

Altaner et al.:

Machine Learning Prediction of Cyanobacterial Toxin (Microcystin) Toxicodynamics in Humans

Supplementary Data

$$\text{Precision} = \frac{\# TP}{\# TP + \# FP} \quad (1) \quad \begin{array}{l} \# TP = \text{Number of True Positive} \\ \# FP = \text{Number of False Positive} \end{array}$$
$$\text{Recall} = \frac{\# TP}{\# TP + \# FN} \quad (2) \quad \# FN = \text{Number of False Negative}$$
$$F - \text{Score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (3)$$

Fig. S1: Calculation of prediction measurements

Precision is defined as the ability of a model to retrieve relevant results, recall as the ability of a model to retrieve relevant results and the F-Score as harmonic mean of precision and recall.

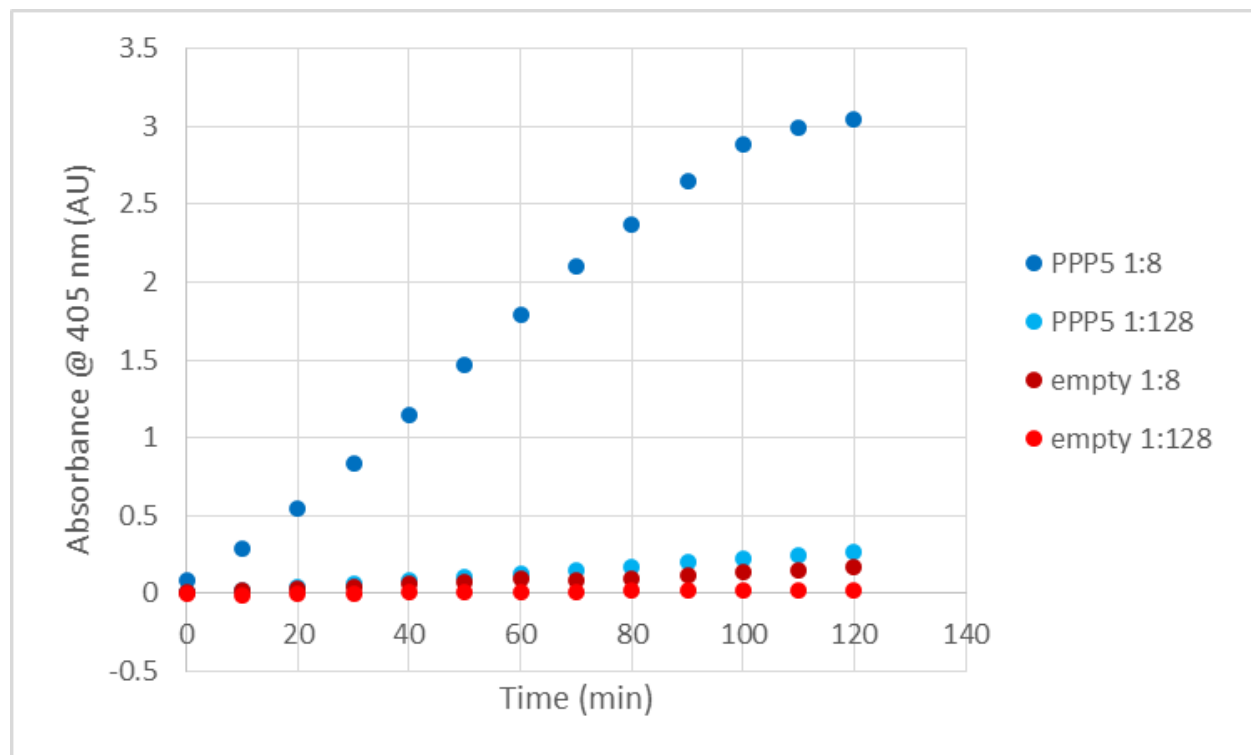


Fig. S2: Comparison of the hPPP5 expression with the empty expression

Different dilutions of the two respective fractions were analyzed towards their activity to dephosphorylate pNPP.

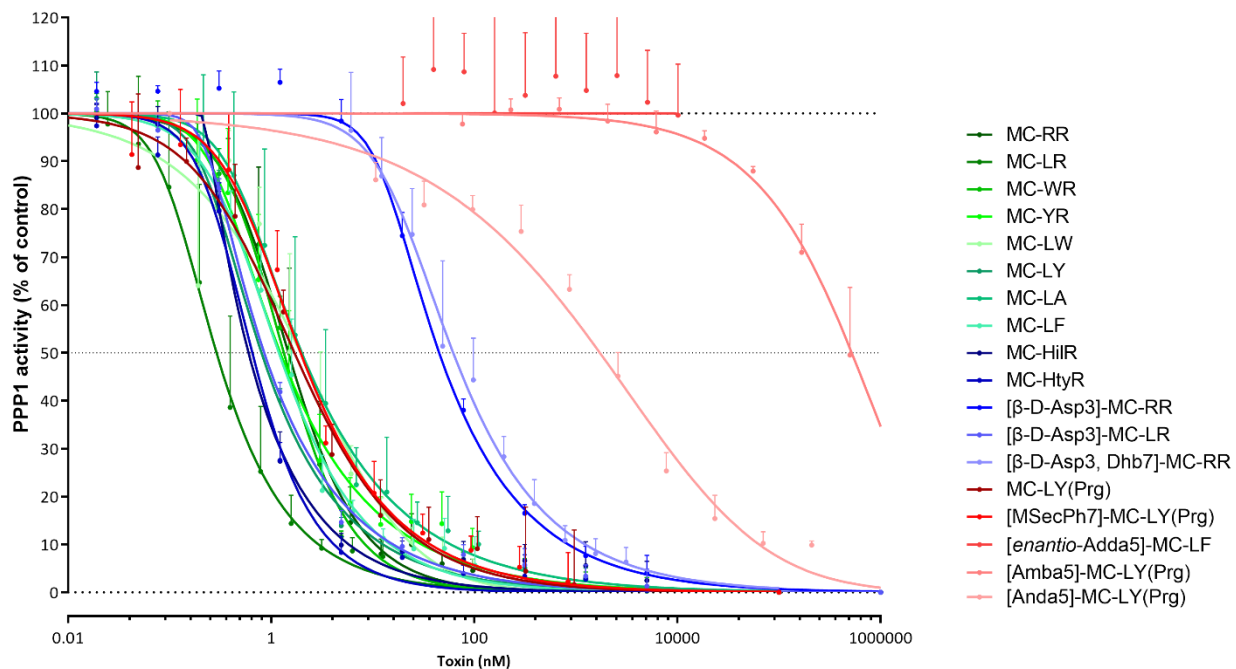


Fig. S3: Inhibition curves for PPP1
 Only upper error bars (SD) are shown for reasons of clarity.

