CHEMBIOCHEM

Supporting Information

DOI: 10.1002/cbic.200800810

Supporting Information

for

The Minimal Size of Liposome-Based Model Cells Brings about a Remarkably Enhanced Entrapment and Protein Synthesis

Tereza Pereira de Souza, Pasquale Stano, and Pier Luigi Luisi*

Content

1. ENTRAPMENT STATISTICS

1.1 The general case – empty and filled vesicles

The Poisson distribution can be used to model the entrapment to water-soluble molecules inside lipid vesicles[1,2].

The a priori probability of entrapment of a molecule ***A*** in a vesicle will be $p = V/V_t$, where V is the volume of a vesicle and V_t the total volume of the solution. The average number of entrapped molecule ***A*** will be $a = p \times N_t$, where N_t is the total number of molecules ***A*** in the volume V_t. Once a is known, it is possible to calculate the Poisson probability of entrapping n molecules of the compound ***A*** in that vesicle (Eq. 1).

(Eq. 1) $$\wp(a,n) = e^{-a}\frac{a^n}{n!}$$

In the following, we will adopt the Poisson distribution in order to calculate the probability of co-entrapping one or more molecules of different molecular species in the same vesicle. Only two assumptions will be used in order to develop a simple mathematical analysis, namely: (*1*) the entrapment events are always considered independent events in statistical sense; (*2*) the volume of entrapped molecules is considered negligible with respect to vesicle volumes.

Given a system composed by k different molecular species (k = 1 to N, N being the number of different species), we first define the Poisson probability for the entrapment of k-th component in n_k copies inside a vesicle (Eq. 2):

(Eq. 2) $$\wp(a_k,n_k) = e^{-a_k}\frac{{a_k}^{n_k}}{n_k!}$$

where a_k is the average number of molecules of *k*-th species. From Eq. 2 it is possible to calculate the probability of finding n_k = 0, 1, .. etc molecules of *k*-th species in a vesicle of certain size. "Empty" vesicles will have $n_k = 0$, whereas "filled" vesicles will have $n_k \geq 1$.
The value a_k depends on the vesicle volume and on the concentration of *k*-th specie, according to Eq. 3:

(Eq. 3) $$a_k = C_k \cdot V = C_k \cdot \frac{4}{3}\pi \left(R - d\right)^3$$

C_k being the molecular concentration of *k*-th species and *V* the vesicle volume. For *spherical unilamellar* vesicles, the vesicle internal volume can be calculated from its external radius *R* (an observable quantity), and from the bilyer thickness *d* (for POPC vesicles, d = 3.8 nm)[3].
The Poisson distribution function is normalized to 1 (Eq. 4).

(Eq. 4) $$\sum_{n_k=0}^{\infty} \wp(a_k, n_k) = 1$$

The latter equation can be rewritten as sum of two contributions (Eq. 5),

(Eq. 5) $$e^{-a_k}\left(1 + \sum_{n_k=1}^{\infty} \frac{{a_k}^{n_k}}{n_k!}\right) = 1$$

since $\frac{{a_k}^0}{0!} = 1$.
It follows that the probability of finding a "filled" vesicles ($n_k \geq 1$) is easily calculated as the difference between 1 and the probability of finding an "empty" vesicle. (Eq. 6)

(Eq. 6) $$\wp\left(a_k, n_k \geq 1\right) = 1 - e^{-a_k}$$

According to the hypothesis of independent entrapment events, the cumulative probability of finding "filled" vesicles containing at least 1 component of each molecular species, can be written as product of the probabilities referred to each single component, as shown in Eq. 7.

(Eq. 7) $$\wp_c = \prod_{k=1}^{N}\left(1 - e^{-a_k}\right)$$

This product is extended to all molecular components of the system. In the case under study, the N = 83 components are the *egfp* gene, the enzymes, t-RNAs, ribosomes, amino acids, ATP, etc., which are needed for the transcription-translation reaction (see Supplementary Table 1). From the numerical viewpoint, however, those components with high C_k (> 100 μM) have probability ~ 1 to be entrapped inside vesicles of any size (R > 30 nm), and therefore do not contribute to the cumulative probability calculated in Eq. 7. In other words, the product in Eq. 7 is extended only to those molecules present at lower concentrations. In PURE system, all macromolecular components are present at C_k < 10 μM, whereas low molecular-weight compounds' concentrations are higher than 100 μM. Eq. 7 is therefore limited to PURE system macromolecules.

