Supporting Information for

SMARTCyp – a 2D-method for Prediction of Cytochrome P450 Mediated Drug Metabolism

Patrik Rydberg, † David E. Gloriam, † Jed Zaretzki, † Curt Breneman, † and Lars Olsen† *

Biostructural Research, Department of Medicinal Chemistry, Faculty of Pharmaceutical Sciences,
University of Copenhagen, Universitetsparken 2, DK-2100 Copenhagen, Denmark. Center for
Biotechnology and Interdisciplinary Studies, Department of Chemistry and Chemical Biology,
Rensselaer Polytechnic Institute, Troy, NY, USA

Table S1. SMARTS rules and corresponding energies.

Table S2. Structures and information on all QM calculations for all SMARTS rules.

Table S1. SMARTS rules and corresponding energies (kJ/mol).

Sulphur atoms	
[SX2H1]	41.5
[\$([SX2H0]);!\$([S][*^2]);!\$([S][CX4H0])]	26.3
[\$([SX2H0][*^2]);!\$([S](~[^2])[^2]);!\$([S][CX4H0])]	34.4
[\$([S][*D4H0]);\$([SX2H0])]	44.4
[\$([SX2H0]([*^2])[*^2]);!\$([S][CX4H0])]	46.9
[\$\(\frac{1}{2}\)\(\f	70.0
[\$([#16X3](=[OX1]));\$([#16]);!\$([#16X3](=[OX1])[#6^2](~[#7^2]))]	30.4
	46.9
[\$([#16X3](=[OX1]));\$([#16]);\$([#16X3](=[OX1])[#6^2](~[#7^2]))]	40.9
Aldohydo sarhon atoms	
Aldehyde carbon atoms	40.2
[\$([CX3H1](=O)[#6])]	40.2
Dheamhan with daubh handad adaban	
Phosphor with double bonded sulphur	42.2
[\$([PX4]);\$([P]=[S])]	13.3
sp3 hybridized carbon atoms	22.4
[\$([CX4]([#6^2])([#6^2]);!\$([CH0])]	33.1
[\$([CX4][N]);!\$([CH0]);!\$([C][N]([*^2])[*^2]);!\$([C][N]=[#6X3]);!\$([CX4][NX3][C]=[O])]	39.8
[!\$([CH0]);\$([CX4]([#6^2])[#6^2]),\$([CX4][#7]=[#6X3]),\$([CX4]([#8])[#8]);!\$([CX4]([#8])[#8][C]=[O])]	48.6
[\$([CX4][S]);!\$([CH0]);!\$([C][S]=[O])]	57.7
[\$([CX4][#6^2]~[#8]),\$([CX4][#6^2](~[#7])~[#7]),\$([CX4][#6^1]),\$([CX4][C^2]=[C^2]-[#6^2]);	60.0
!\$([CH0]);!\$([CX4][C](=[O])[NX3]);!\$([CX4][#6^2](=[#8])-[#8]);!\$([CX4][C^2]([C^2])=[C^2]-[#6^2])] [\$([CX4][O]);!\$([CH0]);!\$([C][O][C]=[O]);!\$([CX4]1[O][C]1)]	62.2
[\$([CX4][#6^2]);!\$([CH0]);!\$([CX4][C](=[O])[NX3])]	66.7
[\$([CX4][#7](~[*^2]);!\$([CH0])]	69.1
	69.5
[\$([CX4][S](=[O])=[O]);!\$([CH0])] [CX4;CH1,CH2;!\$([CX4][NX3H1][C]=[O])]	77.7
	89.6
[CX4H3;!\$([CX4][NX3H1][C]=[O])]	
[\$([CX4][NX3H1][C]=[O]);!\$([CH0])]	94.6
sp2 hybridized carbon atoms	40.4
[\$([CX3H2]);\$([C]=[*^2]-[*^2])]	40.1
[\$([CX3H1]);\$([C]=[*^2]-[*^2]);!\$([C](-[*^2])=[*^2]-[*^2])]	52.4
[\$([cH1]);\$([c](:[#7]):[#7])]	55.8
[\$([ch1]);!\$([ch1]1:[c](-[N^3]-[*^2]):[c]:[c]:[c]:[c]1);!\$([c]1:[c]:[c]:[c]:[c](-[N^3]-[*^2]):[c]:[c]1);	59.5
\$([ch1]1:[c](-[N^3]):[c]:[c]:[c]:[c]1),\$([c]1:[c]:[c]:[c](-[N^3]):[c]:[c]1)] [\$([CX3]);\$([CX3]=[CX3]);!\$([CH0]);!\$([CX3](-[*^2])=[CX3]);!\$([CX3]=[CX3]-[*^2])]	65.6
	67.1
[\$([cH1]);\$([c]:[#16])] [\$([cH1]);\$([c]:[-1:[c]:[c]:[c]:[c]:[c]:[c]:[c]:[c]:[c]:[c]	69.2
[\$([ch1]);\$([c]1:[c]:[c]:[c](~[#7X2H0]~[c^2,C^2]):[c]:[c]1),\$([c]1:[c]:[c]:[c](~[#7X3H1]~[c^2,C^2]):[c]:[c]1); !\$([c]1:[c]:[c]:[c](-[NH]-[C]=[O]):[c]:[c]1)]	09.2
[\$([ch1]);!\$([c]1:[c]:[c]:[c]-[O]-[C]=[O]):[c]:[c]1:[c]:[c]:[c]:[c]-[NH]-[C]=[O]):[c]:[c]1),	75.3
fattan-110-4-trail-1-trail-fatt fatt fatt fattifat-fat-fattfattfattfattfatt fund fattfattfattfatt	, 5.5