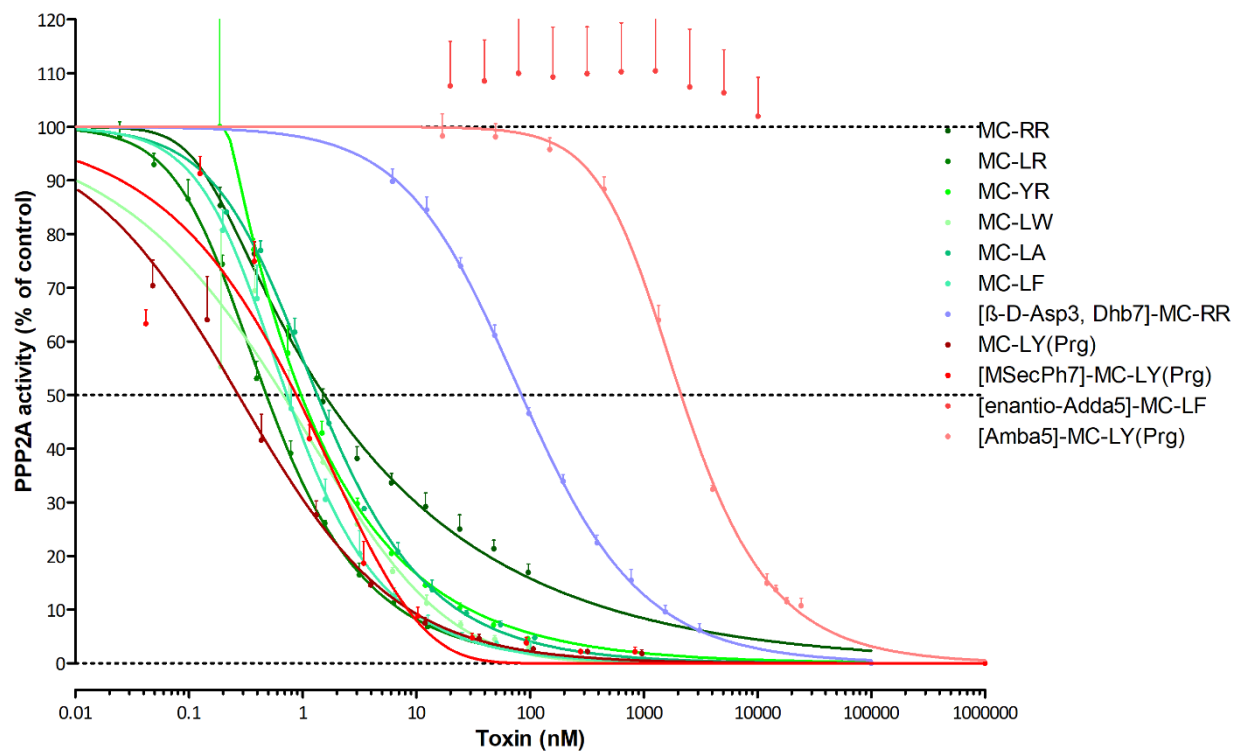


Fig. S4: Inhibition curves for PPP2A
 Only upper error bars (SD) are shown for reasons of clarity.