As shown in Eq. 3, the average number of molecules a_k can be expressed as a function of vesicle (internal) radius. By combining Eqs. 3 and 7, the size-dependence of cumulative probability becomes evident (Eq. 8)

(Eq. 8) $$\wp_c(R) = \prod_{k=1}^{N}\left[1 - \exp\left(-\frac{4\pi C_k}{3}(R-d)^3\right)\right]$$

A plot of $\wp_c(R)$ versus the vesicle external radius R will produce typical sigmoidal curves (see Figure 4 in the main text and Supplementary Figures 6 and 7). The shape of the sigmoidal curves derived from Eq. 8 depends on two parameters: (*i*) the number N of different species that must be simultaneously trapped inside a vesicle; (*ii*) their concentrations, C_k.

1.2 Vesicles are considered "filled" only if $n_k \geq j$.

In paragraph 1.1, "filled" vesicles are defined when $n_k \geq 1$. This definition also implies that, at the aim of this analysis, "filled" vesicles are also functional. This means that even a single molecule of a *k*-th component may allow the occurrence of transcription-translation reactions. However, a more realistic case is when a functional vesicle is obtained only if at least *j* molecules of every *k*-species are simultaneously present inside a vesicle. In other words, there is a numeric threshold for the concentration of molecular components inside vesicles, in order to obtain an EGFP-expressing vesicle.

Eq. 4 can be rewritten as a sum of two contributions, as shown in Eq. 9:

(Eq. 9) $$e^{-a_k}\left(\sum_{n_k=0}^{j-1}\frac{{a_k}^{n_k}}{n_k!}+\sum_{n_k=j}^{\infty}\frac{{a_k}^{n_k}}{n_k!}\right)=1$$

Where the left-hand term in parenthesis concerns the probability of finding less than *j* molecules of the *k*-th chemical species inside a vesicle, and the right-hand term concerns the probability of finding at least *j* molecules of the *k*-th species inside a vesicle. Eq. 9 can be condensed into Eq. 10 and rearranges into Eq. 11:

(Eq. 10) $$e^{-a_k}\left(F_{j-1}+F_j\right)=1$$

(Eq. 11) $$\wp\left(a_k, n_k \geq j\right)=1-F_{j-1}e^{-a_k}$$

where $F_{j\text{-}1}$ is the cumulative probability of finding a vesicle with 0, 1, …*j*-1 copies of the *k*-th component, whereas F_j is the cumulative probability of finding a vesicle with at least *j* copies of the *k*-th component. By comparing Eq. 11 with Eq. 6, it is clear that since $F_{j\text{-}1} > 1$,

F_{j-1} acts to decrease (as intuitively expected) the probability of finding a "filled" vesicle (in this case, "filled" means $n_k \geq j$).

1.3 Association of the *N* components into *N'* aggregates (*N'* < *N*).

In this case, the product shown in Eq. 7 has a minor number of factors (*N*' instead of *N* factors, all factors being < 1) and therefore the cumulative probability is higher than when *N* factors are considered. Clearly, the entrapment of a set of *N*' different molecules is more probable.

1.4 Critical discussions on the assumptions

The first assumption concerns the statistical independence of multiple entrapment events. This assumption cannot be falsified *a priori*. This hypothesis, which is certainly the simplest one, can instead be checked *a posteriori*, by evaluating how experimental data deviates from it. This is exactly the point raised in this work, i.e., the multiple entrapment of macromolecules required for the transcription-translation reaction does not follow simple Poisson statistics for independent events. The reasons for this behavior are not known. We have – however – tried to clarify that simple aggregation of macromolecules (in clusters) before the entrapment is possible but still not explaining observed data. Therefore, we have discussed alternative explanation (see main text).

The second assumption concerns the volume of macromolecules. Their volume has been considered negligible. This has two consequences: firstly, it allows the use of Poisson distribution, which is a distribution with no upper limit (see Eq. 1; *n* can be large at will). If an upper limit to the number of macromolecules is set, other statistical distributions (e.g., beta distribution) has to be used. From the numerical viewpoint, however, the Poisson distribution is perfectly valid even if the macromolecules' volume is considered, since the macromolecules are present at low concentrations and therefore the average occupied volume (inside a vesicle) is very low also for small vesicles. The second consequence is related to the first, and it deals with conditional probabilities. Since the space inside a vesicle is limited, the entrapment molecules will depend on the amount of *free* space. In

other words, the vesicle volume used in Eq. 3 should not be considered constant, but a function of volume occupation by already-entrapped molecules. It is easy to see, however, that our approach actually gives probability that are higher of those obtained when volumes of macromolecules are considered and conditional probabilities are used. Therefore, if our approach gives infinitesimal probability values, these have to be considered as approximated by excess. The third consequence is that if macromolecular volume is negligible, also the effects of excluded volumes (~ 4 times the molecule volume) are negligible. Again, these effects would further reduce the probability of entrapment for reasons similar to those discussed above, and therefore would further decrease the cumulative probability value of finding vesicles which entrap several macromolecular species. In conclusion, neglecting the molecular volume has as consequence of increase the calculated probability of finding a vesicle will all components inside; therefore, our conclusions are not affected by assumption (*2*).

