

Supporting Information

Table SM1. Performance of BDDCS classification models for Random Forest, Naïve Bayes and GTM.

	Class	TP Rate	FP Rate	Precision	F-Measure
<i>Random Forest</i>	1	0.69 ± 0.05 (0.57)	0.25 ± 0.04 (0.55)	0.63 ± 0.05 (0.38)	0.66 ± 0.05 (0.46)
	2	0.66 ± 0.09 (0.27)	0.14 ± 0.03 (0.26)	0.66 ± 0.07 (0.31)	0.66 ± 0.07 (0.29)
	3	0.68 ± 0.07 (0.19)	0.13 ± 0.03 (0.19)	0.66 ± 0.05 (0.28)	0.67 ± 0.06 (0.22)
	4	0.06 ± 0.07 (0.01)	0.004 ± 0.005 (0.005)	0.25 ± 0.30 (0.06)	0.10 ± 0.11 (0.02)
<i>Naïve Bayes</i>	1	0.46 ± 0.06 (0.52)	0.16 ± 0.04 (0.47)	0.64 ± 0.05 (0.39)	0.53 ± 0.04 (0.43)
	2	0.72 ± 0.04 (0.26)	0.25 ± 0.06 (0.26)	0.55 ± 0.05 (0.31)	0.62 ± 0.03 (0.24)
	3	0.44 ± 0.09 (0.23)	0.09 ± 0.02 (0.24)	0.65 ± 0.08 (0.27)	0.52 ± 0.09 (0.22)
	4	0.46 ± 0.15 (0.07)	0.14 ± 0.03 (0.06)	0.17 ± 0.06 (0.10)	0.24 ± 0.07 (0.07)
<i>Bayes GTM</i>	1	0.49 ± 0.05 (0.32)	0.24 ± 0.05 (0.29)	0.56 ± 0.04 (0.40)	0.52 ± 0.04 (0.35)
	2	0.53 ± 0.06 (0.29)	0.17 ± 0.05 (0.28)	0.58 ± 0.09 (0.31)	0.55 ± 0.06 (0.29)
	3	0.53 ± 0.08 (0.30)	0.14 ± 0.03 (0.29)	0.59 ± 0.07 (0.28)	0.55 ± 0.07 (0.29)
	4	0.30 ± 0.14 (0.18)	0.13 ± 0.04 (0.15)	0.12 ± 0.06 (0.07)	0.17 ± 0.08 (0.10)
<i>kNNd GTM</i>	1	0.49 ± 0.05 (0.33)	0.27 ± 0.04 (0.30)	0.53 ± 0.04 (0.40)	0.51 ± 0.03 (0.35)
	2	0.49 ± 0.07 (0.29)	0.18 ± 0.05 (0.29)	0.54 ± 0.10 (0.31)	0.51 ± 0.08 (0.30)

	3	0.53 ± 0.05 (0.30)	0.16 ± 0.04 (0.29)	0.56 ± 0.04 (0.28)	0.54 ± 0.04 (0.29)
	4	0.23 ± 0.10 (0.16)	0.11 ± 0.04 (0.13)	0.11 ± 0.05 (0.08)	0.14 ± 0.06 (0.10)

Performance of BDDCS classification models for RF, Naïve Bayes and GTM. Models are validated using 10-fold cross-validation. Intervals are estimated using t-statistics on the 10 modeling procedures. Numbers between parentheses are an upper estimate of the performances of models trained on y-scrambled data.

Table SM2. 34 compounds or substances were removed from the study. Some compounds did not fit into the BDDCS (“class 0” compounds), and 24 other compounds were discarded for the reasons listed below.

Removed compound	BDDCS class	Reason
Fosfomycin	3	Salt/mixture
Gallium nitrate	3	Salt/mixture
Ferrous sulfate	3	Salt/mixture
Lithium carbonate	3	Salt/mixture
Potassium chloride	3	Salt/mixture
Quinupristin-dalfopristin	4	Salt/mixture
Enfuvirtide	1	Peptide
Exenatide	1	Peptide
Liraglutide	1	Peptide
Nafarelin	1	Peptide
Premlintide	1	Peptide
Dolasetron	1	Alcohol/ketone
Hydrolasetron	3	Alcohol/ketone
Roxatidine	3	Salt/mixture
Roxatidine acetate HCl	1	Salt/mixture
Erythromycin stearate	4	Salt/mixture
Erythromycin A	3	Salt/mixture
Erythromycin lactobionate	3	Salt/mixture
Escitalopram	1	Enantiomer/racemic
Citalopram	2	Enantiomer/racemic
Oseltamivir phosphate	1	Salt/mixture
Oseltamivir	3	Salt/mixture
Norgestrel	1	Enantiomer/racemic

Levonorgestrel	4	Enantiomer/racemic
----------------	---	--------------------