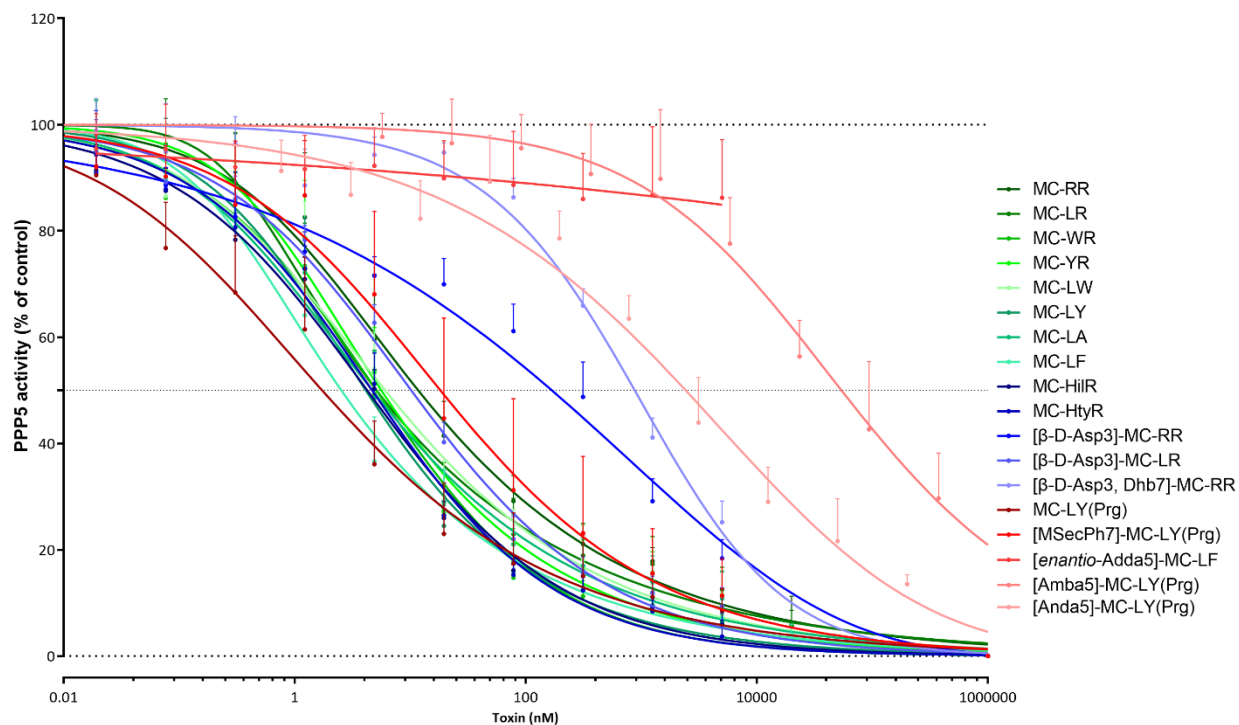


Fig. S5: Inhibition curves for PPP5
Only upper error bars (SD) are shown for reasons of clarity.

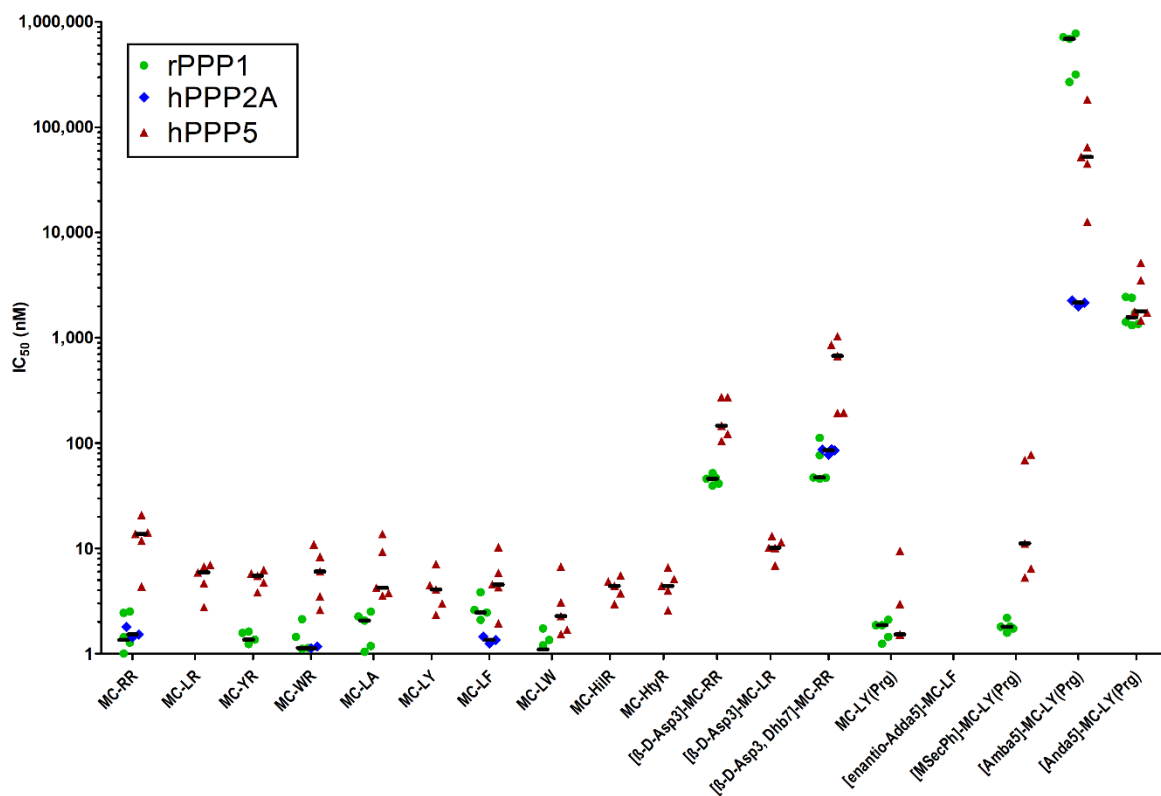


Fig. S6: Comparison IC₅₀s on the three tested phosphatases
Biological replicates were plotted individually using GraphPad Prism 5 and individual IC₅₀ are noted. Black bar shows the median of the IC₅₀.