In conclusion, with the exception of assumption (*1*) – which is actually the hypothesis we want to test, the assumptions (*2*) allows to simple model the statistics of co-entrapment of several macromolecule compounds into vesicles. Due to the fact that assumption (*2*) actually increases the probability values for any vesicle size, our conclusions result strengthened.

2. DLS INTENSITY-WEIGHTED TO NUMBER-WEIGHTED SIZE DISTRIBUTION TRANSFORMATION

Dynamic light scattering measurements produce the so-called *intensity-weighted* size distribution, i.e., large particles dominate the size distribution profile due to their higher efficiency in scattering light. In order to extract the *number-weighted* size distribution (where particles are represented proportionally to their numerical amount in stead of their scattered intensity), the rough DLS profile must be manipulated numerically to take into account: (*i*) the scattering volume, which depends on the vesicle size; (*ii*) the vesicle shape, i.e., the "form factor"; (*iii*) the effect of inter-vesicle interaction (for a full discussion, see Ref. 8, and references therein).

Figure 9 shows the procedure in a schematic way. After the usual regularized Laplace inversion (CONTIN) of the experimentally recorded correlation function ($g_2(\tau)$), the resultant *intensity-weighted* particle size distribution has been fitted by a Schulz function (a log-normal function works well too), so that a size distribution curve can be generated analytically. Data are transformed according to Eq. 12

(Eq. 12) $$N_i \propto \frac{G_{Schulz,i}(R)}{V_{S,i}^2 \cdot P(R_i, q)}$$

where, i is the i-th size class; $G_{\mathrm{Schulz},i}(R)$ is the Schulz function that fits the experimental *intensity-weighted* distribution $G_{\mathrm{i}}(R)$; $V_{S,i} = \frac{4}{3}\pi\left(3R^2 d - 3Rd^2 + d^3\right)$ is the scattering volume (R being the external vesicle radius, and d the bilayer thickness, taken[3] as 3.8 nm); $P(R_i,q)$ is the "form factor" that take into account the size and angular dependence of scattered light. In the case of spherical unilamellar vesicles, the Rayleigh-Gans-Debye form factor is given in Eq. 13.

(Eq. 13)
$$P_i(q,d) = \left\{ \frac{3}{(qR_i)^3 - [q(R_i - d)]^3} \right\}^2 \times$$
$$\times \left\{ \sin(qR_i) - qR_i \cos(qR_i) - \sin[q(R_i - d)] + q(R_i - d)\cos[q(R_i - d)] \right\}^2$$

where q represent the module of the scattering vector, $q = (4\pi n / \lambda)\sin(\theta/2)$ (n is the solvent refractive index, λ the laser wavelength, θ is the scattering angle; 1.33, 633 nm, and 90°, respectively).

The Schulz function has been introduced in order to carry out the transformation indicated in Eq. 12 on an analytical curve containing a large number of points, rather than on the experimental CONTIN size distribution, composed by 32 discrete points. After the intensity-to-number transformation, singular points (data at R = 156 nm, corresponding to $P_i(R,q) \rightarrow 0$) have been removed and substituted by the average between the two adjacent values, and re-normalized to one. Similar procedures have been already published (see references 170-177 of the previously cited review,[8] as well as our previous publication,[9] and in the further reference[10]). Figure 10 shows the size dependence of the form factor (at 90° only), whereas in Figure 11 we report the not-manipulated instrumental *intensity-weighted* size distributions (CONTIN) of V2 and V3 vesicles.

