiveAtomPolarizability	Effective polarizability of a heavy atom
stabilizationPlusCharge	Positive charge stabilization of an atom
sigmaElectronegativity	Gasteiger-Marsili σ-electronegativity of an atom
piElectronegativity	Gasteiger-Marsili π-orbitals electronegativity of an atom
partialSigmaCharge	Gasteiger-Marsili σ-orbitals partial charges
partialTChargeMMFF94	Partial charge of a heavy atom derived from the MMFF94
AtomType	Sybyl atom type, describing the element type and hybridization state
atomDegree	Number of non-hydrogen atoms bonded to an atom
atomValence	Formal valence of an atom
atomHybridizationVSEPR	Hybridization state of an atom according to VSEPR
atomHybridization	Hybridization state of an atom
longestMaxTopDistInMolecule	Maximum number of bonds (topological distance) between two atoms in the molecule (i.e. the maximum in the topological distance matrix for the whole molecule)
highestMaxTopDistInMatrixRow	Maximum number of bonds (topological distance) between two particular atoms in the molecule (maximum in the particular topological distance matrix row)
diffSPAN	longestMaxTopDistInMolecule - highestMaxTopDistInMatrixRow
relSPAN	highestMaxTopDistInMatrixRow / longestMaxTopDistInMolecule

Sicho et al., J Chem Inf Model 2017, 57, 1832–1846. doi: 10.1021/acs.jcim.7b00250