Supplementary Material for

SMARTCyp 3.0: Enhanced cytochrome P450 site-of-metabolism prediction server

Lars Olsen[1,2], Marco Montefiori[1,3], Khanhvi Phuc Tran[1] and Flemming Steen Jørgensen[1,*]

[1]Department of Drug Design and Pharmacology, Faculty of Health and Medical Sciences, University of Copenhagen, Universitetsparken 2, DK-2100, Copenhagen Ø, Denmark, [2]Present address: Protein Engineering, Novozymes A/S, Krogshøjvej 36, DK-2880 Bagsvaerd, Denmark, [3]Present address: CNR-ICRM, Via Mario Bianco 9, 20131 Milan, Italy.

Major differences between SMARTCyp version 2.4 and 3.0

Comparison of key characteristics for SMARTCyp 2.4 and 3.0 are collected in Table S1.

Table S1. Major differences between SMARTCyp version 2.4 and 3.0

SMARTCyp version	2.4.2	3.0
Programming language	Java	Python 3.6
Cheminformatics toolkit	CDK	RDKit
Web interface	PHP	Flask
Fragment rules	55	71
Similarity feature (structure match quality check)	No	Yes
Prediction accuracy of standard model	0.72	0.73

Translating the SMARTS rules from CDK to RDKit

The SMARTS rules in the fragment library, originally made for the CDK software, were evaluated and translated to enable a correct interpretation by the RDKit. Initially 407 SMARTS rules were evaluated in an iterative process. When modifying one rule, several new deviations occurred, and the rules had to be re-evalutated. About 10 iterations were required to achieve identical performance of SMARTCyp 3.0 with the previous version.

The CDK and RDKit software handles aromaticity different. Accordingly, it was necessary to add exceptions or modify some of the rules. Below, we have presented two examples on rules, which were changed from version SMARTCyp 2.4 to 3.0. Changes are shown in bold.

Rule 32 detects ortho and para carbons in aniline in CDK, in which the aniline nitrogen is sp^3 hybridized. The rule was modified to detect all types of hybridizations of aniline nitrogens to match the same carbon atoms in the RDKit, in which the aniline nitrogen is sp^2 hybridized (cf. Table S2).

Table S2. SMARTS rule 32 as defined in SMARTCyp 2.4 and 3.0, respectively.

SMARTCyp 2.4 Rule 32
[$([ch1]1:[c](-[**N^3**]([C^3])[C^3]):[c]:[c]:[c]:[c]1),$([ch1]1:[c]:[c]:[c](-[**N^3**]([C^3])[C^3]):[c]:[c]1), $([ch1]1:[c](-[**N^3H1**][C^3]):[c]:[c]:[c]:[c]1),$([ch1]1:[c]:[c]:[c](-[**N^3H1**][C^3]):[c]:[c]1), $([ch1]1:[c](-[**N^3H2**]):[c]:[c]:[c]:[c]1),$([ch1]1:[c]:[c]:[c](-[**N^3H2**]):[c]:[c]1)]
SMARTCyp 3.0 Rule 32
[$([ch1]1:[c](-[**N**]([C^3])[C^3]):[c]:[c]:[c]:[c]1),$([ch1]1:[c]:[c]:[c](-[**N**]([C^3])[C^3]):[c]:[c]1), $([ch1]1:[c](-[**NH1**][C^3]):[c]:[c]:[c]:[c]1),$([ch1]1:[c]:[c]:[c](-[**NH1**][C^3]):[c]:[c]1), $([ch1]1:[c](-[**NH2**]):[c]:[c]:[c]:[c]1),$([ch1]1:[c]:[c]:[c](-[**NH2**]):[c]:[c]1)]

Rule 34 detects para positioned carbons in two kinds of benzene rings with either a nitrogen bonded to an sp^2 carbon site chain, or as part of an indole moiety. The rule includes an exception preventing it from detecting fragments of a benzene ring with a amide side chain. The rule was modified to prevent detection of fragments of a benzene ring with both aliphatic and aromatic amide side chains (cf. Table S3).

Table S3. SMARTS rule 34 as defined in SMARTCyp 2.4 and 3.0, respectively.

SMARTCyp 2.4 Rule 34
[$([ch1]1:[c]:[c]:[cr6]([#7]~[#6^2]):[c]:[c]1),$([ch1r6]:[cr5]:[nX3r5]); !$([c]1:[c]:[c]:[c](**-[N]-[C]**=[O]):[c]:[c]1)]
SMARTCyp 3.0 Rule 34
[$([ch1]1:[c]:[c]:[cr6]([#7]~[#6^2]):[c]:[c]1),$([ch1r6]:[cr5]:[nX3r5]); !$([c]1:[c]:[c]:[c](**[#7][#6]**=[O]):[c]:[c]1)]

The updated SMARTS rules for version 3.0 are listed in Table S4 (rules 1-60).

