Supporting Information

FAME 3: Predicting the Sites of Metabolism in Synthetic Compounds and Natural Products for Phase 1 and Phase 2 Metabolic Enzymes

Martin Šícho,[1,2] Conrad Stork,[1] Angelica Mazzolari,[3] Christina de Bruyn Kops,[1] Alessandro Pedretti,[3] Bernard Testa,[4] Giulio Vistoli,[3] Daniel Svozil,[2] and Johannes Kirchmair[1,5,6]*

[1] Universität Hamburg, Faculty of Mathematics, Informatics and Natural Sciences, Department of Informatics, Center for Bioinformatics, 20146 Hamburg, Germany

[2] University of Chemistry and Technology Prague, Faculty of Chemical Technology, Department of Informatics and Chemistry, CZ-OPENSCREEN: National Infrastructure for Chemical Biology, 166 28 Prague 6, Czech Republic

[3] Università degli Studi di Milano, Facoltà di Scienze del Farmaco, Dipartimento di Scienze Farmaceutiche "Pietro Pratesi", I- 20133 Milan, Italy

[4] Emeritus Professor, University of Lausanne, 1015 Lausanne, Switzerland

[5] University of Bergen, Department of Chemistry, N-5020 Bergen, Norway

[6] University of Bergen, Computational Biology Unit (CBU), N-5020 Bergen, Norway

[*]J. Kirchmair. E-mail: kirchmair@zbh.uni-hamburg.de. Tel.: +49 40 42838 7303.

Table S1: Reaction Subclasses of Phase 1 Metabolism Annotated in the MetaQSAR Database.

Reaction subclass	Annotation count[a]
Oxidation of aryl compounds to epoxides, phenols or other metabolites	1013
Hydroxylation (or other oxidations) of Csp3 carrying a heteroatom (N, O, S, halo) (including subsequent dealkylation, deamination or dehalogenation)	907
Hydroxylation (or other oxidations) of isolated Csp3	423
Hydroxylation (or other oxidations) of C alpha to an unsaturated system (>C=C<, >C=O, -C±N, aryl)	325
Dehydrogenation of >CH-CH< to >C=C<, of >CH-N< to >C=N- (incl. >C=N+<)	207
Hydrogenation of -CHO to -CH2OH and of >C=O to >CHOH	153
Hydrolysis of alkyl esters	117
Oxygenation of >C=C< bonds to epoxides or other metabolites	113
Other oxidations of phenols and amines (dimerization, quinone-like metabolites, etc)	106
Oxygenation of sulfides to sulfoxides, and of sulfoxides to sulfones	106
Dehydrogenation of -CH2OH groups to -CHO and of >CHOH to >C=O	103
Oxidation of azaarenes to lactams or other metabolites	101
Hydrolysis of anionic and cationic esters	100
Oxidation of cresols and analogs to quinonemethides	85
Hydroxylation of amines to hydroxylamines or intermediates	75
Oxidation of diphenols to quinones	69
Oxidation of tertiary alkylamines and heterocyclic amines to N-oxides or other metabolites	68
Hydrolysis of esters of inorganic acids (nitrates, nitrites, sulfates, sulfamates, phosphates, phosphonates, etc)	67
Other C reductions, e.g. of >C=C< to -CH2-CH2-	66
Hydrolysis of anilides [aryl-N-CO-C~], hydrazides [-CO-NHN<], sulfonamides [-SO2NH-], linear imides [-CO-NR-CO-] and urea derivatives [>N-CO-N<]	61
Hydration of arene and alkene oxides	55
Hydrolytic cleavage of C=N bonds (in imines, hydrazones, imidates, amidines, oximes, oximines, isocyanates, etc), and of C±N bonds (nitriles). Double bond hydration	54
Hydrolysis of alkyl and aryl amides [alkyl-CO-N< and aryl-CO-N<]	49
Oxidation of tertiary arylamines, azaarenes and azo compounds to N-oxides oxides or other metabolites	49
Hydrolysis of linear and cyclic carbamates (>N-CO-OR') and carbonates (RO-CO-OR')	40