3. SUPPLEMENTARY TABLES AND FIGURES

Supplementary Table 1. Composition of PURE system[a] used in this study,[b] inclusive of *egfp* plasmid and calculated concentrations of the t-RNA set.[c]

k	**Name**	**Abbreviation**	**Concentration (C_k), μM**
	EGFP gene (1)		
1	pWM-T7-EGFP	*egfp*	0.02
	Amino acyl tRNA synthases (20)		
2	Alanyl-tRNA synthetase	Ala RS	0.72
3	Arginyl-tRNA synthetase	Arg RS	0.03
4	Asparaginyl-tRNA synthetase	Asn RS	0.42
5	Aspartyl-tRNA synthetase	Asp RS	0.12
6	Cysteinyl-tRNA synthetase	Cys RS	0.02
7	Glutaminyl-tRNA synthetase	Gln RS	0.06
8	Glutamyl-tRNA synthetase	Glu RS	0.23
9	Glycyl-tRNA synthetase alpha subunit	Gly RS	0.28
10	Histidyl-tRNA synthetase	His RS	0.02
11	Isoleucyl-tRNA synthetase	Ile RS	0.38
12	Leucyl-tRNA synthetase	Leu RS	0.04
13	Lysyl-tRNA synthetase	Lys RS	0.11
14	Methionyl-tRNA synthetase	Met RS	0.03
15	Phenylalanyl-tRNA synthetase alpha chain	Phe RS	0.45
16	Prolyl-tRNA synthetase	Pro RS	0.16
17	Seryl-tRNA synthetase	Ser RS	0.04
18	Threonyl-tRNA synthetase	Thr RS	0.08
19	Tryptophanyl-tRNA synthetase	Trp RS	0.03
20	Tyrosyl-tRNA synthetase	Tyr RS	0.01
21	Valyl-tRNA synthetase	Val RS	0.02
	Other enzymes (5)		
22	Inorganic pyrophosphatase	PPiase	0.05
23	Nucleoside diphosphate kinase	NDPK	0.06
24	Creatine kinase B-type	CreK	0.09
25	Adenylate kinase isoenzyme 1 (myokinase)	MyoK	0.14
26	DNA-directed RNA polymerase	T7 RNA Pol	0.10
	Ribosomes and translation factors (11)		
27	Ribosome		1.20
28	Methionyl-tRNA formyltransferase	MetTForm	0.59
29	Translation initiation factor IF-1	IF1	1.23

30	Translation initiation factor IF-2	IF2	0.41
31	Translation initiation factor IF-3	IF3	0.49
32	Peptide chain release factor 1	RF1	0.25
33	Peptide chain release factor 3	RF3	0.17
34	Ribosome recycling factor	RRF	0.48
35	Elongation factor G	EF-G	0.65
36	Elongation factor Ts	EF-Ts	1.65
37	Elongation factor Tu	EF-Tu	2.32
	Calculated t-RNA set (46)[c]		
38	Ala1 tRNA (UGC vs GCU, GCA, GCG)		4.45
39	Ala2 tRNA (GGC vs GCC)		0.84
40	Arg2 tRNA (AGC vs CGU, CGC, CGA)		6.51
41	Arg3 tRNA (CCG vs CGG)		0.87
42	Arg4 tRNA (UCU vs AGA)		1.19
43	Arg5 tRNA (CCU vs AGG)		0.58
44	Asn tRNA (GUU vs AAC, AAU)		1.63
45	Asp1 tRNA (GUC vs GAC, GAU)		3.28
46	Cys tRNA (GCA vs UGC, UGU)		2.17
47	Gln1 tRNA (UUG vs CAA)		1.05
48	Gln2 tRNA (CUG vs CAG)		1.21
49	Glu2 tRNA (UUC vs GAA, GAG)		6.46
50	Gly1 tRNA (CCC vs GGG)		1.46
51	Gly2 tRNA (UCC vs GGA, GGG)		1.46
52	Gly3 tRNA (GCG vs GGC, GGU)		5.97
53	His tRNA (GUG vs CAC, CAU)		0.87
54	Ile1 tRNA (GAU vs AUC, AUU)		2.38
55	Ile2 tRNA (CAU vs AUA)		2.38
56	Leu1 tRNA (CAG vs CUG)		6.12
57	Leu2 tRNA (GAG vs CUC, CUU)		1.29
58	Leu3 tRNA (UAG vs CUA, CUG)		0.91
59	Leu4 tRNA (CAA vs UUG)		2.62
60	Leu5 tRNA (UAA vs UUA, UUG)		1.41
61	Lys tRNA (UUU vs AAA, AAG)		2.63
62	Metf1 tRNA (CAU vs AUG)		1.66
63	Metf2 tRNA (CAU vs AUG)		0.98
64	Metm tRNA (CAU vs AUG)		0.97
65	Phe tRNA (GAA vs UUC, UUU)		1.42
66	Pro1 tRNA (CGG vs CCG)		1.23
67	Pro2 tRNA (GGG vs CCC, CCU)		0.99
68	Pro3 Trna (UGG vs CCA, CCU, CCG)		0.80
69	Sec tRNA (UCA vs UGA)		0.30
70	Ser1 tRNA (UGA vs UCA, UCU, UCG)		1.77
71	Ser2 tRNA (CGA vs UCG)		0.47
72	Ser3 tRNA (GCU vs AGC, AGU)		1.93
73	Ser5 tRNA (GGA vs UCC, UCU)		1.05