\$([c]1:[c]:[c]:[c](-[O,SX2]):[c]:[c]1)]	
[\$([cH1]);\$([c]1:[c]:[#16]:[c]1)]	76.7
[\$([ch1]);\$([ch1]1:[c](~[#7X2H0]~[c^2,C^2]):[c]:[c]:[c]:[c]1),\$([ch1]1:[c](~[#7X3H1]~[c^2,C^2]):[c]:[c]:[c]1) ,\$([ch1]:[c]-[O,SX2]);!\$([ch1]:[c]-[O]-[C]=[O])]	77.3
[\$([CX3H1,cX3H1]);!\$([ch1]1:[c]:[c]:[c]:[c]1);!\$([ch1]1:[c]:[#7]:[c]:[c]1);!\$([ch1]1:[c]:[#7]:[c]:[c]1); !\$([ch1]:[#7])]	82.3
[\$([ch1]1:[c]:[c]:[c]:[c]1),\$([ch1]1:[c]:[#7]:[c]:[c]1),\$([ch1]1:[c]:[#7]:[c]:[c]1),\$([ch1]:[#7])]	89.6
Nitrogen atoms	
[\$([N^3H0]);!\$([N^3][*^2])]	41.0
[\$([N^3]);\$([H1,H2])]	54.1
[\$([N]([#6^2]1)[#6^2]=[#6^2][#6^3][#6^2]=1)]	61.9
[\$([N^3H0]);\$([N^3][*^2]);!\$([N^3]([*^2])[*^2])]	63.9
[\$([nr6]),\$([N^2]=[C])]	75.6
[\$([N]);\$([NX3H0]([*^2])[*^2]),\$([N^2H1][C]=[O])]	89.6
[nr5H0]	92.1

Table S2. Structures and information on all QM calculations for all SMARTS rules. Spin state in parenthesis after compound name (d is doublet, q is quartet). If there are multiple possible sites the computed site is marked with the number 1 unless otherwise written.