Reduction of nitro compounds to nitroso compounds	38
Reduction of quinones and analogs	33
Other N-reductions (e.g. azo compounds to hydrazines, hydrazines to amines, reductive N-rings opening)	31
Oxidation of -CHO to -COOH	30
Hydrolytic dehalogenations	30
Oxidation of amino- and amido-phenols to quinoneimines or quinoneimides, resp.	29
Hydrolysis of lactams, cyclic imides and cyclic ureides	28
Cyclizations by intramolecular nucleophilic substitution (with elimination of amine, phenol, halide or H2O)	23
Bond order increases by elimination of H2O or RSH or other Nu-H	21
Reductive dehalogenations	21
Hydrolysis of aryl esters	20
Oxidative desulfurations of >C=S to ketones, and of -P=S to -P=O groups	19
Hydrolysis of thioesters (RCO-SR' and RCS-SR') and thiolactones	19
Reduction of hydroxylamines and hydroxylamides (incl. spontaneous dismutation)	18
Hydrolysis of glucuronides and other glycosides (including N- and S-glycosides)	18
Reduction of nitroso compounds and oximes to hydroxylamines	18
Hydrolytic opening of other ring systems (1,2-oxazoles, etc)	17
Other N-oxidations (1,4-dihydropyridines, etc)	17
Other Csp3 oxidations (organometallic dealkylation, C-C cleavage, etc)	17
Any other non-redox, non-conjugation reaction	15
Oxygenation of thiones (>C=S) or thioamides to sulfines, and of sulfines to sulfenes	15
Oxygenation of -C-C±C-H and -C-C±C-C- bonds	14
Reduction of N-oxides	12
Hydroxylation of amides to hydroxylamides	12
Reversible hydrolytic opening of lactone rings	11
Other ether hydrolyses (benzhydryl ethers, acetals, etc)	10
Oxidation of thiols to sulfenic acids or disulfides	10
Reduction of sulfoxides to sulfides	6

Hydrolysis of peptide bonds	5
Hydrolysis of linear Mannich bases (N-, O- and S-Mannich bases) and cyclic Mannich bases (imidazolidines, oxazolidines, etc)	4
S-Oxygenations of disulfides, thiosulfinates (-SO-S-), alpha-disulfoxides (-SO-SO-) and thiosulfonates (-SO2-S-)	4
Oxidation of primary hydroxylamines to nitroso compounds or oximes (incl. spontaneous dismutation), then to nitro compounds	4
Oxidative decarboxylation	3
Oxidation of silicon, phosphorus, arsenic and other atoms	2
Oxygenation of sulfenic acids to sulfinic acids, and of sulfinic acids to sulfonic acids	2
Reduction of disulfides to thiols	2
Reduction of Se, P, Hg, As and other atoms	2

[a] Number of SoM annotations per reaction subclass.

Table S2: Reaction Subclasses of CYP-mediated Metabolism Annotated in the MetaQSAR Database.

Reaction Subclass	Annotation count[a]
Oxidation of aryl compounds to epoxides, phenols or other metabolites	814
Hydroxylation (or other oxidations) of Csp3 carrying a heteroatom (N, O, S, halo) (including subsequent dealkylation, deamination or dehalogenation)	810
Hydroxylation (or other oxidations) of isolated Csp3	346
Hydroxylation (or other oxidations) of C alpha to an unsaturated system (>C=C<, >C=O, -C±N, aryl)	321
Oxygenation of sulfides to sulfoxides, and of sulfoxides to sulfones	100
Dehydrogenation of >CH-CH< to >C=C<, of >CH-N< to >C=N- (incl. >C=N+<)	88
Oxygenation of >C=C< bonds to epoxides or other metabolites	73
Oxidation of azaarenes to lactams or other metabolites	56
Oxidation of tertiary alkylamines and heterocyclic amines to N-oxides or other metabolites	53
Oxidation of tertiary arylamines, azaarenes and azo compounds to N-oxides oxides or other metabolites	51
Hydroxylation of amines to hydroxylamines or intermediates	51

