Table S2. Key features that selected for HydPred

Hydroxyproline

Rank	Feature	Category	Note
1	G0T	I	Residue pair composition of G and T pair that are being together
2	G3T	I	Residue pair composition of G and T pair that are separated by 3 residues
3	posPWM.8	III	Positive position weighted matrices value at position 8
4	H1G	I	Residue pair composition of H and G pair that are separated by 1 residue
5	H6D	I	Residue pair composition of H and D pair that are separated by 6 residues
6	H2T	I	Residue pair composition of H and T pair that are separated by 2 residues
7	locAA_8_G	IV	Amino acid binary localization encoding for G at position 8
8	G0H	I	Residue pair composition of G and H pair that are attached each other
9	posPWM.11	III	Positive position weighted matrices value at position 11
10	negPWM.9	III	Negative position weighted matrices value at position 9
11	locAA_2_G	IV	Amino acid binary localization encoding for G at position 2
12	locAA_5_G	IV	Amino acid binary localization encoding for G at position 5
13	posPWM.5	III	Positive position weighted matrices value at position 5
14	locAA_3_G	IV	Amino acid binary localization encoding for G at position 3
15	locAA_9_G	IV	Amino acid binary localization encoding for G at position 9
16	H3P	I	Residue pair composition of H and P pair that are separated by 3 residues
17	locAA_11_G	IV	Amino acid binary localization encoding for G at position 11
18	negPWM.6	III	Negative position weighted matrices value at position 6
19	I9P	I	Residue pair composition of I and P pair that are separated by 9 residues
20	negPWM.3	III	Negative position weighted matrices value at position 3
21	negPWM.8	III	Negative position weighted matrices value at position 8
22	H4G	I	Residue pair composition of H and G pair that are separated by 4 residues
23	T	I	Single amino acid composition of T
24	locAA_6_G	IV	Amino acid binary localization encoding for G at position 6
25	I8T	I	Residue pair composition of I and T pair that are separated by 8 residues
26	G3F	I	Residue pair composition of G and F pair that are separated by 3 residues
27	G1E	I	Residue pair composition of G and E pair that are separated by 1 residue
28	I7G	I	Residue pair composition of I and G pair that are separated by 7 residues
29	T5R	I	Residue pair composition of T and R pair that are separated by 5 residues
30	posPWM.2	III	Positive position weighted matrices value at position 2
31	E3G	I	Residue pair composition of E and G pair that are separated by 3 residues
32	G7D	I	Residue pair composition of G and D pair that are separated by 7 residues
33	G9T	I	Residue pair composition of G and T pair that are separated by 9 residues
34	H0N	I	Residue pair composition of H and N pair that are attached each other
35	C	I	Single amino acid composition of C
36	P1H	I	Residue pair composition of P and H pair that are separated by 1 residue
37	G4P	I	Residue pair composition of G and P pair that are separated by 4 residues
38	P2N	I	Residue pair composition of P and N pair that are separated by 2 residues
39	D2D	I	Residue pair composition of D and D pair that are separated by 2 residues
40	N1T	I	Residue pair composition of N and T pair that are separated by 1 residue

41	T6D	I	Residue pair composition of T and D pair that are separated by 6 residues
42	negPWM.12	III	Negative position weighted matrices value at position 12
43	N0G	I	Residue pair composition of N and G pair that are attached each other
44	negPWM.11	III	Negative position weighted matrices value at position 8
45	I4G	I	Residue pair composition of I and G pair that are separated by 4 residues
46	P0G	I	Residue pair composition of P and G pair that are attached each other
47	G6T	I	Residue pair composition of G and T pair that are separated by 6 residues
48	locAA_10_N	IV	Amino acid binary localization encoding for N at position 10
49	GearynAuto_11_2	II	Geary autocorrelation of the molecular volume factor with d value as 2

Hydroxylysine

Ranking	Feature	Category	Note
1	G3E	I	Residue pair composition of G and E pair that are separated by 3 residues
2	posPWM.2	III	Positive position weighted matrices value at position 2
3	G0E	I	Residue pair composition of G and E pair that are attached each other
4	posPWM.11	III	Positive position weighted matrices value at position 11
5	locAA_5_G	IV	Amino acid binary localization encoding for G at position 5
6	G1K	I	Residue pair composition of G and E pair that are seperated by 1 residue
7	posPWM.3	III	Positive position weighted matrices value at position 3
8	K1E	I	Residue pair composition of K and E pair that are separated by 1 residue
9	P3G	I	Residue pair composition of P and G pair that are seperated by 3 residues
10	posPWM.6	III	Positive position weighted matrices value at position 6
11	locAA_8_G	IV	Amino acid binary localization encoding for G at position 8
12	posPWM.8	III	Positive position weighted matrices value at position 8
13	locAA_2_G	III	Amino acid binary localization encoding for G at position 2
14	G6E	I	Residue pair composition of G and E pair that are seperated by 6 residues
15	G4K	I	Residue pair composition of G and K pair that are seperated by 4 residues
16	E2E	I	Residue pair composition of E and E pair that are seperated by 2 residue
17	K0G	I	Residue pair composition of K and G pair that are attached each other
18	K3G	I	Residue pair composition of K and G pair that are seperated by 3 residues
19	negPWM.11	III	Negative position weighted matrices value at position 11
20	locAA_11_G	IV	Amino acid binary localization encoding for G at position 11
21	locAA_9_E	IV	Amino acid binary localization encoding for E at position 9
22	E1G	I	Residue pair composition of E and G pair that are seperated by 1 residue
23	negPWM.2	III	Negative position weighted matrices value at position 2
24	posPWM.5	III	Positive position weighted matrices value at position 5
25	G8G	I	Residue pair composition of G and G pair that are seperated by 8 residues
26	negPWM.8	III	Negative position weighted matrices value at position 8
27	GearyAuto_6_1	II	Geary autocorrelation of the average volumes of residues with d value as 1
28	posPWM.9	III	Positive position weighted matrices value at position 9
29	NorMBAuto_2_6	II	Normalized Moreau-Broto autocorrelation of the average flexibility index with d value as 6

30	P1E	I	Residue pair composition of P and E pair that are seperated by 1 residue
31	P6G	I	Residue pair composition of P and G pair that are seperated by 6 residues
32	G6D	I	Residue pair composition of G and D pair that are seperated by 6 residues
33	K4D	I	Residue pair composition of K and D pair that are seperated by 4 residues
34	P0G	I	Residue pair composition of P and G pair that are attached each other
35	GearyAuto_2_6	II	Geary autocorrelation of the average flexibility index with d value as 6
36	P5K	I	Residue pair composition of P and K pair that are seperated by 5 residues
37	G5G	I	Residue pair composition of G and G pair that are seperated by 5 residues
38	S	I	Single amino acid composition of S
39	P7D	I	Residue pair composition of P and D pair that are seperated by 7 residues
40	G2G	I	Residue pair composition of G and G pair that are seperated by 2 residues
41	P9G	I	Residue pair composition of P and G pair that are seperated by 9 residues
42	negPWM.3	III	Negative position weighted matrices value at position 3
43	GearyAuto_6_2	II	Geary autocorrelation of the average volumes of residues with d value as 2
44	G0L	I	Residue pair composition of G and L pair that are attached each other