Sulphur Oxidations		
[SX2H1]		Energy (kJ/mol)
Methylsulfane (d)	——SH	41.5
[\$([SX2H0]);!\$([S][*^2]);!\$([S]	[CX4H0])]	Energy (kJ/mol)
Methylethylsulfane (d)	s	28.7ª
Dimethylsulfane (d)	S	23.8 ^b
	Average:	26.3
	Standard deviation:	3.5
[\$([SX2H0][*^2]);!\$([S](~[^2])	[^2]);!\$([S][CX4H0])]	Energy (kJ/mol)
Phenylmethylsulfane (d)	S	33.8 ^a
Vinylmethylsulfane (d)	S	35.0 ^a
	Average:	34.4
	Standard deviation:	0.9
[\$([S][*D4H0]);\$([SX2H0])]		Energy (kJ/mol)
Tertbutylmethylsulfane (d)	S	44.4 ^a
[\$([SX2H0]([*^2])[*^2]);!\$([S][CX4H0])]		Energy (kJ/mol)
Divinylsulfane (d)	S	45.1 ^a

Dibenzo[b,f]thiepine (d)	S	48.8 ^a
	Average:	46.9
	Standard deviation:	2.6
[sX2r5]		Energy (kJ/mol)
Thiophene (d)	S	69.1
Thiazole (d)	S	74.0
Isothiazole (d)	SN	66.9
	Average:	70.0
	Standard deviation:	3.6
[\$([#16X3](=[OX1]));\$([#16]);!	\$([#16X3](=[OX1])[#6^2](~[#7^2]))]	Energy (kJ/mol)
Dimethylsulfoxide (d)		27.3 ^b
Methylsulfinylethane (d)	S	25.3ª
Methylsulfinylisopropane (d)	S	33.6 ^a

Methylsulfinyltertbutane (d)	S S	34.8 ^a
Methylsulfinylbenzene (d)	S S	31.3 ^a
Methylsulfinylethene (d)	S	27.4ª
Divinylsulfoxide (d)	0 s //	32.8 ^a
	Average:	30.4
	Standard deviation:	3.7
[\$([#16X3](=[OX1]));\$([#16]);	6([#16X3](=[OX1])[#6^2](~[#7^2]))]	Energy (kJ/mol)
2-(ethylsulfinyl)-1H- benzo[d]imidazole (d)	S S	46.9 ^a
	Aldehyde oxidation (of C=O carbon)	
[\$([CX3H1](=O)[#6])]		Energy (kJ/mol)
Propionaldehyde (q)	○	37.0
Acrylaldehyde (q)	№ 0	42.5
Benzaldehyde (q)	0	41.1
	Average:	40.2
	Standard deviation:	2.8

Phosphor with double bonded sulphur (desulphurization)		
[\$([PX4]);\$([P]=[S])]		Energy (kJ/mol)
O,O,O-trimethyl phosphorothioate (d)	S / O / P - O / O / O / O / O / O / O / O / O / O	13.3°
	Nitrogen oxidations	
[\$([N^3H0]);!\$([N^3][*^2])]		Energy (kJ/mol)
Trimethylamine (d)	N	39.8 ^b
<i>N</i> -ethyl-dimethylamine (d)	N	42.3
	Average:	41.0
	Standard deviation:	1.7
[\$([N^3]);\$([H1,H2])]		Energy (kJ/mol)
Propane-2-amine (d)	NH ₂	51.6 ^a
Propene-2-amine (d)	NH ₂	55.9ª
Aniline (d)	NH ₂	56.3
<i>N</i> -methylaniline (d)	N _H	50.0
1,3-dimethylguanidine (d)	¹NH N N H H	56.9ª

	Average:	54.1
	Standard deviation:	3.1
[\$([N]([#6^2]1)[#6^2]=[#6^2][#6^3][#6^2]=1)]	Energy (kJ/mol)
1-methyl-4-hydropyridine (d)	N	61.9
[\$([N^3H0]);\$([N^3][*^2]);!\$([N^3]([*^2])[*^2])]	Energy (kJ/mol)
<i>N,N</i> -dimethylaniline (d)	N N	62.5
<i>N,N</i> -dimethylethenamine (d)	N.	65.3
	Average:	63.9
	Standard deviation:	2.0
[\$([nr6]),\$([N^2]=[C])]		Energy (kJ/mol)
Quinoline (d)	N	72.7
5-fluoro-4-methylpyrimidine (d)	FN_	83.5
Pyridine (d)	N N	71.0