Other oxidations of phenols and amines (dimerization, quinone-like metabolites, etc)	49
Oxidation of diphenols to quinones	19
Oxidation of cresols and analogs to quinonemethides	16
Reduction of nitro compounds to nitroso compounds	13
Oxidation of amino- and amido-phenols to quinoneimines or quinoneimides, resp.	12
Oxidative desulfurations of >C=S to ketones, and of -P=S to -P=O groups	11
Oxidation of -CHO to -COOH	9
Other N-oxidations (1,4-dihydropyridines, etc)	9
Hydrogenation of -CHO to -CH2OH and of >C=O to >CHOH	8
Other Csp3 oxidations (organometallic dealkylation, C-C cleavage, etc)	6
Oxygenation of -C-C±C-H and -C-C±C-C- bonds	6
Other C reductions, e.g. of >C=C< to -CH2-CH2-	6
Reductive dehalogenations	5
Hydroxylation of amides to hydroxylamides	5
Oxidation of thiols to sulfenic acids or disulfides	5
Oxidation of silicon, phosphorus, arsenic and other atoms	5
Other N-reductions (e.g. azo compounds to hydrazines, hydrazines to amines, reductive N-rings opening)	4
Bond order increases by elimination of H2O or RSH or other Nu-H	4
Reduction of N-oxides	3
Dehydrogenation of -CH2OH groups to -CHO and of >CHOH to >C=O	2
Oxidative decarboxylation	2
Reduction of hydroxylamines and hydroxylamides (incl. spontaneous dismutation)	2
Oxygenation of sulfenic acids to sulfinic acids, and of sulfinic acids to sulfonic acids	2
Oxygenation of thiones (>C=S) or thioamides to sulfines, and of sulfines to sulfenes	2
S-Oxygenations of disulfides, thiosulfinates (-SO-S-), alpha-disulfoxides (-SO-SO-) and thiosulfonates (-SO2-S-)	2
Reduction of quinones and analogs	1
Reduction of sulfoxides to sulfides	1
Reduction of Se, P, Hg, As and other atoms	1

Hydrolysis of alkyl esters	1
Hydrolysis of linear and cyclic carbamates (>N-CO-OR') and carbonates (RO-CO-OR')	1
Hydrolysis of thioesters (RCO-SR' and RCS-SR') and thiolactones	1
Hydrolysis of anilides [aryl-N-CO-C~], hydrazides [-CO-NHN<], sulfonamides [-SO2NH-], linear imides [-CO-NR-CO-] and urea derivatives [>N-CO-N<]	1
Hydrolytic dehalogenations	1

[a] Number of SoM annotations per reaction subclass.

Table S3: Reaction Subclasses of Phase 2 Metabolism Annotated in the MetaQSAR Database.

Reaction Subclass	**Annotation Count[a]**
O-Glucuronidation of phenols	203
Nucleophilic additions of glutathione (to a,ß-unsaturated carbonyls, quinones and analogues, isocyanates and isothiocyanates, epoxides, etc)	112
O-Glucuronidation of alcohols	110
O-Glucuronidation of carboxylic acids (including subsequent rearrangements)	108
N-Glucuronidation of linear and cyclic amines (including =N- in azaarenes)	106
Reactions of glutathione addition-elimination (at C-X groups, acyl halides, halogenated olefins, etc)	96
O-Sulfonation of phenols	85
O-Methylation of catechols and other hydroxy groups	68
N-Acetylation of aromatic amines	42
Conjugation with glycine, glutamic acid, taurine and other amino acids or short peptides	41
O-Sulfonation of alcohols	33
Other acetylations (N-acetylation of alkylamines, O-acetylation, etc)	31
Other reactions of conjugation, transaminations	24
Other reactions (O-sulfonation of hydroxylamines, N-sulfonation of amines, etc)	22
Chain elongation by 2C, beta- or alpha- oxidation (loss of 2C or 1C, resp.), other sequels	22
Any conjugations with glucose or other sugars	18
Non-enzymatic formation of hydrazones, binding of CO2 to form carbamates	16
N-Glucuronidation of amides	14

Reactions of phosphorylation	14
Formation of hybrid glycerides, conjugation with cholesterol or other sterols	13
S-Methylation of thiols	9
N-Methylation of exocyclic and endocyclic amino groups (including =N- in azaarenes)	9
O-Glucuronidation of hydroxylamines and hydroxylamides	9
Conjugation with carnitine	8
Reactions of acylation (formylation, formation of fatty acyl esters, etc)	7
Conjugations with other thiols (Cys, N-Ac-Cys, etc)	6
S-Glucuronidation of thiols and thioacids, C-Glucuronidation of acidic enols	4
Metabolic processing of glutathione conjugates up to thiols	4
Unidirectional chiral inversion of profens and analogues	3
N-Acetylation of hydrazines and hydrazides	2
Reductions following glutathione conjugations	1

[a] Number of SoM annotations per reaction subclass.