74	Thr1 tRNA (GGU vs ACC, ACU)		0.14
75	Thr2 tRNA (CGU vs ACG)		0.74
76	Thr3 tRNA (GGU vs ACC, ACU)		1.50
77	Thr4 tRNA (UGU vs ACA, ACU, ACG)		1.25
78	Trp tRNA (CCA vs UGG)		1.29
79	Tyr1 tRNA (GUA vs UAC, UAU)		1.05
80	Tyr2 tRNA (GUA vs UAC, UAU)		1.73
81	Val1 tRNA (UAC vs GUA, GUG, GUU)		5.26
82	Val2A tRNA (GAC vs GUC, GUU)		0.86
83	Val2B tRNA (GAC vs GUC, GUU)		0.87

[a] Data taken from references[4,5].

[b] When needed, the actual concentrations used in this study have been substituted to the values reported in previous publications[6].

[c] t-RNA set of PURE system is supplied as tRNAmix (Roche) of unknown composition and concentration. Concentrations of individual t-RNA species have been tentatively calculated on the basis of published data on average t-RNA composition in *E. coli*[7]. Each t-RNA species has been reported in the table with its anticodon vs recognized codon(s).

Supplementary Table 2. Summary of EGFP yields in lipid vesicles.

Entry #	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Transcription-translation kit	E. coli extracts	E. coli extracts	E. coli extracts	E. coli extracts	PURE system	PURE system	PURE system	PURE system
Lipid [a]	POPC	POPC/CH	POPC	POPC	POPC	POPC	POPC	POPC
Lipid composition		8/2						
[Lipid], mM	10	6.6	15	8.5	15	8.5	15	15
Liposome class [b]	V1	V1	V2	V3	V1	V1	V2	V2
Liposome formation method [c]	EI	EI	EI VET100	EI VSEC	EI	EI	EI VET100	EI VET200
Radius,[d] nm	From 200 to 2000	150 and 1000	102 ± 23	124 ± 21	From 1000 to 5000	100 and 1000	112 ± 11	100 and 500
Polydispersity [e]	0.25	0.25	0.10 ± 0.03	0.16 ± 0.05	0.16	0.24	0.12 ± 0.02	0.24
EGFP relative yield (%) [f]	1.2	3.8	10.0 ± 7.0	0.3 ± 0.1	1.2	0.5	9.9 ± 5.2	7.4
Internal vesicle volume (%) [g]	n.a.	n.a.	10.2	7.2	n.a.	n.a.	11.3	n.a.
pmol EGFP/mmol lipid	186 ± 175	481 ± 242	309 ± 211	27 ± 6	11	9	103 ± 45	71
EGFP molecules/vesicle	n.a.	n.a.	0.11 ± 0.10	0.01 ± 0.006	n.a.	n.a.	0.05 ± 0.02	n.a.
Number of experiments (replicates) [h]	2	2	3	3	1	1	3	1

[a] POPC = 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphatidylcholine; CH = cholesterol

[b] V1 = vesicles obtained by injecting lipids in ethanol, at high concentration; V2 = vesicles obtained by extrusion of V1; V3 = vesicles obtained directly by injecting lipids in ethanol, at lower concentrations (see Figure 1 of the main text).

[c] EI = ethanol injection method; VET100 (orVET200) = vesicles obtained by the extrusion technique, using 100-nm (or 200-nm) pore filters; VSEC = vesicles obtained after size exclusion chromatography

[d] Hydrodynamic radius as calculated by cumulant analysis of dynamic light scattering (DLS) data. Polydisperse samples are indicated by the size range (e.g., from 200 to 2000 nm); bimodal (two-peaks) samples are reported by indicating the positions of the two peaks (e.g., 150 and 1000 nm); monomodal samples are indicated by the peak position. Data are ± standard deviation *between* replicate experiments (last line); (i.e., the standard deviation of the measurement is less than 2 nm).