Pyridazine (d)	N N	75.0
	Average:	75.6
	Standard deviation:	5.5
[\$([N]);\$([NX3H0]([*^2])[*^2]),\$([N^2H1][C]=[O])]	Energy (kJ/mol)
N,N-divinylmethaneamine (d)	N N	93.1
N-methylformamide (d)	0 1	86.2
	Average:	89.6
	Standard deviation:	4.9
[nr5H0]		Energy (kJ/mol)
Imidazole (d)	HN N 1	85.8
Thiazole (d)	S	88.4
Isothiazole (d)	S _N	102.1
	Average:	92.1
	Standard deviation:	9.4
	sp ³ carbons	
[\$([CX4]([#6^2])([#6^2])[#6^2]);!\$([CH0])]	Energy (kJ/mol)
3-vinylpenta-1,4-diene (q)		33.1
[\$([CX4][N]);!\$([CH0]);!\$([C][!\$([C][N]=[#6X3]);!\$([CX4][NX		Energy (kJ/mol)

1-ethyl-4-methylpiperidine (q) (1)		35.9
1-ethyl-4-methylpiperidine (q) (2)	1	37.5
	N	
1,4-diethylpiperazine (q) (1)	1	44.5
1,4-diethylpiperazine (q) (2)	2	39.6
Trimethylamine (q)		38.4 ^b
N-ethyl –dimethylamine (q) (1)	N—2	38.8
N-ethyl –dimethylamine (q) (2)	N—2	37.1
N,N-dimethyl aniline (q)	N N	40.7 ^d
N-methyl aniline (q)	- N H	44.0 ^d
Dimethylpropenamine (q)	1_N	32.9

N-ethyl-methylamine (q) (1)	1	41.9
	HN—²	
N-ethyl-methylamine (q) (2)	1	39.5
	HN—²	
		to d
Dimethylamine (q)	, N	42.6 ^d
Ethanamine (q) (1)	H_2N	43.5
	Average:	39.8
	Standard deviation:	3.3
[!\$([CH0]);\$([CX4]([#6^2])[#6 \$([CX4]([#8])[#8]);!\$([CX4]([#		Energy (kJ/mol)
<i>N</i> -ethylideneethaneamine (q)	N	48.6
Penta-1,4-diene (q)		48.4
1,3-dioxolane (q)	0 1	48.7
	Average:	48.6
	Standard deviation:	0.1
[\$([CX4][S]);!\$([CH0]);!\$([C][S]=[O])]	Energy (kJ/mol)
Dimethylsulfane (q)	S	58.1 ^b
Methyl(phenyl)sulfane (q)	S	57.3 ^d
	Average:	57.7
	Standard deviation:	0.5
\$([CX4][C^2]=[C^2]-[#6^2]);!	#6^2](~[#7])~[#7]),\$([CX4][#6^1]), \$([CH0]);!\$([CX4][C](=[O])[NX3]); ([CX4][C^2]([C^2])=[C^2]-[#6^2])]	Energy (kJ/mol)
Pent-2-yne (q)	1	62.3

Penta-1,3-diene (q)		57.7
2-methyl-1 <i>H</i> -imidazole (q)	N NH	59.1
Propionaldehyde (q)	0 1 1	61.3
Prop-1-en-2-ol (q)	ОН	59.4 ^d
	Average:	60.0
	Standard deviation:	1.8
[\$([CX4][O]);!\$([CH0]);!\$([C][O][C]=[O]);!\$([CX4]1[O][C]1)]	Energy (kJ/mol)
Isopropanol (q)	OH 1	55.7
Dimethylether (q)	_0_	63.1 ^d
Anisole (q)		68.0 ^d
	Average:	62.2
	Standard deviation:	6.2
[\$([CX4][#6^2]);!\$([CH0]);!\$	[[CX4][C](=[O])[NX3])]	Energy (kJ/mol)
Propene (q)		67.0 ^d
2-fluoro-prop-1-ene (q)	F	69.1 ^d
Toluene (q)		67.3 ^d
Ethylbenzene (q)		64.6 ^d
1-methylethylbenzene (q)		69.7 ^d
Para-xylene (q)		65.9 ^d

