

Supporting Information

Niwa et al. 10.1073/pnas.0811922106

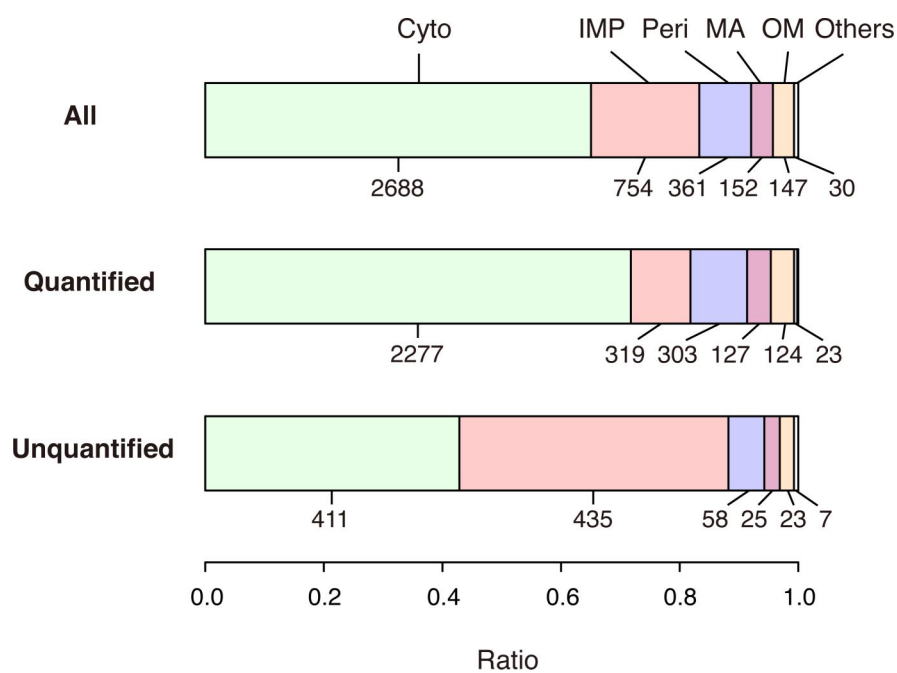


Fig. S1. Subcellular locations for all of the quantified and unquantified proteins. The number of proteins in each location is shown below. Cyto, cytoplasmic proteins; IMP, integral membrane proteins; Peri, periplasmic proteins; MA, membrane anchored proteins; OM, outer membrane lipoproteins and β -barrel proteins.

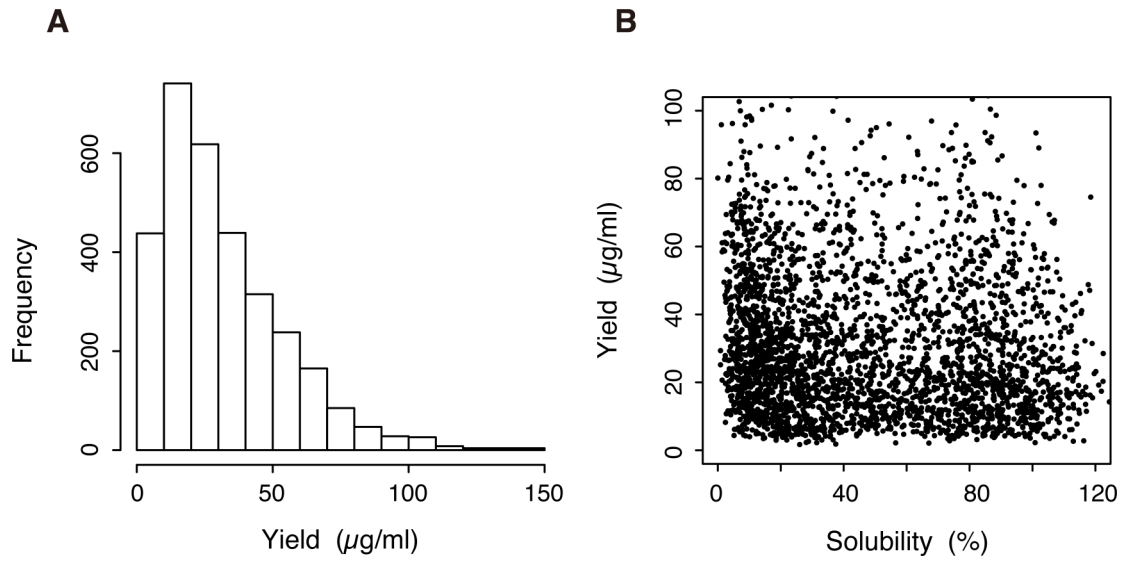


Fig. S2. (A) Histogram of synthesis yields for all quantified proteins, including integral membrane proteins ($n = 3,173$). Average was $\approx 33 \mu\text{g/mL}$. (B) Scatter plot of synthesis yield versus solubility.