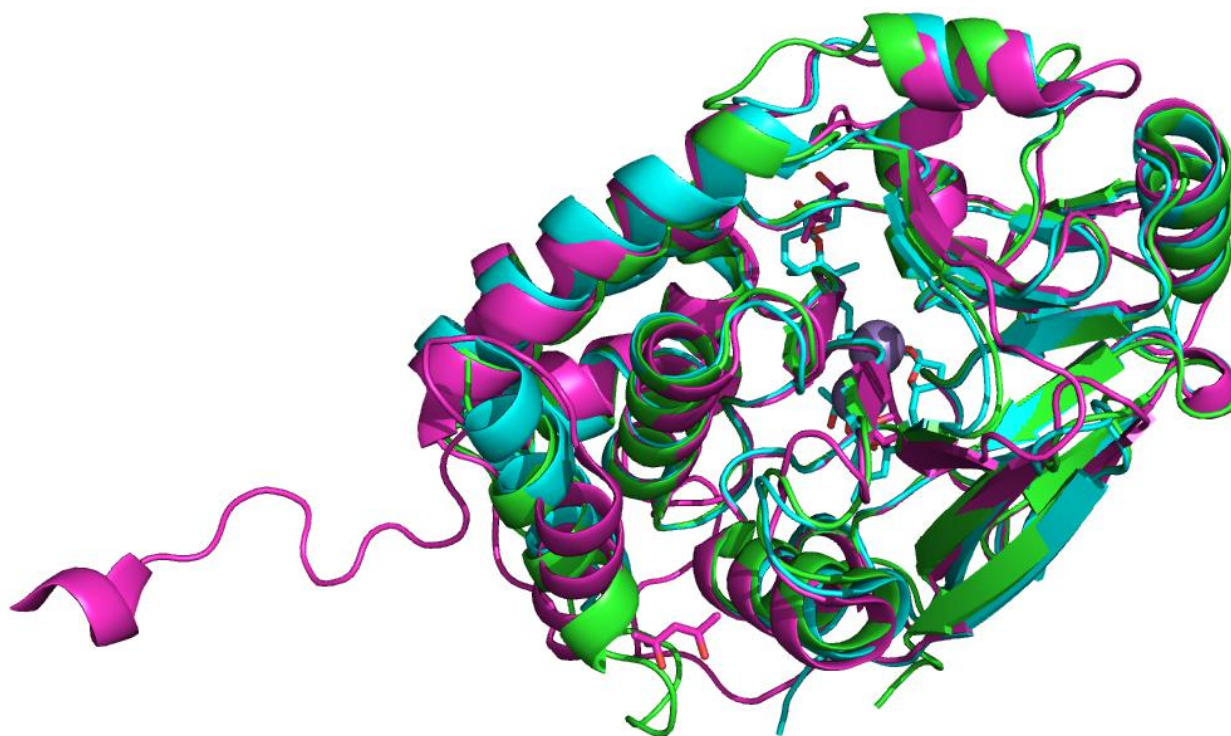


Fig. S7: Overlay of the structures of the used PPP. PyMOL was used using the PDB entries for PPP1 (4MOV, green), PPP2A (2IE4, teal) and PPP5 (4ZX2, magenta).

Tab. S1: Congener-dependent modifications

Hil (homisoleucine) and Hty (homotyrosine) are variants of isoleucine and tyrosine respectively, which have an additional CH₂ compared to the parent amino acids. For structures of amino acids see Figure 1.

Congener	X (2)	Z (4)	Position 3	Position 7	Position 5
MC-RR	Arginine	Arginine	β -D-MeAsp	Mdha	Adda
MC-LR	Leucine	Arginine	β -D-MeAsp	Mdha	Adda
MC-YR	Tyrosine	Arginine	β -D-MeAsp	Mdha	Adda
MC-WR	Tryptophane	Arginine	β -D-MeAsp	Mdha	Adda
MC-LA	Leucine	Alanine	β -D-MeAsp	Mdha	Adda
MC-LY	Leucine	Tyrosine	β -D-MeAsp	Mdha	Adda
MC-LF	Leucine	Phenylalanine	β -D-MeAsp	Mdha	Adda
MC-LW	Leucine	Tryptophane	β -D-MeAsp	Mdha	Adda
MC-HilR	Homisoleucine	Arginine	β -D-MeAsp	Mdha	Adda
MC-HtyR	Homotyrosine	Arginine	β -D-MeAsp	Mdha	Adda
[β -D-Asp3]-MC-RR	Arginine	Arginine	β -D-Asp	Mdha	Adda
[β -D-Asp3]-MC-LR	Leucine	Arginine	β -D-Asp	Mdha	Adda
[β -D-Asp3, Dhb7]-MC-RR	Arginine	Arginine	β -D-Asp	Dhb	Adda
MC-LY(Prg)	Leucine	Tyrosine(Prg)	β -D-MeAsp	Mdha	Adda
[<i>enanti</i> -Adda5]-MC-LF	Leucine	Phenylalanine	β -D-MeAsp	Mdha	<i>enanti</i> -Adda
[M _{Sec} Ph7]-MC-LY(Prg)	Leucine	Tyrosine(Prg)	β -D-MeAsp	M _{Sec} Ph	Adda
[Amba5]-MC-LY(Prg)	Leucine	Tyrosine(Prg)	β -D-MeAsp	Mdha	Amba
[Anda5]-MC-LY(Prg)	Leucine	Tyrosine(Prg)	β -D-MeAsp	Mdha	Anda

Tab. S2: Results of the mass spectrometric analysis

First two hits were hPPP5 and thioredoxin, which are both part of the expressed fusion protein. Other proteins were found in lower quantity.