1-methyl-4-nitro-benzene (q)	O N+	62.6 ^d
Propionic acid (q)	O 1 OH	67.2
	Average:	66.7
	Standard deviation:	2.3
[\$([CX4][#7](~[*^2])~[*^2]);!\$	G([CH0])]	Energy (kJ/mol)
1-methyl-1 <i>H</i> -tetrazole (q)	N=N- N=N	65.7
Caffeine (q) (1)		75.8 ^e
Caffeine (q) (2)		70.1 ^e
Caffeine (q) (3)		64.7 ^e
	Average:	69.1
	Standard deviation:	5.0
[\$([CX4][S](=[O])=[O]);!\$([CH	10])]	Energy (kJ/mol)
Methylethylsulfone (q)		69.5
[CX4;CH1,CH2;!\$([CX4][NX3	BH1][C]=[O])]	Energy (kJ/mol)

Ethyl propionate (q)	0 1	81.9
N-methylacetamide (q)		81.7
N,N-diethylacetamide (q)	1 N	79.6
2,3-diethyloxirane (q)	0	73.0
Dimethylsulphoxide (q)	O Ś	77.0 ^b
Propane (q)	1	77.0 ^d
Isobutane (q)	1	74.7 ^d
Fluoroethane (q)		76.5 ^d
	Average:	77.7
	Standard deviation:	2.3
[CX4H3;!\$([CX4][NX3H1][C]=[O])]		Energy (kJ/mol)
Propane (q)	1	88.4 ^d
Fluoroethane (q)	₁ ~ F	93.0 ^d
Ethylbenzene (q)	1	87.4 ^d
Average:		89.6
Standard deviation:		3.0
[\$([CX4][NX3H1][C]=[O]);!\$([CH0])]		Energy (kJ/mol)

N-ethylacetamide (q)	1 NH	94.6
[\$([CX3H2]);\$([C]=[*^2]-[*	sp ² carbons	Energy (kJ/mol)
Penta-1,3-diene (q)	1	44.0
Hepta-1,3,5-triene (q)		29.3
Acrylaldehyde (q)	0	42.2
Styrene (q)		44.9
	Average:	40.1
	Standard deviation:	7.3
[\$([CX3H1]);\$([C]=[*^2]-[*	*^2]);!\$([C](-[*^2])=[*^2]-[*^2])]	Energy (kJ/mol)
Penta-1,3-diene (q)		52.5
Methylstyrene (q)		52.3
	Average:	52.4
	Standard deviation:	0.1
[\$([cH1]);\$([c](:[#7]):[#7])]		Energy (kJ/mol)

Caffeine (d)		54.4 ^e
Theophylline (d)		57.1 ^e
	Average:	55.8
	Standard deviation:	1.9
[\$([ch1]);!\$([ch1]1:[c](-[N^3]-[*^2]):[c]:[c]:[c]:[c]1); !\$([c]1:[c]:[c]:[c](-[N^3]-[*^2]):[c]:[c]1); \$([ch1]1:[c](-[N^3]):[c]:[c]:[c]1),\$([c]1:[c]:[c]:[c](-[N^3]):[c]:[c]1)]		Energy (kJ/mol)
<i>N,N</i> -dimethylaniline (d) (1)	2 1 N	58.8 ^f
N,N-dimethylaniline (d) (2)	2 1 N	60.2 ^f
	Average:	59.5
	Standard deviation:	0.9
[\$([CX3]);\$([CX3]=[CX3]);!\$([CH0]);!\$([CX3](-[*^2])=[CX3]); !\$([CX3]=[CX3]-[*^2])]		Energy (kJ/mol)
Ethene (q)	=	67.5 ^f
2-butene (d)		63.6 ^f
Average:		65.6
Standard deviation:		2.7
[\$([cH1]);\$([c]:[#16])]		Energy (kJ/mol)