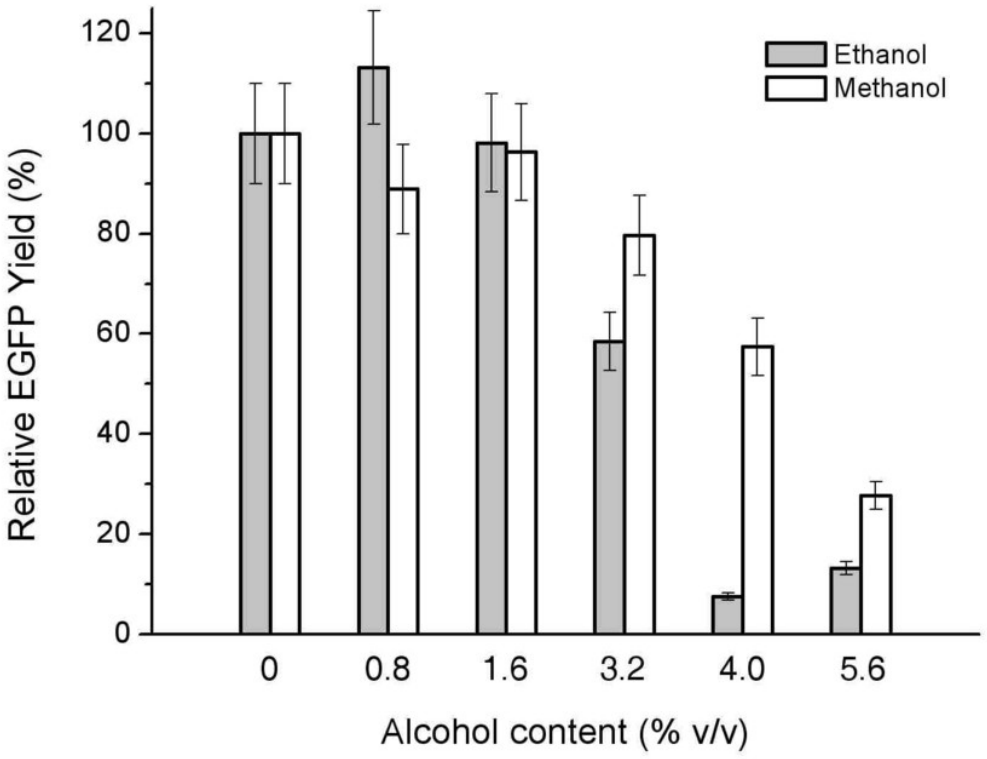
[e] Polydispersity index as calculated by the cumulant analysis of DLS data.

[f] Relative EGFP yield (%) is calculated as the ratio between fluorescence due to internal synthesis and fluorescence due to total (internal + external) synthesis. See also Supplementary Figure 8.

[g] Internal vesicle volume has been calculated only for monodisperse samples, considering: (*i*) unilamellar spherical vesicles; (*ii*) bilayer thickness: 3.8 nm;[3] (*iii*) POPC headgroup area: 0.72 nm^2.[3]