Accession	Description	Score	Coverage	# Proteins	# Unique Peptides	# Peptides	# PSMs	# AAs	MW [kDa]	calc. pI
P53041	Serine/threonine-protein phosphatase 5 OS=Homo sapiens GN=PPP5C PE=1 SV=1 - [PPP5_HUMAN]	18779.21	69.74	10	35	35	626	499	56.8	6.28
C6EG32	Thioredoxin OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_4259 PE=4 SV=1 - [C6EG32_ECOBD]	5021.36	66.97	1	6	6	128	109	11.8	4.88
C6EG74	Glutamine--fructose-6-phosphate aminotransferase [isomerizing] OS=Escherichia coli (strain B / BL21-DE3) GN=glmS PE=3 SV=1 - [C6EG74_ECOBD]	2972.68	67.16	1	33	33	89	609	66.9	5.87
C6ECY2	Chaperonin GroEL OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_3888 PE=3 SV=1 - [C6ECY2_ECOBD]	2072.26	72.26	1	29	29	68	548	57.3	4.94
C6EB40	Chaperone protein DnaK OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_3605 PE=3 SV=1 - [C6EB40_ECOBD]	1381.54	52.98	1	23	26	43	638	69.1	4.97
C6EE39	Translation elongation factor Tu OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_4053 PE=3 SV=1 - [C6EE39_ECOBD]	1286.33	59.39	2	16	18	45	394	43.3	5.45
C6ECZ5	Lysine--tRNA ligase OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_3901 PE=3 SV=1 - [C6ECZ5_ECOBD]	1226.75	45.35	1	13	20	44	505	57.8	5.24
C6EKZ1	Chaperone protein HtpG OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_3183 PE=3 SV=1 - [C6EKZ1_ECOBD]	1221.43	57.05	1	24	27	42	624	71.4	5.21
C6E9U5	Bifunctional polymyxin resistance protein ArnA OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_1404 PE=3 SV=1 - [C6E9U5_ECOBD]	1112.67	48.18	1	21	21	37	660	74.2	6.87
C6EAU2	Pyruvate dehydrogenase complex dihydrolipoamide acetyltransferase OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_3504 PE=4 SV=1 - [C6EAU2_ECOBD]	1019.86	40.16	1	16	16	30	630	66.1	5.17
C6EI71	Ribosomal protein S1 OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_2684 PE=4 SV=1 - [C6EI71_ECOBD]	975.04	45.42	1	21	21	31	557	61.1	4.98
C6EJZ8	ATP-dependent chaperone ClpB OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_1092 PE=4 SV=1 - [C6EJZ8_ECOBD]	841.83	31.39	1	19	19	25	857	95.5	5.52
C6ECU7	Superoxide dismutase OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_1987 PE=4 SV=1 - [C6ECU7_ECOBD]	815.45	80.31	1	9	9	26	193	21.3	5.95
C6EL63	Cysteine synthase A OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_1267 PE=4 SV=1 - [C6EL63_ECOBD]	812.89	56.04	1	12	12	17	323	34.5	6.06
C6EJ93	Enolase OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_0950 PE=3 SV=1 - [C6EJ93_ECOBD]	672.19	46.30	1	12	12	19	432	45.6	5.48
C6EII3	Phosphoglycerate kinase OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_0812 PE=3 SV=1 - [C6EII3_ECOBD]	575.89	42.12	1	11	12	23	387	41.1	5.22
C6ECN8	Threonine--tRNA ligase OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_1926 PE=3 SV=1 - [C6ECN8_ECOBD]	537.73	18.22	1	11	11	18	642	74.0	6.19
C6EIL8	Lysine--tRNA ligase OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_0847 PE=3 SV=1 - [C6EIL8_ECOBD]	529.29	34.06	1	8	15	25	505	57.6	5.24
C6EAU1	Dihydrolipoamide dehydrogenase OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_3503 PE=4 SV=1 - [C6EAU1_ECOBD]	522.12	32.49	1	11	11	14	474	50.7	6.15
C6EJL4	Succinate dehydrogenase flavoprotein subunit OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_2937 PE=3 SV=1 - [C6EJL4_ECOBD]	484.14	23.30	1	10	10	13	588	64.4	6.27

C6EA90	Proline--tRNA ligase OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_3424 PE=3 SV=1 - [C6EA90_ECOBD]	446.70	25.70	1	10	10	12	572	63.6	5.19
P00761	Trypsin OS=Sus scrofa PE=1 SV=1 - [TRYF_PIG]	442.41	25.11	1	4	4	14	231	24.4	7.18
C6ECS9	Pyruvate kinase OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_1969 PE=4 SV=1 - [C6ECS9_ECOBD]	426.45	27.02	1	10	10	17	470	50.7	6.09
C6EC50	Glyceraldehyde-3-phosphate dehydrogenase OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_1865 PE=3 SV=1 - [C6EC50_ECOBD]	397.55	29.61	1	6	7	11	331	35.5	7.11
C6EL29	Trigger factor OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_3221 PE=3 SV=1 - [C6EL29_ECOBD]	383.65	34.26	1	11	11	13	432	48.2	4.88
C6EJ92	CTP synthase OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_0949 PE=3 SV=1 - [C6EJ92_ECOBD]	380.09	16.88	1	8	8	13	545	60.3	5.94
C6EFU4	Short-chain dehydrogenase/reductase SDR OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_2329 PE=1 SV=1 - [C6EFU4_ECOBD]	374.98	25.19	1	4	5	8	262	27.8	5.87
C6EA50	Aminoacyl-histidine dipeptidase OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_3384 PE=4 SV=1 - [C6EA50_ECOBD]	348.95	25.98	1	8	9	12	485	52.9	5.52
C6EBG1	Transaldolase OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_3610 PE=3 SV=1 - [C6EBG1_ECOBD]	334.15	38.49	2	9	9	12	317	35.2	5.21
C6EGV5	Isocitrate dehydrogenase, NADP-dependent OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_2463 PE=4 SV=1 - [C6EGV5_ECOBD]	322.82	22.36	1	6	6	11	416	45.7	5.40
C6EGZ6	Malate dehydrogenase, NAD-dependent OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_0511 PE=3 SV=1 - [C6EGZ6_ECOBD]	315.07	23.08	1	5	5	7	312	32.3	5.77
C6EL61	Phosphoenolpyruvate-protein phosphotransferase OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_1265 PE=4 SV=1 - [C6EL61_ECOBD]	311.30	14.43	1	6	6	8	575	63.5	4.87
C6EGG1	Ribosomal protein L6 OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_0446 PE=3 SV=1 - [C6EGG1_ECOBD]	302.07	41.81	1	6	6	12	177	18.9	9.70
C6EBI7	Purine nucleoside phosphorylase OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_3636 PE=3 SV=1 - [C6EBI7_ECOBD]	296.62	37.24	1	6	6	10	239	25.9	5.66
C6EIH4	Transketolase OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_0803 PE=4 SV=1 - [C6EIH4_ECOBD]	296.10	20.97	1	7	8	11	663	72.2	5.67
C6EB22	4-hydroxy-tetrahydronicotinate reductase OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_3585 PE=3 SV=1 - [C6EB22_ECOBD]	252.77	17.95	1	3	3	6	273	28.7	5.76
C6EI79	Formate acetyltransferase OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_2692 PE=4 SV=1 - [C6EI79_ECOBD]	245.29	11.97	1	7	7	8	760	85.3	6.01
C6EI52	Asparaginyl-tRNA synthetase OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_2665 PE=3 SV=1 - [C6EI52_ECOBD]	240.13	23.82	1	7	8	10	466	52.5	5.31
C6EFH1	Phosphoenolpyruvate carboxykinase (ATP) OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_0342 PE=3 SV=1 - [C6EFH1_ECOBD]	235.16	15.37	1	6	6	9	540	59.6	5.71
C6EG71	ATP synthase F1, beta subunit OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_4300 PE=3 SV=1 - [C6EG71_ECOBD]	233.94	25.87	1	9	9	9	460	50.3	5.01
C6E9W2	Glycerophosphoryl diester phosphodiesterase OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_1421 PE=4 SV=1 - [C6E9W2_ECOBD]	232.32	24.30	1	6	6	6	358	40.8	5.60

