

Table S1. The list of *Arabidopsis thaliana* ESTs containing known Cyt-b561 coding sequences

Cyt-b561 coding sequences	No. of ESTs	EST accession numbers
[Single-domain Cyt-b561 proteins]		
At1g14730	1	AI996604
At1g26100	3	BP587463, CF652328, BE522695
At425570	9	AV819134, BP571487, AV788828, BP621624, BX835859, BP611349, BE844714, BP571951, T04807
At5g38630	5	AU238085, BP623885, BP632139, BP663600, BP573945
[Multiple-domain proteins containing the Cyt-b561 domain]		
At3g07570	3	BP576524, BP585760, AV786182
At3g61750	1	CF773415
At5g47530	7	AV567140, AV565093, AV565001, AV562405, AV566417, AV567065, BP573319

Table S2. Loadings of the physico-chemical properties of amino acids for the five principal components^a

Amino acid properties	Loadings for each principal component (% variance)				
	PC1 (40.4)	PC2 (28.7)	PC3 (10.9)	PC4 (8.9)	PC5 (4.2)
Mass	0.34	0.34	0.13	-0.05	0.07
Volume	0.33	0.13	0.17	-0.33	0.19
Surface area	0.39	0.24	0.09	0.08	0.05
Hydrophobicity	0.19	-0.47	-0.03	-0.18	-0.34
Hydrophilicity	-0.38	0.24	0.01	-0.11	0.30
Refractivity	0.36	0.23	0.06	-0.04	-0.20
Isoelectric point	0.03	0.20	0.03	0.85	-0.18
Energy of water to ethanol	-0.39	0.21	-0.17	0.04	0.22
Non-polar surface	0.03	-0.51	-0.12	0.19	0.02
Frequency of alpha-helix	0.12	0.15	-0.76	-0.14	0.07
Frequency of beta-sheet	0.21	0.32	0.20	0.19	0.77
Frequency of reverse-turn	-0.32	0.06	0.52	-0.16	-0.19

^aLoadings reflect the relative contribution of each property to the five principal components.

Table S3. The number of PLS components and the predictive ability of PLS-ACC from the leave-one-out cross validation procedure.

Training sets ^a	Number of PLS components	Q ²
10	1	0.58
20	4	0.74
50	3	0.65
100	5	0.70
200	7	0.74

^aSee Table 1 for the details on each training set.

Table S4. The number of PLS components and the predictive ability of PLS-AA from the leave-one-out cross validation procedure.

Training sets ^a	Number of PLS components	Q ²
10	3	0.74
20	4	0.76
50	4	0.77
100	4	0.91
200	6	0.88

^aSee Table 1 for the details on each training set.

Table S5. The number of PLS components and the predictive ability of PLS-ACC and PLS-AA from the leave- one-out cross validation procedure obtained from Cyt-b561 training.

Training datasets	Number of PLS components	Q ²
[PLS-ACC]		
Arabidopsis only	4	0.65
Plants	4	0.62
Plants and animals	6	0.57
[PLS-AA]		
Arabidopsis only	3	0.62
Plants	4	0.82
Plants and animals	4	0.91