Supplementary Material

Table S1. The original values of the seven physicochemical properties for each amino acid.

code	H_1	H_2	V	P_1	P_2	SASA	NCI
A	0.62	-0.5	27.5	8.1	0.046	1.181	0.007187
C	0.29	-1	44.6	5.5	0.128	1.461	-0.03661
D	-0.9	3	40	13	0.105	1.587	-0.02382
E	-0.74	3	62	12.3	0.151	1.862	0.006802
F	1.19	-2.5	115.5	5.2	0.29	2.228	0.037552
G	0.48	0	0	9	0	0.881	0.179052
H	-0.4	-0.5	79	10.4	0.23	2.025	-0.01069
I	1.38	-1.8	93.5	5.2	0.186	1.81	0.021631
K	-1.5	3	100	11.3	0.219	2.258	0.017708
L	1.06	-1.8	93.5	4.9	0.186	1.931	0.051672
M	0.64	-1.3	94.1	5.7	0.221	2.034	0.002683
N	-0.78	2	58.7	11.6	0.134	1.655	0.005392
P	0.12	0	41.9	8	0.131	1.468	0.239531
Q	-0.85	0.2	80.7	10.5	0.18	1.932	0.049211
R	-2.53	3	105	10.5	0.291	2.56	0.043587
S	-0.18	0.3	29.3	9.2	0.062	1.298	0.004627
T	-0.05	-0.4	51.3	8.6	0.108	1.525	0.003352
V	1.08	-1.5	71.5	5.9	0.14	1.645	0.057004
W	0.81	-3.4	145.5	5.4	0.409	2.663	0.037977
Y	0.26	-2.3	117.3	6.2	0.298	2.368	0.023599

H_1, hydrophobicity; H_2, hydrophilicity; V, volume of side chains; P_1, polarity; P_2, polarizability; SASA, solvent accessible surface area; NCI, net charge index of side chains.

Table S2. The prediction results of the test sets based on the non-redundant dataset that was excluded the homologous protein pairs with $\geq$ 40% pairwise sequence identity.

Test set	TP	FN	TN	FP	Sensitivity (%)	Precision (%)	Accuracy (%)
1	2055	183	1759	479	91.82	81.10	85.21
2	2077	161	1747	491	92.81	80.88	85.43
3	1933	305	2016	222	86.37	89.70	88.23
4	1735	503	2085	153	78.33	91.90	85.34
5	1920	318	2044	194	85.79	90.82	88.56
Average	1944	294	1930	308	86.86	86.32	86.55±1.68

TP, true positive; FP, false positive; TN, true negative; FN, false negative; AC, auto covariance; ACC, auto and cross covariance. Psub is the negative dataset non-interacting pairs of non co-localized proteins.

Table S3. The prediction result of non-interacting protein pairs using the model based on the original training set.

Subcellular localization	No. of non-interacting pairs	Correctly predicted pairs
Cytoplasm	8000	6172 (77.15%)
Nucleus	8000	6468 (80.85%)
Mitochondrion	8284	7294 (88.05%)
Endoplasmic reticulum	1953	1349 (69.07%)
Golgi apparatus	300	279 (93.00%)
Peroxisome	171	146 (85.38%)
Vacuole	496	453 (91.33%)
Total	27204	22161 (81.46%)

Table S4. The prediction result of non-interacting protein pairs using the model based on the non-redundant training set.

Subcellular localization	No. of non-interacting pairs	Correctly predicted pairs
Cytoplasm	8000	6305 (78.81%)
Nucleus	8000	5516 (69.00%)
Mitochondrion	8284	7545 (91.08%)
Endoplasmic reticulum	1953	1493 (76.45%)
Golgi apparatus	300	278 (92.67%)
Peroxisome	171	158 (92.4%)
Vacuole	496	457 (92.14)
Total	27204	21572 (79.96%)