New rules in SMARTCyp 3.0

Another prominent problem was when an atom did not match any rule, and therefore was assigned wrong activation energy either by matching a wrong SMARTS string, or matching no strings at all, which deems the atom not reactive.

In total, 16 new rules were added from different sources: 11 rules were derived from DFT calculations on new fragments (Table S1, rules 61-71), 14 energies from the work by Leth et al.[1] were grouped into a single rule (Table S1, rule 56), and 4 rules from the work of Bonomo et al.[2] (Table S1, rules 57-60). The DFT determined activation energies are the difference between the transition state energy and the energy of the reactant complex. The hydrogen abstraction complexes were calculated as the quartet spin states, all other complexes were calculated as the doublet spin states, and all energies were corrected for zero-vibrational energy. Further details on the DFT calculations can be found in Leth et al.[1]

Prediction of isoform specific metabolism

The SMARTCyp program predicts the site of metabolism for three of the most prevalent CYP isoforms. The SMARTCyp scores for prediction of 3A4 metabolism is based on the DFT determined activation energies corrected for the accessibility of the atom as described previously [3]:

$$\text{3A4 score} = \text{DFT energy} - 8 * \text{accessibility} - 0.04 * \text{SASA}$$

The SMARTCyp scores for prediction of 2D6 and 2C9 metabolism are slightly more complex:

$$\text{2D6 score} = \text{DFT energy} + 6.7 * (8 - \text{N+dist} + \text{Span2End})$$

$$\text{2C9 score} = \text{DFT energy} + 5.9 * (8 - \text{COO-dist} + \text{Span2End}) - 0.04 * \text{SASA}$$

where N+dist and COO-dist are distance corrections. The basis for the isoform specific corrections has been described in previous papers.[4,5]

Performance

A comparison of the performance of SMARTCyp 3.0 with the previous the version 2.4 based on the 375 3A4 compound dataset from Zaretzki et al.[6] revealed that the prediction is identical for the two SMARTCyp versions. Addition of the 16 new rules resulted in essentially equal performance (cf. Table S4). Dissecting the predictions lead to the following number of correct predicted SOMs (cf. Table S5).

Table S5. Comparison of performance of SMARTCyp 2.4 and 3.0 without and with new rules.

Predicted accuracy measured by AUC for the three SMARTCyp versions			
Isoform	v. 2.4	v. 3.0	v. 3.0 with new rules
3A4	0.72	0.72	0.73
2D6	0.78	0.78	0.78
2C9	0.75	0.75	0.76
Predicted true positive SOM			
Isoform	v. 2.4	v. 3.0	v. 3.0 with new rules
3A4	340	340	342
2D6	209	208	208
2C9	168	168	169

References

1. Leth R, Rydberg P, Jørgensen FS, Olsen L. Density functional theory study on the formation of reactive benzoquinone imines by hydrogen abstraction. J Chem Inf Model 2015;55:660-6.

2. Bonomo S, Jorgensen FS, Olsen L. Dissecting the Cytochrome P450 1A2- and 3A4-Mediated Metabolism of Aflatoxin B1 in Ligand and Protein Contributions. Chemistry 2017;23:2884-93.

3. Rydberg P, Rostkowski M, Gloriam DE, Olsen L. The Contribution of Atom Accessibility to Site of Metabolism Models for Cytochromes P450. Molecular Pharmaceutics 2013;10:1216-23.

4. Rydberg P, Olsen L. Predicting Drug Metabolism by Cytochrome P450 2C9: Comparison with the 2D6 and 3A4 Isoforms. ChemMedChem 2012;7:1202-9.

5. Rydberg P, Olsen L. Ligand-Based Site of Metabolism Prediction for Cytochrome P450 2D6. ACS Medicinal Chemistry Letters 2012;3:69-73.

6. Zaretzki J, Rydberg P, Bergeron C, Bennett KP, Olsen L, Breneman CM. RS-Predictor Models Augmented with SMARTCyp Reactivities: Robust Metabolic Regioselectivity Predictions for Nine CYP Isozymes. J Chem Inform Model 2012;52:1637-59.