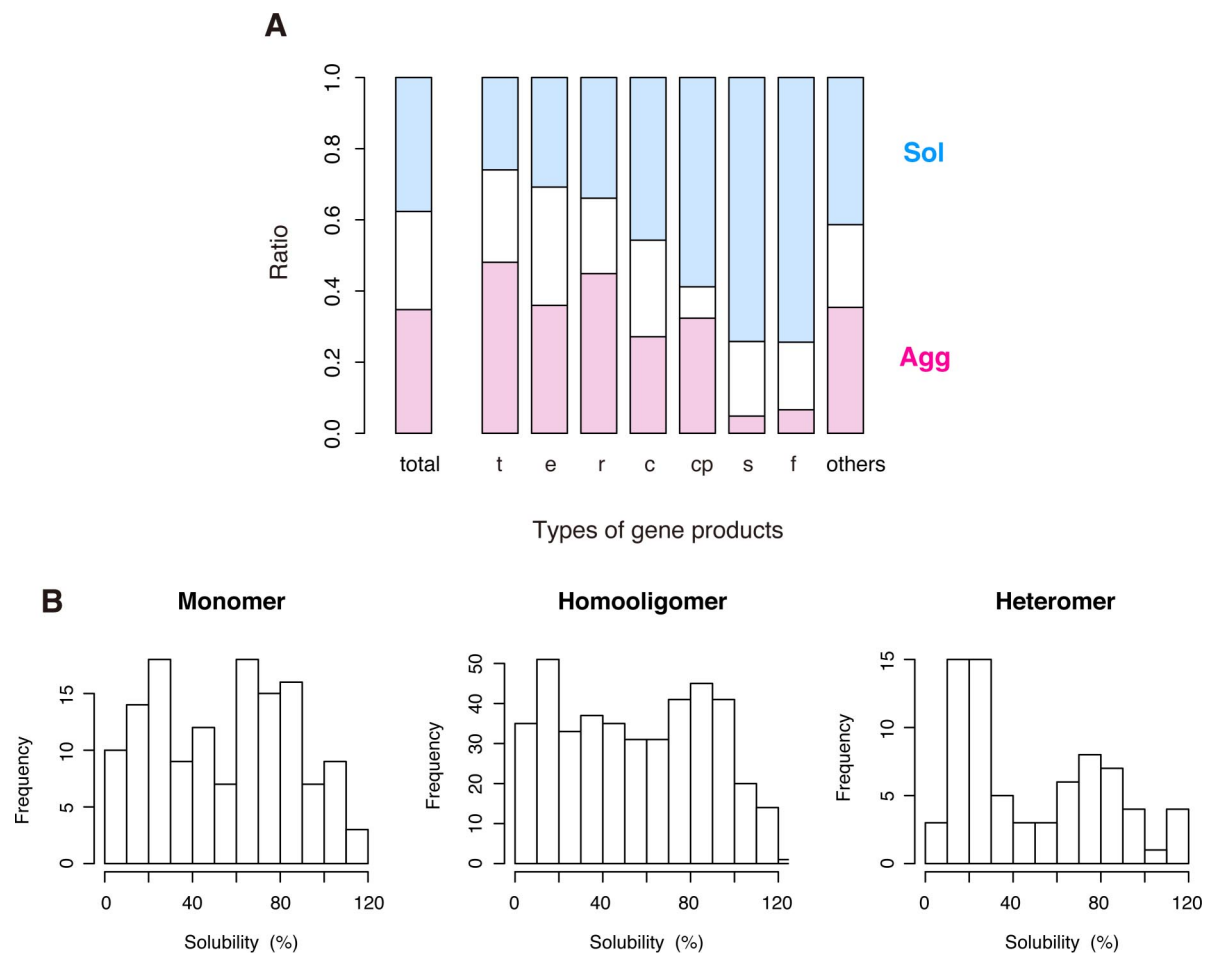


Fig. 55. (A) Relationship between the solubility and the protein function for cytoplasmic proteins. Functional assignments of gene products, derived from GenoBase, are shown as follows: t, transporter; e, enzyme; r, regulator; c, carrier; cp, cell process; f, factor; s, structural component. All predicted proteins (predicted enzyme, predicted transporter, and so on) are included in this analysis. The "others" category includes "pseudogene in common," "phages/IS in common," "partial information," and "unknown function." The "membrane" and "lipoprotein" categories are not shown, because their numbers were too small for the analysis. The f category contains transcription and translation factors, chaperones, and proteases. Note that the enrichment of the Sol groups in the s category was mainly attributed to the presence of a number of ribosomal proteins. The total number for each column is as follows: t, 104; e, 1104; r, 283; c, 70; cp, 34; f, 121; s, 62; others, 486. (B) Histograms of solubility for each oligomeric state. Information about the oligomeric states (Monomer, Homooligomer, and Heteromer) was obtained from the *SUBUNIT* annotation in Uniprot database (www.uniprot.org/). The total number of each annotation is as follows: Monomer, 138; Homooligomer, 415; Heteromer, 74.