Thiophene (d)	S 1	67.1
[\$([ch1]);\$([c]1:[c]:[c](~[#7X2H0]~[c^2,C^2]):[c]:[c]1), \$([c]1:[c]:[c](~[#7X3H1]~[c^2,C^2]):[c]:[c]1); !\$([c]1:[c]:[c](-[NH]-[C]=[O]):[c]:[c]1)]		Energy (kJ/mol)
Diphenylamine (d)	HN 1	64.9
Tacrine (d)	NH ₂	68.6 ^e
2,6-dichloro-N-o-tolylaniline (d) (1)	CI H N 2	69.3 ^f
2,6-dichloro-N-o-tolylaniline (d) (2)	CI H N 2	73.9 ^f
	Average:	69.2
Standard deviation:		3.7
[\$([ch1]);!\$([c]1:[c]:[c]:[c](-[O]-[C]=[O]):[c]:[c]1); \$([c]1:[c]:[c](-[NH]-[C]=[O]):[c]:[c]1), \$([c]1:[c]:[c](-[O,SX2]):[c]:[c]1)]		Energy (kJ/mol)
Anisole (d)	1	74.4

Methyl(phenyl)sulfane (d)	S	74.3
Phenol (d)	OH 1	75.9
N-phenylacetamide (d)	HN O	76.7
Average:		75.3
Standard deviation:		1.2
[\$([cH1]);\$([c]1:[c]:[#16]:[c]:[c	[]1)]	Energy (kJ/mol)
Thiophene (d)	S1	76.7
[\$([ch1]);\$([ch1]1:[c](~[#7X2H0]~[c^2,C^2]):[c]:[c]:[c]:[c]1), \$([ch1]1:[c](~[#7X3H1]~[c^2,C^2]):[c]:[c]:[c]:[c]1), \$([ch1]:[c]-[O,SX2]);!\$([ch1]:[c]-[O]-[C]=[O])]		Energy (kJ/mol)
Anisole (d)		78.2
Methyl(phenyl)sulfane (d)	S	77.2

Tacrine (d)	NH ₂	76.5 ^e
7-ethoxyresorufin (d)		77.3 ^e
	Average:	77.3
	Standard deviation:	0.7
[\$([CX3H1,cX3H1]);!\$([ch1]1: !\$([ch1]1:[c]:[#7]:[c]:[c]:[c]1);! !\$([ch1]:[#7])]		Energy (kJ/mol)
Penta-1,3-diene (d)	_	81.8
Hepta-1,3,5-triene (q)		77.0
Anisole (d)		85.8
Methyl(phenyl)sulfane (d)	S	87.1
Tacrine (d)	NH ₂	82.0 ^e
Benzene (d)		87.3 ^f

Warfarin (d) (1)	HO 1 3	83.4 ^f
Warfarin (d) (2)	O 4 3 2	80.8 ^f
Warfarin (d) (3)	HO 1 3	81.0 ^f
Warfarin (d) (4)	O 4 3 2	78.7 ^f
N,N-dimethylaniline (d)		86.7 ^f
Nitrobenzene (q) (1)		78.5 ^f
Nitrobenzene (d) (2)	O 1 2 3	84.5 ^f

Nitrobenzene (q) (3)		79.2 ^f
2-fluoro-4-methylbiphenyl (d)	F 1	80.7 ^f
	Average:	82.3
	Standard deviation:	3.4
[\$([ch1]1:[c]:[c]:[#7]:[c]:[c]1),\$ \$([ch1]1:[c]:[#7]:[c]:[c]1),\$([ch	6([ch1]1:[c]:[#7]:[c]:[c]:[c]1), 1]:[#7])]	Energy (kJ/mol)
Pyridine (q) (1)	2 N	90.6
Pyridine (d) (2)	2 N	85.5
Pyridine (q) (3)	2 N 3 N	92.6
	Average:	89.6
Standard deviation:		3.6

^a Transition state geometry from scan crossing point.

^b Taken from the publication: P. Rydberg, U. Ryde, L. Olsen, *J. Chem. Theory Comput.* 2008, 4, 1369-1377

^c Energy taken from the calculation with the small basis set relative to reactant complex.

^d Taken from the publication: L. Olsen, P. Rydberg, TH. Rod, U. Ryde, *J. Med. Chem.* 2006, 49, 6489-6499.

^e Taken from the publication: P. Rydberg, P. Vasanthanathan, C. Oostenbrink, L. Olsen, *ChemMedChem* 2009, 4, 2070-2079.

^f Taken from the publication: P. Rydberg, U. Ryde, L. Olsen, *J. Phys. Chem. A* 2008, 112, 13058-13065.