Table S4: Optimum Parameters Selected by Grid Search During 10-fold Cross-Validation.[a]

Model	Descriptor Set	max_features	max_features_ANOVA
P1+P2	CDK [0]	0.6	15
P1+P2	circCDK + ATF [1]	0.3	200
P1+P2	circCDK + ATF [2]	0.3	400
P1+P2	circCDK + ATF [3]	0.3	400
P1+P2	circCDK + ATF [4]	0.6	400
P1+P2	circCDK + ATF [5]	0.6	400
P1+P2	circCDK + ATF [6]	0.6	400
P1+P2	circCDK + ATF [7]	0.6	400
P1+P2	circCDK + ATF [8]	0.9	400
P1+P2	circCDK + ATF [9]	0.9	400
P1+P2	circCDK + ATF [10]	0.6	400
CYP	CDK [0]	0.9	15
CYP	circCDK + ATF [1]	0.3	200
CYP	circCDK + ATF [2]	0.3	400
CYP	circCDK + ATF [3]	0.9	400
CYP	circCDK + ATF [4]	0.6	400

CYP	circCDK + ATF [5]	0.9	400
CYP	circCDK + ATF [6]	0.9	400
CYP	circCDK + ATF [7]	0.9	400
CYP	circCDK + ATF [8]	0.6	400
CYP	circCDK + ATF [9]	0.9	400
CYP	circCDK + ATF [10]	0.9	400
P1	CDK [0]	0.9	15
P1	circCDK + ATF [1]	0.9	200
P1	circCDK + ATF [2]	0.6	400
P1	circCDK + ATF [3]	0.6	400
P1	circCDK + ATF [4]	0.9	400
P1	circCDK + ATF [5]	0.9	400
P1	circCDK + ATF [6]	0.9	400
P1	circCDK + ATF [7]	0.6	400
P1	circCDK + ATF [8]	0.6	400
P1	circCDK + ATF [9]	0.6	400
P1	circCDK + ATF [10]	0.9	400
P2	CDK [0]	0.3	15
P2	circCDK + ATF [1]	0.9	200
P2	circCDK + ATF [2]	0.9	200
P2	circCDK + ATF [3]	0.3	400
P2	circCDK + ATF [4]	0.9	400
P2	circCDK + ATF [5]	0.6	400
P2	circCDK + ATF [6]	0.6	400
P2	circCDK + ATF [7]	0.6	400
P2	circCDK + ATF [8]	0.6	400
P2	circCDK + ATF [9]	0.9	400
P2	circCDK + ATF [10]	0.6	400
P1+P2 100+	CDK [0]	0.6	15
P1+P2 100+	circCDK + ATF [1]	0.3	200
P1+P2 100+	circCDK + ATF [2]	0.9	400
P1+P2 100+	circCDK + ATF [3]	0.6	400
P1+P2 100+	circCDK + ATF [4]	0.9	400
P1+P2 100+	circCDK + ATF [5]	0.6	400
P1+P2 100+	circCDK + ATF [6]	0.9	400
P1+P2 100+	circCDK + ATF [7]	0.6	400
P1+P2 100+	circCDK + ATF [8]	0.6	400
P1+P2 100+	circCDK + ATF [9]	0.6	400
P1+P2 100+	circCDK + ATF [10]	0.9	400

CYP 100+	CDK [0]	0.6	15
CYP 100+	circCDK + ATF [1]	0.9	200
CYP 100+	circCDK + ATF [2]	0.3	400
CYP 100+	circCDK + ATF [3]	0.9	400
CYP 100+	circCDK + ATF [4]	0.9	400
CYP 100+	circCDK + ATF [5]	0.9	400
CYP 100+	circCDK + ATF [6]	0.9	400
CYP 100+	circCDK + ATF [7]	0.3	400
CYP 100+	circCDK + ATF [8]	0.6	400
CYP 100+	circCDK + ATF [9]	0.6	400
CYP 100+	circCDK + ATF [10]	0.6	400
P1 100+	CDK [0]	0.9	15
P1 100+	circCDK + ATF [1]	0.3	200
P1 100+	circCDK + ATF [2]	0.6	200
P1 100+	circCDK + ATF [3]	0.3	400
P1 100+	circCDK + ATF [4]	0.9	400
P1 100+	circCDK + ATF [5]	0.6	400
P1 100+	circCDK + ATF [6]	0.9	400
P1 100+	circCDK + ATF [7]	0.6	400
P1 100+	circCDK + ATF [8]	0.6	400
P1 100+	circCDK + ATF [9]	0.9	400
P1 100+	circCDK + ATF [10]	0.9	400
P2 100+	CDK [0]	0.3	15
P2 100+	circCDK + ATF [1]	0.9	200
P2 100+	circCDK + ATF [2]	0.6	200
P2 100+	circCDK + ATF [3]	0.9	400
P2 100+	circCDK + ATF [4]	0.6	200
P2 100+	circCDK + ATF [5]	0.3	200
P2 100+	circCDK + ATF [6]	0.3	400
P2 100+	circCDK + ATF [7]	0.9	400
P2 100+	circCDK + ATF [8]	0.6	400
P2 100+	circCDK + ATF [9]	0.3	400
P2 100+	circCDK + ATF [10]	0.6	400