Supplementary Material Table S4. SMARTS rules for SMARTCyp 3.0

Rule	SMARTS	Energy [kJ/mol]
1	[SX2H1]	41.5
2	[$([SX2H0]);!$([S][*^2]);!$([S][CX4H0]);!$([S][P])]	26.3
3	[$([SX2H0][*^2]),$([S][P]=[S]);!$([S](~[^2])[^2]);!$([S][CX4H0])]	34.5
4	[$([S][*D4H0]);$([SX2H0])]	44.4
5	[$([SX2H0]([*^2])[*^2]);!$([S][CX4H0])]	46.9
6	[$([sX2r5]1:[c]2:[c]:[c]:[c]:[c]:[c]2:[c]:[c]1);$([sX2r5])]	56.9
7	[sX2r5]	70.0
8	[$([#16X3](=[OX1]));$([#16]);!$([#16X3](=[OX1])[#6^2](~[#7^2]));!$([#16X3](=[OX1])[OH1])]	30.4
9	[$([#16X3](=[OX1]));$([#16]);$([#16X3](=[OX1])[#6^2](~[#7^2]))]	46.9
10	[$([CX3H1](=O)[#6])]	40.2
11	[$([S]=[PX4])]	50.9
12	[$([S]=[PX4][S])]	40.0
13	[$([S]=[C])]	43.5
14	[$([CX4][N^3R2r6]1[C]2[C][C][C][C]1[C][C][C]2);!$([CH0])]	22.9
15	[$([CX4]1[N^3][C][C]=[C]1);!$([CH0])]	28.3
16	[!$([CH0]);$([CX4]([#6^2])([#6^2])[#6^2]),$([CX4][N^3R0][C]=[C]),$([CX4R0,CX4R1][N^3r5]);!$([CX4][N][*^2]);!$([CX4][NX3][#16X4](=[OX1])(=[OX1]))]	32.8
17	[$([CX4][N]);!$([CH0]);!$([C][N]([*^2])[*^2]);!$([C][N]=[#6X3]);!$([CX4][NX3][C]=[O]);!$([CX4][NX3][#16X4](=[OX1])(=[OX1]));!$([CX4][NX3][N]=[O]);!$([CX4]1[C][C]2[C][C][N]1[C][C]2);!$([CX4H1](@[C])(@[C])@[N^3]);!$([CX4]1[N][C][C][N]([*^2])[C]1)]	41.1
18	[!$([CH0]);$([CX4]([C^2]~[C])[C^2]~[C]),$([CX4][#7]=[#6X3]),$([CX4]([#8])[#8]);!$([CX4]([#8])[#8][C]=[O]);!$([C]1[O][C]=[C][C]1)]	48.5
19	[$([CX4][S]);!$([CH0]);!$([C][S]=[O]);!$([C][S][P]=[S])]	57.7
20	[$([CX4][#6^2]~[#8]),$([CX4][cr5]),$([CX4]([c])[c]),$([CX4]([c])[C]=[C]),$([CX4][#6^1]),$([CX4][C^2]=[C^2]-[#6^2]),$([CX4][NX3][N]=[O]),$([CX4]1[N][C][C][N]([*^2])[C]1);!$([CH0]);!$([CX4][C](=[O])[NX3]);!$([CX4][#6^2](=[#8])-[#8]);!$([CX4][C^2]([C^2])=[C^2]-[#6^2]);!$([CX4][#6^2](=[#8])[#6^2])]	59.9
21	[$([CX4][O]);!$([CH0]);!$([C][O][C]=[O]);!$([CX4]1[O][C]1);!$([C][O][P]=[S])]	62.2