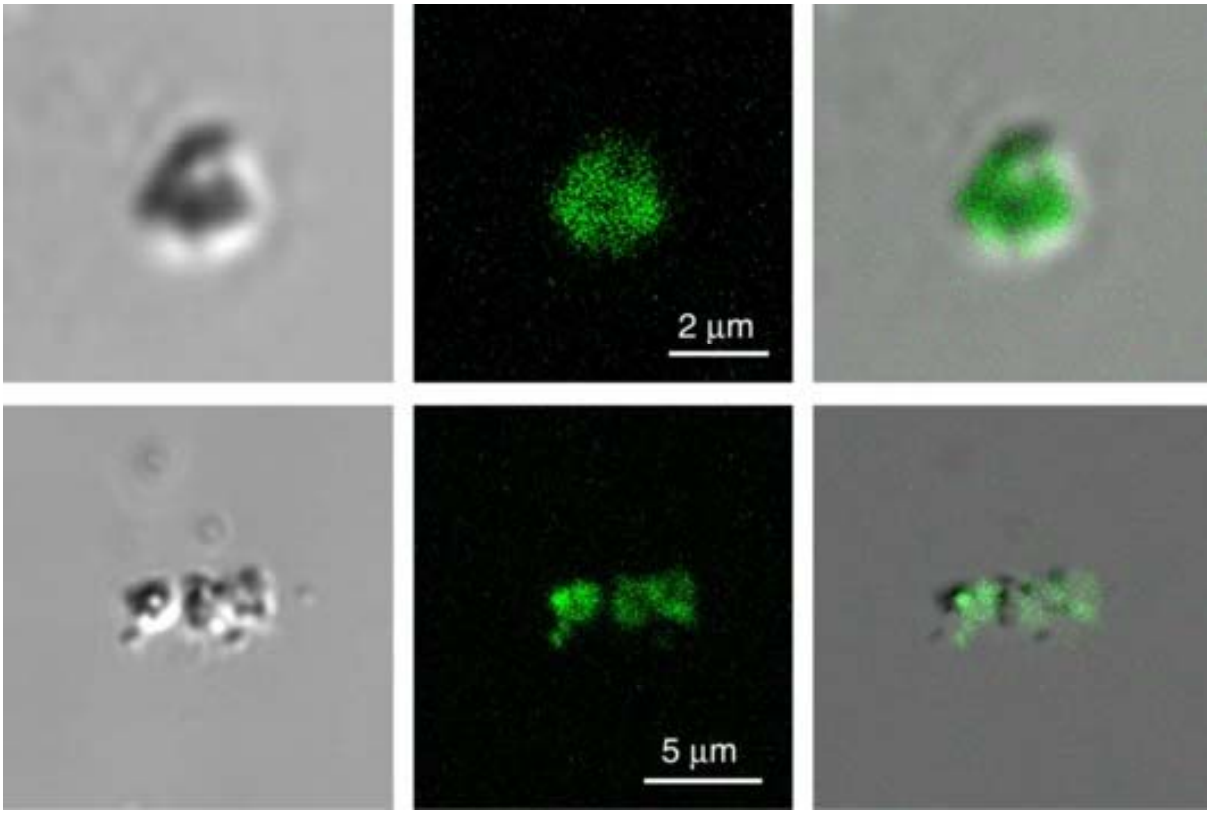
[h] Number of experiments (replicates) that have been used to construct Supplementary Table 2. Additional preliminary experiments were also performed to optimize liposome formation and EGFP detection. Data reported as mean value ± standard deviation are based on the measure of replicates, in order to show reproducibility of experiments.

Supplementary Figure 1



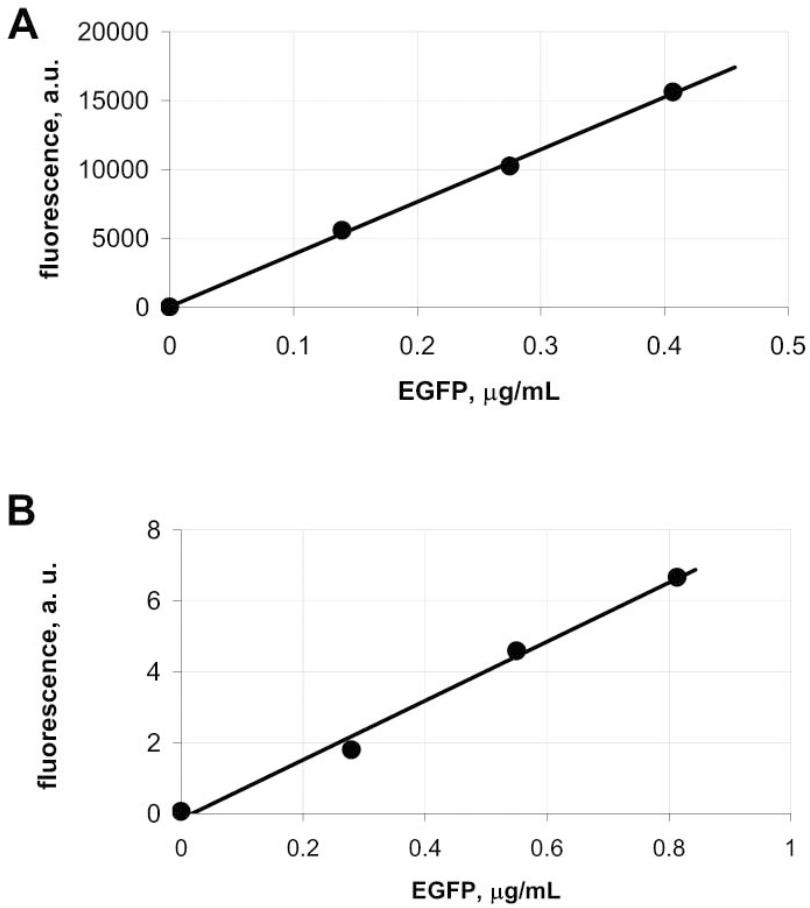
Supplementary Figure 1. Variation of EGFP productivity in *E. coli* cell extracts as a function of the ethanol or methanol concentrations.