[a] Selected values of these parameters are shown for all combinations of models and descriptor sets developed in this study.

Table S5: Average Performance of the Optimal "P1+P2 Model" during 10-fold Cross-Validation.[a]

Descriptor Set	Top-2 (avg)	Top-2 (std)	Top-3 (avg)	Top-3 (std)	MCC (avg)	MCC (std)	AUC (avg)	AUC (std)
CDK [0]	78%	3%	85%	2%	0.47	0.03	0.87	0.02
circCDK + ATF [1]	81%	2%	87%	2%	0.49	0.03	0.88	0.02
circCDK + ATF [2]	81%	3%	88%	2%	0.50	0.03	0.89	0.01
circCDK + ATF [3]	83%	2%	89%	1%	0.51	0.03	0.89	0.01
circCDK + ATF [4]	82%	3%	89%	2%	0.51	0.03	0.89	0.01
circCDK + ATF [5]	82%	1%	88%	2%	0.51	0.03	0.89	0.01
circCDK + ATF [6]	82%	2%	88%	2%	0.51	0.02	0.89	0.01
circCDK + ATF [7]	82%	2%	88%	2%	0.51	0.02	0.89	0.01
circCDK + ATF [8]	82%	1%	88%	2%	0.51	0.02	0.89	0.01
circCDK + ATF [9]	82%	2%	88%	2%	0.50	0.02	0.89	0.01
circCDK + ATF [10]	82%	1%	88%	1%	0.51	0.02	0.89	0.01

[a] For each model, average MCC, AUC and Top-2 and Top-3 success rates are shown. Standard deviations across the ten folds are also shown.

Table S6: Average Performance of the Optimal "CYP Model" during 10-fold Cross-Validation.[a]

Descriptor Set	Top-2 (avg)	Top-2 (std)	Top-3 (avg)	Top-3 (std)	MCC (avg)	MCC (std)	AUC (avg)	AUC (std)
CDK [0]	77%	4%	86%	2%	0.47	0.05	0.87	0.02
circCDK + ATF [1]	80%	3%	88%	3%	0.50	0.05	0.88	0.02
circCDK + ATF [2]	82%	3%	87%	3%	0.51	0.05	0.89	0.02
circCDK + ATF [3]	82%	4%	89%	3%	0.51	0.05	0.89	0.02
circCDK + ATF [4]	81%	4%	90%	4%	0.51	0.04	0.89	0.02
circCDK + ATF [5]	82%	3%	89%	2%	0.52	0.05	0.90	0.02
circCDK + ATF [6]	83%	3%	89%	3%	0.52	0.05	0.90	0.02
circCDK + ATF [7]	83%	3%	89%	2%	0.51	0.05	0.90	0.02
circCDK + ATF [8]	84%	2%	89%	3%	0.51	0.04	0.90	0.02
circCDK + ATF [9]	83%	3%	89%	2%	0.52	0.04	0.90	0.02
circCDK + ATF [10]	83%	3%	89%	2%	0.51	0.04	0.90	0.02

[a] For each model, average MCC, AUC and Top-2 and Top-3 success rates are shown. Standard deviations across the ten folds are also shown.

