Construction and comparison of three HMM profiles for the annotation of proteins based on the presence of Kunitz-type protease inhibitor domain

Denys Carbini - LM Bioinformatics - LB1 Project

SUPPLEMENTARY MATERIALS

PFAM	CATH	INTERPRO
1AAP:A	1AAP:A	1AAP:A
1BUN:B	1BUN:B	1BUN:B
1DTX:A	1D0D:A	1D0D:A
1FAK:I	1DTX:A	1DTX:A
1KTH:A	1FAK:I	1FAK:I
1YC0:I	1KTH:A	1KTH:A
1ZR0:B	1YC0:I	1YC0:I
3BYB:A	1ZR0:B	1ZR0:B
3M7Q:B	2UUY:B	2FMA:A
4BQD:A	2W8X:A	2UUY:B
4DTG:K	3AUB:A	3BYB:A
4ISO:B	3BYB:A	3M7Q:B
4NTW:B	3M7Q:B	4BQD:A
4U30:X	4BQD:A	4DTG:K
4U32:X	4DTG:K	4ISO:B
5M4V:A	4ISO:B	4NTW:B
5NX1:D	4NTW:B	4U30:X
5PTI:A	4U30:X	4U32:X
5YV7:A	4U32:X	5M4V:A
6Q61:A	5M4V:A	5NX1:D
6YHY:A	5NX1:D	5PTI:A
	5YV7:A	5YV7:A
	6Q61:A	6Q61:A
		6YHY:A

Table 1: List of PDB ids used to generate the multiple structure alignment and HMMs divided by the external reference used to retrieve them from RCSB PDB

SwissProt ids
sp—C5H8E7—TICK1_RHIAP
sp—G9I929—VKTA_MICTN
sp—O43278—SPIT1_HUMAN
sp—O43291—SPIT2_HUMAN
sp—P00974—BPT1_BOVIN
sp—P00980—VKTHA_DENAN
sp—P00989—VKTH2_BUNMU
sp—P02760—AMBP_HUMAN
sp—P05067—A4_HUMAN
sp—P08592—A4_RAT
sp—P0C1X2—VKTS1_CONST
sp—P10646—TFPI1_HUMAN
sp—P12023—A4_MOUSE
sp—P12111—CO6A3_HUMAN
sp—P17726—TAP_ORNMO
sp—P29216—A4_MACMU
sp—P31713—VKT1_STIHL
sp—P48307—TFPI2_HUMAN
sp—P53601—A4_MACFA
sp—P79307—A4_PIG
sp—P81658—VKTHC_DENAN
sp—Q1RPT0—VKTH4_BUNMU
sp—Q28864—TFPI1_MACMU
sp—Q5IS80—A4_PANTR
sp—Q60495—A4_CAVPO
sp—Q6ITB4—VKT2_OXYMI
sp—Q6ITB5—VKT1_OXYMI
sp—Q6ITB6—VKT2_OXYSC
sp—Q6ITB7—VKT1_OXYSC
sp—Q6ITB9—VKT3_PSEAU
sp—Q75S50—VKTH3_BUNCA
sp—Q8AY44—VKTH7_BUNCA
sp—Q8AY45—VKBH2_BUNCA
sp—Q90WA1—VKT1_PSETT
sp—Q95241—A4_SAISC

Table 2: List of SwissProt ids that have to be removed from the set of positives downloaded from UniProt and then used for the optimization and the testing of the method.

Train	Test	Threshold	Acc	MCC	TPR	PPV	F1 Score
Set 1,2,3	Set 4	1e-06	1.0	1.0	1.0	1.0	1.0
Set 1,2,4	Set 3	1e-06	1.0	1.0	1.0	1.0	1.0
Set 1,3,4	Set 2	1e-06	0.99998999	0.99215177	0.984375	1.0	0.99212598
Set 2,3,4	Set 1	1e-06	0.99998999	0.99215177	0.984375	1.0	0.99212598

Table 3: PFAM summary of the 4-folds cross-validation

Train	Test	Threshold	Acc	MCC	TPR	PPV	F1 Score
Set 1,2,3	Set 4	1e-06	0.99998999	0.99215177	0.984375	1.0	0.99212598
Set 1,2,4	Set 3	1e-06	1.0	1.0	1.0	1.0	1.0
Set 1,3,4	Set 2	1e-05	0.99998999	0.99215177	0.984375	1.0	0.99212598
Set 2,3,4	Set 1	1e-05	0.99997999	0.98436499	0.984375	0.984375	0.984375

Table 4: CATH summary of the 4-folds cross-validation

Train	Test	Threshold	Acc	MCC	TPR	PPV	F1 Score
Set 1,2,3	Set 4	1e-06	0.99998999	0.99215177	0.984375	1.0	0.9921259
Set 1,2,4	Set 3	1e-06	0.99998999	0.99227291	1.0	0.9846153	0.99224806
Set 1,3,4	Set 2	1e-06	0.99997999	0.98424113	0.96875	1.0	0.98412698
Set 2,3,4	Set 1	1e-06	0.99998999	0.99215177	0.984375	1.0	0.99212598

Table 5: INTERPRO summary of the 4-folds cross-validation