22	[$([CX4][NX3H1][C]=[O]),$([CX4][#7](~[*^2])~[*^2]);!$([CH0]);!$([CX4][#7X3H0][#6]=[O]);!$([CX4][nX3H0][c]=[O])]	63.9
23	[$([CX4][S][P]),$([CX4]1[N]([C][C]2)[C][C][C]2[C]1);!$([CH0]);!$([C][S]=[O])]	66.9
24	[$([CX4][#6^2]);!$([CH0]);!$([CX4][C](=[O])[NX3])]	66.4
25	[$([CX4][S](=[O])=[O]),$([CX4][NX3][#16X4](=[OX1])(=[OX1])),$([CX4][#6^2](=[#8])[#6^2]);!$([CH0]);!$([CX4][S](=[OX1])(=[OX1])[c])]	72.3
26	[CX4;CH1,CH2,$([CH3][#7X3,#6^2]),$([CH3][O][C]=[O]),$([CH3][S]=[O])]	75.9
27	[CX4H3]	89.6
28	[$([CX3H2]);$([C]=[*^2]-[*^2])]	40.1
29	[$([CX3H1]);$([C]=[C^2]-[C^2,c]);!$([C](-[*^2,a])=[*^2]-[*^2])]	52.4
30	[$([ch1r5]:[#8]),$([ch1r5](:[c]):[nH1])]	52.9
31	[$([ch1r5](:[nX3]):[nX2])]	57.9
32	[$([ch1]1:[c](-[N]([C^3])[C^3]):[c]:[c]:[c]:[c]1),$([ch1]1:[c]:[c]:[c]:[c](-[N]([C^3])[C^3]):[c]:[c]1),$([ch1]1:[c](-[NH1][C^3]):[c]:[c]:[c]:[c]1),$([ch1]1:[c]:[c]:[c](-[NH1][C^3]):[c]:[c]1),$([ch1]1:[c](-[NH2]):[c]:[c]:[c]:[c]1),$([ch1]1:[c]:[c]:[c](-[NH2]):[c]:[c]1)]	59.5
33	[$([CX3]);$([CX3]=[CX3]);!$([CH0]);!$([CX3](-[C^2])=[CX3]);!$([CX3]=[CX3]-[C^2])]	65.6
34	[$([ch1]1:[c]:[c]:[cr6]([#7]~[#6^2]):[c]:[c]1),$([ch1r6]:[cr5]:[nX3r5]);!$([c]1:[c]:[c]:[c](~[#7][#6]=[O]):[c]:[c]1);!$([ch1r6]:[cr5]:[nr5]:[c](=[O])[nr5][nr5]);!$([ch1r6]:[cr5]:[nr5]:[c](=[O])[cr6])]	68.2
35	[$([ch1r5]:[#16X2]:[c]),$([ch1r5]:[c]:[nX3])]	69.4
36	[$([ch1]);!$([c]1:[c]:[c]:[c](-[O]-[C^2]~[O]):[c]:[c]1);$([c]1:[c]:[c]:[c](~[#7H][#6]=[O]):[c]:[c]1),$([c]1:[c]:[c]:[c](-[O,SX2]):[c]:[c]1),$([ch1r6]:[cr6]:[cr5]:[nX3r5]),$([ch1r6]:[cr5]:[cr5]:[nX3r5,oX2r5,sX2r5])]	74.1
37	[$([ch1]);$([ch1]1:[c](~[#7X2]~[#6^2]):[c]:[c]:[c]:[c]1),$([ch1r6]:[cr6]-[O,SX2]),$([ch1r6]:[cr5]:[sX2r5]),$([ch1r6]:[cr6]:[cr5]:[or5,sX2r5]),$([ch1r6]:[cr6]:[cr6]:[cr5]:[nX3r5]);!$([ch1r6]:[cr6][#8][#6^2]~[O])]	77.2