Table S7: Average Performance of the Optimal P1 Model during 10-fold Cross-Validation.[a]

Descriptor Set	Top-2 (avg)	Top-2 (std)	Top-3 (avg)	Top-3 (std)	MCC (avg)	MCC (std)	AUC (avg)	AUC (std)
CDK [0]	79%	4%	85%	2%	0.47	0.04	0.87	0.02
circCDK + ATF [1]	82%	3%	88%	3%	0.50	0.04	0.88	0.02
circCDK + ATF [2]	83%	3%	88%	3%	0.51	0.03	0.89	0.02
circCDK + ATF [3]	83%	3%	89%	4%	0.52	0.03	0.90	0.02
circCDK + ATF [4]	83%	4%	90%	3%	0.52	0.04	0.90	0.01
circCDK + ATF [5]	84%	3%	90%	3%	0.53	0.04	0.90	0.01
circCDK + ATF [6]	83%	3%	89%	4%	0.53	0.03	0.90	0.01
circCDK + ATF [7]	82%	4%	88%	3%	0.53	0.04	0.90	0.01
circCDK + ATF [8]	83%	4%	88%	3%	0.52	0.04	0.90	0.01
circCDK + ATF [9]	82%	4%	89%	4%	0.52	0.04	0.90	0.02
circCDK + ATF [10]	83%	4%	89%	4%	0.52	0.04	0.90	0.01

[a] For each model, average MCC, AUC and Top-2 and Top-3 success rates are shown. Standard deviations across the ten folds are also shown.

Table S8: Average Performance of the Optimal P2 Model during 10-fold Cross-Validation.[a]

Descriptor Set	Top-2 (avg)	Top-2 (std)	Top-3 (avg)	Top-3 (std)	MCC (avg)	MCC (std)	AUC (avg)	AUC (std)
CDK [0]	89%	5%	93%	3%	0.70	0.06	0.96	0.02
circCDK + ATF [1]	89%	4%	93%	3%	0.72	0.06	0.97	0.01
circCDK + ATF [2]	90%	4%	94%	2%	0.71	0.06	0.96	0.01
circCDK + ATF [3]	91%	4%	94%	2%	0.71	0.06	0.97	0.01
circCDK + ATF [4]	92%	4%	94%	3%	0.72	0.05	0.97	0.01
circCDK + ATF [5]	90%	5%	94%	4%	0.72	0.05	0.97	0.01
circCDK + ATF [6]	91%	4%	93%	3%	0.72	0.05	0.97	0.01
circCDK + ATF [7]	89%	4%	93%	2%	0.72	0.05	0.97	0.01
circCDK + ATF [8]	89%	5%	93%	4%	0.72	0.04	0.97	0.01
circCDK + ATF [9]	89%	5%	92%	2%	0.72	0.04	0.97	0.01
circCDK + ATF [10]	90%	5%	93%	4%	0.72	0.05	0.97	0.01

[a] For each model, average MCC, AUC and Top-2 and Top-3 success rates are shown. Standard deviations across the ten folds are also shown.

Table S9: Average Performance of the Optimal P1+P2 100+ Model during 10-fold Cross-Validation.[a]

Descriptor Set	Top-2 (avg)	Top-2 (std)	Top-3 (avg)	Top-3 (std)	MCC (avg)	MCC (std)	AUC (avg)	AUC (std)
CDK [0]	78%	4%	84%	4%	0.48	0.02	0.89	0.01
circCDK + ATF [1]	81%	5%	87%	4%	0.51	0.02	0.90	0.01
circCDK + ATF [2]	81%	6%	89%	5%	0.51	0.03	0.91	0.01
circCDK + ATF [3]	82%	3%	89%	4%	0.52	0.02	0.91	0.01
circCDK + ATF [4]	83%	3%	88%	4%	0.53	0.01	0.91	0.01
circCDK + ATF [5]	83%	3%	88%	3%	0.54	0.01	0.91	0.01
circCDK + ATF [6]	84%	3%	89%	3%	0.53	0.01	0.91	0.01
circCDK + ATF [7]	83%	5%	89%	3%	0.54	0.02	0.91	0.01
circCDK + ATF [8]	83%	4%	89%	3%	0.54	0.02	0.91	0.01
circCDK + ATF [9]	82%	3%	88%	3%	0.54	0.02	0.91	0.01
circCDK + ATF [10]	83%	4%	88%	4%	0.54	0.02	0.91	0.01

[a] For each model, average MCC, AUC and Top-2 and Top-3 success rates are shown. Standard deviations across the ten folds are also shown.