C6EI14	Fructose-bisphosphate aldolase, class II OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_0813 PE=4 SV=1 - [C6EI14_ECOBD]	231.68	20.61	1	5	5	8	359	39.1	5.86
C6EAB5	Translation elongation factor Ts OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_3449 PE=3 SV=1 - [C6EAB5_ECOBD]	222.18	24.73	1	6	6	10	283	30.4	5.29
C6ECY6	Aspartate ammonia-lyase OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_3892 PE=4 SV=1 - [C6ECY6_ECOBD]	206.81	17.78	1	6	7	10	478	52.3	5.29
C6EE34	50S ribosomal protein L10 OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_4048 PE=3 SV=1 - [C6EE34_ECOBD]	200.67	31.52	1	4	4	4	165	17.7	8.98
C6ECA9	Alpha, alpha-phosphotrehalase OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_3794 PE=4 SV=1 - [C6ECA9_ECOBD]	195.51	11.62	1	5	5	6	551	63.8	5.87
C6EI89	Seryl-tRNA synthetase OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_2702 PE=3 SV=1 - [C6EI89_ECOBD]	192.63	10.47	1	4	4	5	430	48.4	5.50
C6EE35	Ribosomal protein L1 OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_4049 PE=3 SV=1 - [C6EE35_ECOBD]	191.29	29.06	1	5	5	7	234	24.7	9.64
C6EK89	Peroxiredoxin OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_3047 PE=1 SV=1 - [C6EK89_ECOBD]	190.58	32.09	1	4	4	7	187	20.7	5.17
C6EEX2	Glycerol kinase OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_4098 PE=3 SV=1 - [C6EEX2_ECOBD]	184.75	12.95	1	6	6	8	502	56.2	5.50
C6ELH0	Riboflavin biosynthesis protein RibD OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_3247 PE=4 SV=1 - [C6ELH0_ECOBD]	182.02	16.08	1	4	4	7	367	40.2	7.83
C6EJP6	Phosphoglucomutase, alpha-D-glucose phosphate-specific OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_2973 PE=4 SV=1 - [C6EJP6_ECOBD]	170.33	13.92	1	5	5	6	546	58.4	5.71
C6EI54	Aspartate transaminase OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_2667 PE=4 SV=1 - [C6EI54_ECOBD]	166.81	14.14	1	5	5	5	396	43.5	5.77
H6VRF8	Keratin 1 OS=Homo sapiens GN=KRT1 PE=3 SV=1 - [H6VRF8_HUMAN]	164.03	6.52	8	4	4	4	644	66.0	8.12
C6EE66	Catalase/peroxidase HPI OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_4081 PE=3 SV=1 - [C6EE66_ECOBD]	160.52	11.16	1	5	6	6	726	80.0	5.31
C6EGH1	DNA-directed RNA polymerase, alpha subunit OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_0456 PE=3 SV=1 - [C6EGH1_ECOBD]	160.52	28.57	1	5	7	8	329	36.5	5.06
C6EK88	Alkyl hydroperoxide reductase, F subunit OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_3046 PE=3 SV=1 - [C6EK88_ECOBD]	147.66	10.55	1	4	4	4	531	57.4	5.69
C6EDZ9	Glucose-6-phosphate isomerase OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_4012 PE=3 SV=1 - [C6EDZ9_ECOBD]	144.54	10.75	1	3	4	4	549	61.5	6.29
C6EG37	Ketol-acid reductoisomerase OS=Escherichia coli (strain B / BL21-DE3) GN=ilvC PE=3 SV=1 - [C6EG37_ECOBD]	137.36	11.61	1	5	5	5	491	54.0	5.31
C6EKK1	Inosine-5'-monophosphate dehydrogenase OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_1178 PE=3 SV=1 - [C6EKK1_ECOBD]	135.63	15.57	1	4	4	5	488	52.0	6.42
C6EGH0	Ribosomal protein S4 OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_0455 PE=3 SV=1 - [C6EGH0_ECOBD]	135.16	20.87	1	4	4	5	206	23.5	10.05
C6EE78	Heat shock protein HslVU, ATPase subunit HslU OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_4093 PE=3 SV=1 - [C6EE78_ECOBD]	134.17	9.71	1	3	4	6	443	49.6	5.35