Supplementary Material Table S4. SMARTS rules for SMARTCyp 3.0

38	[$([ch1r5]);!$([ch1r5]:[nX2]:[#16X2])]	78.1
39	[$([ch1]1:[c]([#6^2,#6^1,#8]):[c]:[c]:[c]:[c]1),$([ch1]1:[c]:[c]:[c]([#6^2,#6^1,#8]):[c]:[c]1),$([ch1]1:[c]([NX3](~[O])~[O]):[c]:[c]:[c]:[c]1),$([ch1]1:[c]:[c]:[c]([NX3](~[O])~[O]):[c]:[c]1),$([C^2h1](-[C^2])=[C^2]),$([ch1r6]:[cr5]:[or5]),$([ch1r6]:[cr6]:[cr6]:[cr5]:[or5,sX2r5])]	80.8
40	[$([ch1]1:[c]([F,Cl,I,Br]):[c]:[c]:[c]:[c]1),$([ch1]1:[c]:[c]:[c]([F,Cl,I,Br]):[c]:[c]1),$([ch1]1:[c]:[c]([NX3](~[O])~[O]):[c]:[c]:[c]1),$([ch1]1:[c]:[c]:[c]([S]=O):[c]:[c]1)]	84.1
41	[$([#6^2h1]);!$([ch1r6]:[nr6]);!$([ch1]:[c]-[S](=O)(=O)-[NX3])]	86.3
42	[$([ch1r6]:[nr6]),$([ch1]:[c]-[S](=O)(=O)-[NX3])]	92.0
43	[$([N^3H0]);!$([N^3][*^2]);!$([NX3][#16X4](=[OX1])(=[OX1]));!$([N]1([CX4])[C][C][N]([*^3])[C][C]1);!$([N]1([CX4])[C][C][NH1][C][C]1);!$([NR1]1([CX4])[C][C][C][C][C]1);!$([NR2]1([CX4])[C]2[C][C][C][C]1[C][C]2);!$([NR2r6](@[C])(@[C])@[C])]	42.6
44	[$([N^3H0]);$([N]1([CX4])[C][C][C,N][C][C]1);!$([NR2r6](@[C])(@[C])@[C])]	50.3
45	[$([N^2,N^3]);$([H1,H2]);!$([NX3][#16X4](=[OX1])(=[OX1]));!$([N^2H1,N^3H1]([*^2])[*^2]);!$([N^2][C]=[O])]	54.1
46	[$([NX3H0]([#6^2]1)[#6^2]=[#6^2][#6^3][#6^2]=1)]	61.9
47	[$([N]([C^3])=[C^2])]	61.9
48	[$([NX3H0]);$([NX3][*^2]);!$([N^3]([*^2])[*^2]);!$([N^2]([*^2])[*^2]);!$([NX3][#16X4](=[OX1])(=[OX1]));!$([N;+1]);!$([N^2][C]=[O])]	63.9
49	[$([N^3H0]);$([NR2r6](@[C])(@[C])@[C])]	68.4
50	[$([NX3H1]([*^2])[*^2]);!$([nr5H0]);!$([NX3H1][C]=[O])]	72.0
51	[$([nD2r6]),$([N](-[#6^2])=[#6^2]);!$([nr5H0]);!$([nr6]([#6])([c])[c])]	75.6
52	[$([#7]);$([#7X3H0]([*^2,c])[*^2,c]),$([#7^2][#6]=[O]);!$([N;+1]);!$([nH0r])]	89.6
53	[nr5H0]	92.1
54	[$([NX3]);$([NX3][#16X4](=[OX1])(=[OX1]))]	94.4
55	[$([NX3H1]([#6^2]1)[#6^2]=[#6^2][#6^3][#6^2]=1)]	10.4
56	[$([O]);$([OH1][c]1[c][c][c]([NH][C]~[O,#7])[c][c]1),$([OH][c]1[c][c]([NHx0][C]=[O])[c][c][c]1),$([OH][c]1[c]([NHx0][C]=[O])[c][c][c][c]1),$([OH][c]1[c][c][c]([NR6,cr5])[c][c]1),$([OH][c]1[c][c]([sr5])[c][c][c]1),$([OH][c]1[c]([Cl])[c][c]([N][#6r6])[c][c]1),$([OH][c]1[c]([C]#[C])[c][c]([N][#6r6])[c][c]1),$([OH1][c]1[c][c][c]([O])[c][c]1);!$([OH1][c]1[c][c]([S])[c]([O])[c][c]1)]	15.0
57	[CH1;$([C]=1[O][C][C][C]=1)]	49.0
58	[CH1;$([C]=1[C][C][O][C]=1)]	41.9
59	[CH1;$([C]1[O][C]=[C][C]1)]	73.0
60	[CH1;$([C]1[C]=[C][O][C]1)]	46.2
61	[$([ch1]);$([c]1[c]([Cl])[c]([Cl])[c]([N]2[C][C][N]([C])[C][C]2)[c][c]1),$([c]1[c]([N]2[C][C][N]([C])[C][C]2)[c]([Cl])[c]([Cl])[c][c]1)]	67.7
62	[$([ch1]);$([c]1[c][ch1][c]([N]2[C][C][N]([C])[C][C]2)[ch1][c]1),$([c]1[c]([N]2[C][C][N]([C])[C][C]2)[ch1][c][c][c]1)]	37.6
63	[$([CH2]);$([C]([Cl])[c]1[c][c]([o][c]2[c]1[c][c][c]([c]2)[O][C][C])=[O])]	62.6
64	[$([CH1]);$([C]1([CR0])[N]([CR0])[C][C][C]1)]	36.9
65	[$([CH2]);$([CH2]1[CH2][CH2][CH2]([CR0])[CH2][CH2]1)]	74.4
66	[$([CH2]);$([C]1[C][C]([O][c]2[c][c][c][c][c]2)[C][C]1)]	72.2
67	[$([ch1r5]);$([c]1[nH1][c]2[c]([cR1]1)[c][c][c][c]2)]	50.6
68	[$([ch1r5]);$([c]1[c][nH1][c]2[c]1[c][c][c][c]2)]	59.3
69	[$([s]);$([s]1[c](=[O])[n][c](=[O])[c]([C])1)]	18.0
70	[$([n]);$([n](C)1[c](=[O])[n][c][n]1)]	136.5
71	[$([nH0]);$([n]1[c](=[O])[n][c][n]1)]	103.7