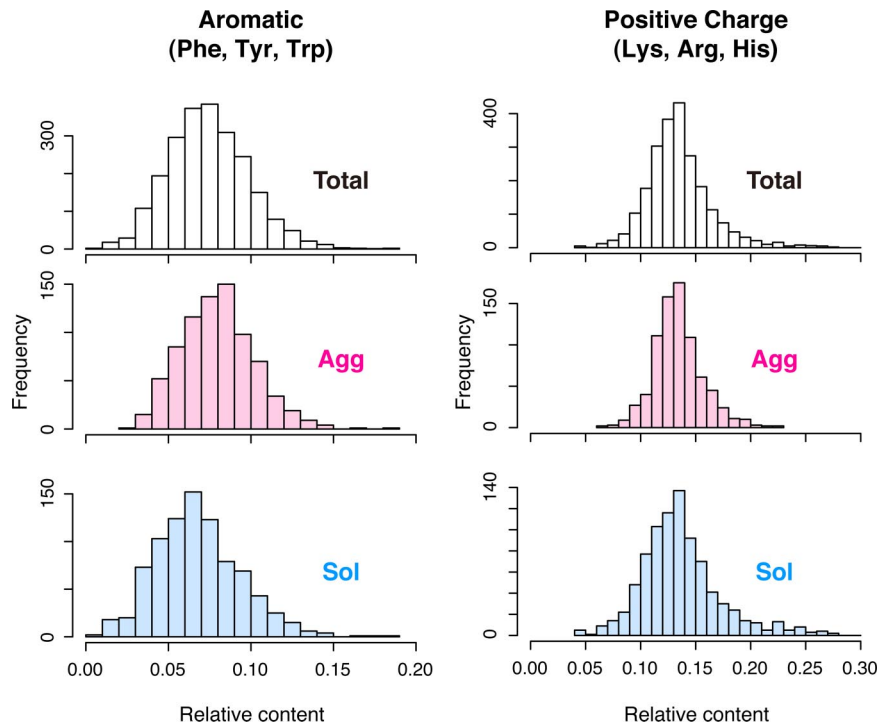


Fig. S6. Histograms of the relative contents of aromatic residues (Phe, Tyr, and Trp) (*Left*) and positively charged residues (Lys, Arg, and His) (*Right*) in the Total, Agg, and Sol groups (cytoplasmic proteins only). The statistical test is shown in [Table S1](#).

Table S1. The results of Welch's *t* test between the Agg and Sol groups

Value	<i>t</i>	d.f.	<i>P</i>
Molecular mass	16.844	1,498	1.968E−58
Amino acid contents			
Negative charge (Asp, Glu)	16.420	1,439	1.149E−55
Hydrophobic (Val, Leu, Ile)	0.514	1,588	6.072E−01
Aromatic (Tyr, Phe, Trp)	10.102	1,640	2.552E−23
Positive charge (Lys, Arg, His)	2.120	1,377	3.419E−02
PSIPRED			
Coil	0.268	1,569	7.891E−01
Helix	2.397	1,522	1.664E−02
Sheet	3.653	1,505	2.680E−04

d.f., degree of freedom.

Table S2. SCOP fold name and the number of each fold

Abbreviation	SCOP fold name	Number			χ^2 Test		
		Total	Agg	Sol	χ^2	d.f.	P
c94	Periplasmic binding protein-like II	42	35	1	26.169	1	3.13E-07
c67	PLP-dependent transferases	35	25	2	15.057	1	1.04E-04
a4	DNA/RNA-binding 3-helical bundle	160	95	31	24.810	1	6.33E-07
c1	TIM β/α -barrel	123	65	18	20.150	1	7.16E-06
c3	FAD/NAD(P)-binding domain	26	13	4	2.677	1	1.02E-01
c55	Ribonuclease H-like motif	44	20	12	0.659	1	4.17E-01
c37	P-loop containing nucleoside triphosphate hydrolases	198	88	39	12.662	1	3.73E-04
c66	S-adenosyl-L-methionine-dependent methyltransferases	39	15	7	1.305	1	2.53E-01
c2	NAD(P)-binding Rossmann-fold domains	86	33	28	0.000	1	1.00E + 00
c26	Adenine nucleotide alpha hydrolase-like	24	7	8	0.091	1	7.63E-01
d58	Ferredoxin-like	63	18	23	1.297	1	2.55E-01
a35	λ -Repressor-like DNA-binding domains	24	6	8	0.314	1	5.75E-01
c72	Ribokinase-like	21	5	7	0.314	1	5.75E-01
c56	Phosphorylase/hydrolase-like	23	5	12	3.202	1	7.35E-02
b40	OB-fold	34	5	23	13.453	1	2.45E-04
c23	Flavodoxin-like	68	9	44	28.589	1	8.95E-08
c47	Thioredoxin fold	20	2	15	10.613	1	1.12E-03
Total cytoplasmic proteins for SCOP analysis		2,081	781	669			

d.f., degree of freedom.