C6EG69	ATP synthase F1, alpha subunit OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_4298 PE=3 SV=1 - [C6EG69_ECOBD]	131.99	11.31	1	4	4	6	513	55.2	6.13
C6EH56	NusA antitermination factor OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_0571 PE=3 SV=1 - [C6EH56_ECOBD]	131.09	5.25	1	2	2	4	495	54.8	4.64
C6EGF2	Ribosomal protein S3 OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_0437 PE=3 SV=1 - [C6EGF2_ECOBD]	124.61	30.47	1	3	4	4	233	26.0	10.27
C6EAJ7	Methionine--tRNA ligase OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_1544 PE=3 SV=1 - [C6EAJ7_ECOBD]	118.56	7.68	1	4	4	4	677	76.2	5.94
C6EJK5	Cytochrome bd ubiquinol oxidase subunit I OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_2928 PE=4 SV=1 - [C6EJK5_ECOBD]	116.12	8.24	1	3	3	3	522	58.2	6.81
C6EGQ9	Ribose-phosphate pyrophosphokinase OS=Escherichia coli (strain B / BL21-DE3) GN=prs PE=3 SV=1 - [C6EGQ9_ECOBD]	113.00	7.42	1	2	2	2	337	36.6	5.74
C6EGG9	Ribosomal protein S11 OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_0454 PE=3 SV=1 - [C6EGG9_ECOBD]	112.01	18.60	1	1	2	3	129	13.8	11.33
C6EG40	Dihydroxy-acid dehydratase OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_4268 PE=3 SV=1 - [C6EG40_ECOBD]	111.31	6.01	1	2	3	4	616	65.5	6.01
C6EGM4	Extracellular solute-binding protein family 5 OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_2379 PE=4 SV=1 - [C6EGM4_ECOBD]	109.66	6.63	1	3	3	3	558	62.6	6.37
C6EG30	Transcription termination factor Rho OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_4257 PE=3 SV=1 - [C6EG30_ECOBD]	109.07	14.32	1	4	4	4	419	47.0	7.25
C6EJ19	2,3-bisphosphoglycerate-dependent phosphoglycerate mutase OS=Escherichia coli (strain B / BL21-DE3) GN=gpmA PE=3 SV=1 - [C6EJ19_ECOBD]	108.50	14.00	1	3	3	3	250	28.5	6.18
C6EBA1	Molecular chaperone Hsp31 and glyoxalase 3 OS=Escherichia coli (strain B / BL21-DE3) GN=hchA PE=2 SV=1 - [C6EBA1_ECOBD]	107.23	7.77	1	2	2	2	283	31.2	6.02
C6EJQ4	Glutaminyl-tRNA synthetase OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_2981 PE=3 SV=1 - [C6EJQ4_ECOBD]	104.01	10.65	1	4	5	6	554	63.4	6.28
C6EGE9	Ribosomal protein L2 OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_0434 PE=3 SV=1 - [C6EGE9_ECOBD]	100.49	7.69	1	2	2	2	273	29.8	10.93
C6ECC2	Inorganic diphosphatase OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_3807 PE=3 SV=1 - [C6ECC2_ECOBD]	95.13	30.68	1	4	5	6	176	19.7	5.17
C6EGB5	Peptidylprolyl isomerase FKBP-type OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_0400 PE=4 SV=1 - [C6EGB5_ECOBD]	95.00	11.73	1	2	2	2	196	20.8	5.05
C6ECH2	Adenylosuccinate synthetase OS=Escherichia coli (strain B / BL21-DE3) GN=purA PE=3 SV=1 - [C6ECH2_ECOBD]	94.31	10.88	1	3	4	5	432	47.3	5.49
C6EJL0	Succinyl-CoA synthetase, beta subunit OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_2933 PE=3 SV=1 - [C6EJL0_ECOBD]	92.27	6.19	1	2	2	2	388	41.4	5.52
C6EBX6	Glucose-6-phosphate 1-dehydrogenase OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_1787 PE=3 SV=1 - [C6EBX6_ECOBD]	90.65	9.57	1	4	4	4	491	55.7	5.76

C6EJL1	Dihydrolipoyllysine-residue succinyltransferase component of 2-oxoglutarate dehydrogenase complex OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_2934 PE=3 SV=1 - [C6EJL1_ECOBD]	90.47	6.67	1	2	2	2	405	44.0	5.81
C6EAL9	Alcohol dehydrogenase GroES domain protein OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_1566 PE=4 SV=1 - [C6EAL9_ECOBD]	86.89	6.07	1	2	2	2	346	37.4	6.38
C6EKL0	Polyphosphate kinase OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_1187 PE=3 SV=1 - [C6EKL0_ECOBD]	79.87	4.36	1	2	2	2	688	80.4	8.92
C6EE36	Ribosomal protein L11 OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_4050 PE=3 SV=1 - [C6EE36_ECOBD]	75.65	16.20	1	2	2	3	142	14.9	9.63
C6EE00	Aspartate kinase OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_4013 PE=4 SV=1 - [C6EE00_ECOBD]	72.24	4.90	1	2	2	2	449	48.5	5.11
C6EBW1	Aspartyl-tRNA synthetase OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_1772 PE=4 SV=1 - [C6EBW1_ECOBD]	72.19	3.90	1	2	2	3	590	65.9	5.69
C6EI04	Glutathionylspermidine amidase OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_0750 PE=4 SV=1 - [C6EI04_ECOBD]	71.67	4.36	1	2	2	2	619	70.5	5.26
C6EKK2	GMP synthase [glutamine-hydrolyzing] OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_1179 PE=3 SV=1 - [C6EKK2_ECOBD]	63.84	7.24	1	2	3	4	525	58.6	5.39
C6EFE8	Glycerol-3-phosphate dehydrogenase OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_0316 PE=3 SV=1 - [C6EFE8_ECOBD]	62.14	4.59	1	2	2	2	501	56.7	7.44
C6EKF6	Glycine hydroxymethyltransferase OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_1133 PE=3 SV=1 - [C6EKF6_ECOBD]	60.73	7.43	1	2	2	2	417	45.3	6.48
C6EBX3	Pyruvate kinase OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_1784 PE=4 SV=1 - [C6EBX3_ECOBD]	60.20	7.50	1	1	3	3	480	51.3	6.68
C6EGH2	Ribosomal protein L17 OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_0457 PE=3 SV=1 - [C6EGH2_ECOBD]	57.55	18.11	1	2	2	2	127	14.4	11.05
C6EJL8	Citrate synthase OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_2941 PE=3 SV=1 - [C6EJL8_ECOBD]	57.31	5.15	1	1	2	2	427	48.0	6.68
C6EHZ8	Cytochrome-c3 hydrogenase OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_0744 PE=4 SV=1 - [C6EHZ8_ECOBD]	57.27	6.53	1	2	2	2	567	62.5	6.28
C6EAB9	2,3,4,5-tetrahydropyridine-2,6-dicarboxylate N-succinyltransferase OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_3453 PE=3 SV=1 - [C6EAB9_ECOBD]	56.81	8.03	1	2	2	2	274	29.9	5.74
C6EHB9	3-oxoacyl-(Acyl-carrier-protein) synthase 2 OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_2506 PE=4 SV=1 - [C6EHB9_ECOBD]	55.41	13.56	1	1	3	3	413	43.0	6.09
C6EKZ0	Adenylate kinase OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_3182 PE=3 SV=1 - [C6EKZ0_ECOBD]	54.54	10.28	1	1	2	2	214	23.6	5.76
C6EB56	6-phosphogluconate dehydrogenase, decarboxylating OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_1630 PE=3 SV=1 - [C6EB56_ECOBD]	50.74	5.98	1	3	3	3	468	51.5	5.07
C6EH63	DEAD/DEAH box helicase domain protein OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_0578 PE=3 SV=1 - [C6EH63_ECOBD]	46.89	3.82	1	2	2	2	629	70.5	8.72
C6EGF8	Ribosomal protein L5 OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_0443 PE=3 SV=1 - [C6EGF8_ECOBD]	44.64	26.26	1	2	3	3	179	20.3	9.48