Table S10: Average Performance of the Optimal CYP 100+ Model during 10-fold Cross-Validation.[a]

Descriptor Set	Top-2 (avg)	Top-2 (std)	Top-3 (avg)	Top-3 (std)	MCC (avg)	MCC (std)	AUC (avg)	AUC (std)
CDK [0]	80%	5%	88%	4%	0.50	0.03	0.90	0.01
circCDK + ATF [1]	82%	4%	89%	3%	0.53	0.03	0.91	0.01
circCDK + ATF [2]	86%	4%	91%	3%	0.54	0.03	0.91	0.01
circCDK + ATF [3]	85%	3%	91%	3%	0.54	0.03	0.92	0.01
circCDK + ATF [4]	86%	4%	92%	4%	0.53	0.04	0.92	0.01
circCDK + ATF [5]	86%	4%	92%	4%	0.55	0.03	0.92	0.01
circCDK + ATF [6]	87%	4%	92%	4%	0.54	0.03	0.92	0.01
circCDK + ATF [7]	85%	4%	92%	4%	0.54	0.04	0.92	0.01
circCDK + ATF [8]	87%	5%	93%	4%	0.55	0.04	0.92	0.01
circCDK + ATF [9]	87%	5%	94%	3%	0.54	0.04	0.92	0.01
circCDK + ATF [10]	87%	4%	92%	4%	0.54	0.04	0.92	0.01

[a] For each model, average MCC, AUC and Top-2 and Top-3 success rates are shown. Standard deviations across the ten folds are also shown.

Table S11: Average Performance of the Optimal P1 100+ Model During 10-fold Cross-Validation.[a]

Descriptor Set	Top-2 (avg)	Top-2 (std)	Top-3 (avg)	Top-3 (std)	MCC (avg)	MCC (std)	AUC (avg)	AUC (std)
CDK [0]	82%	4%	89%	3%	0.52	0.04	0.90	0.01
circCDK + ATF [1]	84%	4%	90%	4%	0.54	0.04	0.91	0.01
circCDK + ATF [2]	86%	4%	92%	3%	0.57	0.04	0.92	0.01
circCDK + ATF [3]	86%	5%	92%	2%	0.56	0.05	0.92	0.01
circCDK + ATF [4]	85%	4%	92%	2%	0.57	0.04	0.93	0.01
circCDK + ATF [5]	86%	4%	92%	2%	0.57	0.05	0.93	0.01
circCDK + ATF [6]	88%	4%	93%	1%	0.57	0.04	0.93	0.01
circCDK + ATF [7]	87%	3%	93%	2%	0.58	0.03	0.93	0.01
circCDK + ATF [8]	87%	3%	92%	3%	0.58	0.04	0.93	0.01
circCDK + ATF [9]	88%	2%	93%	2%	0.58	0.04	0.93	0.01
circCDK + ATF [10]	88%	4%	93%	2%	0.58	0.04	0.93	0.01

[a] For each model, average MCC, AUC and Top-2 and Top-3 success rates are shown. Standard deviations across the ten folds are also shown.

Table S12: Average Performance of the Optimal P2 100+ Model During 10-fold Cross-Validation.[a]

Descriptor Set	Top-2 (avg)	Top-2 (std)	Top-3 (avg)	Top-3 (std)	MCC (avg)	MCC (std)	AUC (avg)	AUC (std)
CDK [0]	89%	6%	93%	4%	0.70	0.06	0.97	0.01
circCDK + ATF [1]	91%	5%	95%	3%	0.71	0.04	0.97	0.01
circCDK + ATF [2]	94%	3%	96%	3%	0.73	0.06	0.97	0.01
circCDK + ATF [3]	93%	4%	96%	3%	0.74	0.06	0.97	0.01
circCDK + ATF [4]	92%	4%	95%	3%	0.74	0.04	0.98	0.01
circCDK + ATF [5]	91%	5%	95%	4%	0.75	0.05	0.97	0.01
circCDK + ATF [6]	92%	5%	95%	4%	0.75	0.05	0.98	0.01
circCDK + ATF [7]	92%	5%	95%	3%	0.75	0.05	0.98	0.01
circCDK + ATF [8]	93%	5%	95%	3%	0.76	0.05	0.98	0.01
circCDK + ATF [9]	92%	4%	95%	3%	0.76	0.06	0.98	0.01
circCDK + ATF [10]	92%	5%	95%	4%	0.75	0.06	0.98	0.01

[a] For each model, average MCC, AUC and Top-2 and Top-3 success rates are shown. Standard deviations across the ten folds are also shown.