P02533	Keratin, type I cytoskeletal 14 OS=Homo sapiens GN=KRT14 PE=1 SV=4 - [K1C14_HUMAN]	44.54	3.81	2	1	2	3	472	51.5	5.16
C6EGE5	Ribosomal protein S10 OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_0430 PE=3 SV=1 - [C6EGE5_ECOBD]	39.15	20.39	1	2	2	2	103	11.7	9.69
C6ECB1	Anaerobic ribonucleoside-triphosphate reductase OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_3796 PE=4 SV=1 - [C6ECB1_ECOBD]	36.70	2.11	1	1	2	2	712	80.0	6.83
C6EAB6	Ribosomal protein S2 OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_3450 PE=3 SV=1 - [C6EAB6_ECOBD]	31.42	7.88	1	2	2	3	241	26.7	7.14
C6EBJ7	Peptide chain release factor 3 OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_3646 PE=3 SV=1 - [C6EBJ7_ECOBD]	31.18	4.73	1	1	2	2	529	59.5	5.97
C6EKI4	PepB aminopeptidase OS=Escherichia coli (strain B / BL21-DE3) GN=ECBD_1161 PE=3 SV=1 - [C6EKI4_ECOBD]	30.83	4.45	1	1	2	2	427	46.2	5.77

Tab. S3: Summary of settings of different machine learning models trained

The following abbreviations were used: ML: Machine Learning, RF: Random Forests, GBM: Gradient Boosting Machines.

Feature	ML algorithm	Tuned hyper-parameters	
		Number of estimators	Maximum tree depth
Mol2vec + ProtVec	RF	50	5
Mol2vec	RF	50	5
Mol2vec	GBM	100	3

Tab. S4: Evaluation of machine learning performance (mean \pm standard deviation) on different classes

CV	Model	Class	precision	recall	F-score
		0 (non-toxic)			
CV	Combined Model	1 (less toxic)	0.81 \pm 0.15	0.80 \pm 0.16	0.79 \pm 0.10
		2 (toxic)	0.87 \pm 0.11	0.87 \pm 0.15	0.86 \pm 0.10
			0.92 \pm 0.10	0.87 \pm 0.17	0.88 \pm 0.10
	RF (Mol2vec+ ProtVec)	0 (non-toxic)	0.80 \pm 0.14	0.81 \pm 0.16	0.79 \pm 0.11
		1 (less toxic)	0.89 \pm 0.11	0.90 \pm 0.13	0.88 \pm 0.09
		2 (toxic)	0.91 \pm 0.11	0.83 \pm 0.16	0.86 \pm 0.11
	RF (Mol2vec)	0 (non-toxic)	0.80 \pm 0.15	0.80 \pm 0.16	0.78 \pm 0.10
		1 (less toxic)	0.87 \pm 0.11	0.85 \pm 0.15	0.85 \pm 0.10
		2 (toxic)	0.92 \pm 0.10	0.87 \pm 0.17	0.88 \pm 0.11
	XGB (Mol2vec)	0 (non-toxic)	0.81 \pm 0.14	0.78 \pm 0.16	0.78 \pm 0.11
		1 (less toxic)	0.86 \pm 0.11	0.87 \pm 0.14	0.85 \pm 0.10
		2 (toxic)	0.92 \pm 0.10	0.87 \pm 0.15	0.88 \pm 0.10
80/20	Combined Model	0 (non-toxic)	0.80 \pm 0.15	0.79 \pm 0.17	0.77 \pm 0.11
		1 (less toxic)	0.82 \pm 0.16	0.86 \pm 0.14	0.83 \pm 0.11
		2 (toxic)	0.91 \pm 0.11	0.88 \pm 0.15	0.88 \pm 0.11
	RF (Mol2vec+ ProtVec)	0 (non-toxic)	0.80 \pm 0.17	0.80 \pm 0.16	0.78 \pm 0.12
		1 (less toxic)	0.84 \pm 0.16	0.90 \pm 0.15	0.85 \pm 0.12
		2 (toxic)	0.89 \pm 0.12	0.83 \pm 0.16	0.85 \pm 0.12
	RF (Mol2vec)	0 (non-toxic)	0.79 \pm 0.16	0.79 \pm 0.18	0.77 \pm 0.11
		1 (less toxic)	0.83 \pm 0.16	0.85 \pm 0.14	0.83 \pm 0.10
		2 (toxic)	0.91 \pm 0.11	0.88 \pm 0.15	0.88 \pm 0.11
	XGB (Mol2vec)	0 (non-toxic)	0.81 \pm 0.16	0.78 \pm 0.17	0.78 \pm 0.11
		1 (less toxic)	0.82 \pm 0.16	0.87 \pm 0.13	0.83 \pm 0.10
		2 (toxic)	0.89 \pm 0.12	0.88 \pm 0.14	0.88 \pm 0.11