Supplementary Figure 2



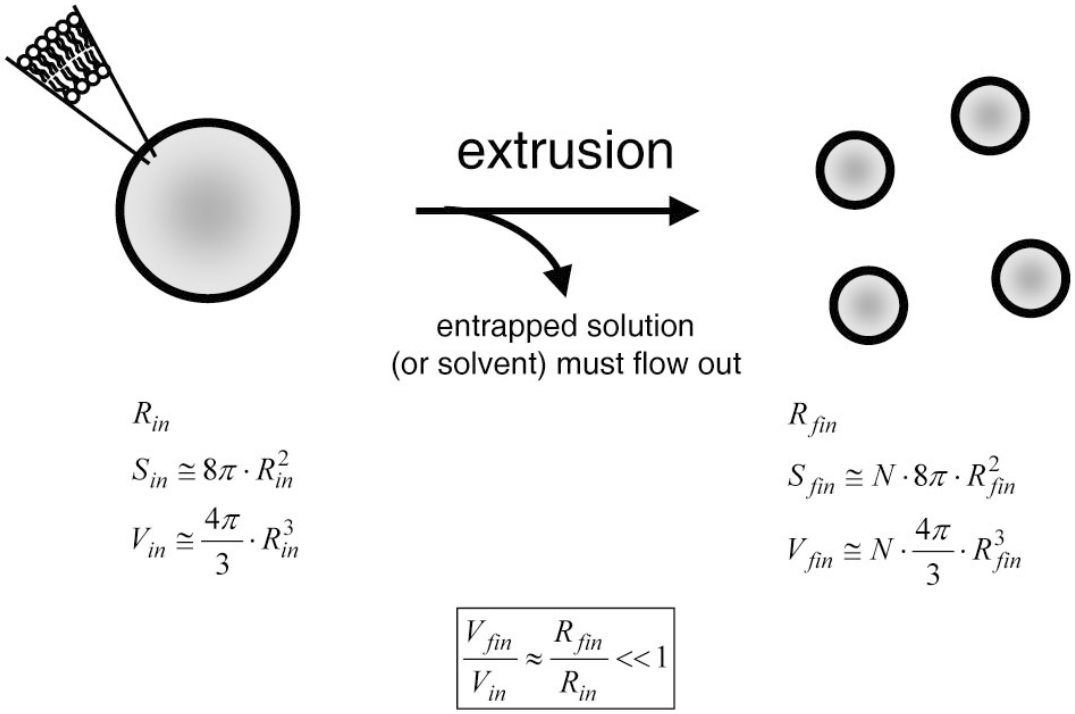
Supplementary Figure 2. Micrographs of EGFP-expressing vesicles (V1) as visualized by confocal microscope. Nomarski (left); fluorescence (centre); overlapping of the two images (right). The non-spherical shapes of vesicles V1 reflect their structures, often composed by vesicle clusters.

Supplementary Figure 3



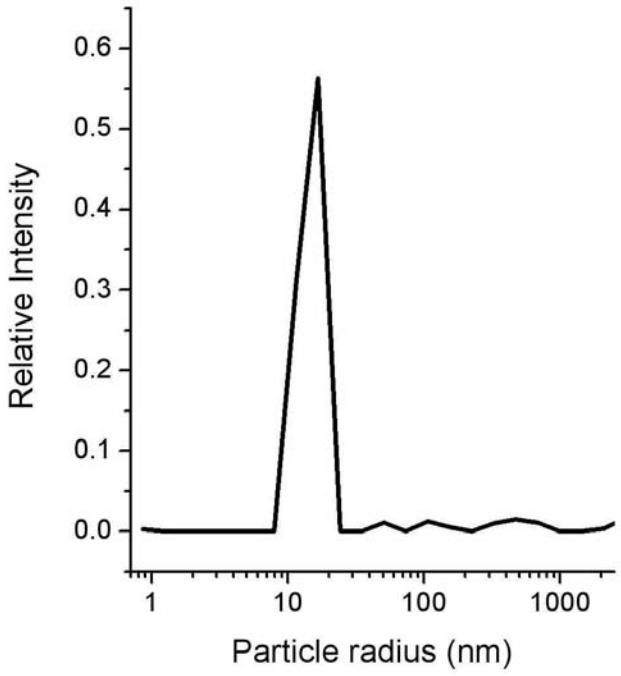
Supplementary Figure 3. EGFP calibration lines of (A) plate reader, in the presence of vesicles and sodium cholate; (B) RT-PCR instrument, in the presence of vesicles (gain = 5). The fluorescence values are considered as difference between measurements and negative control.

Supplementary Figure 4



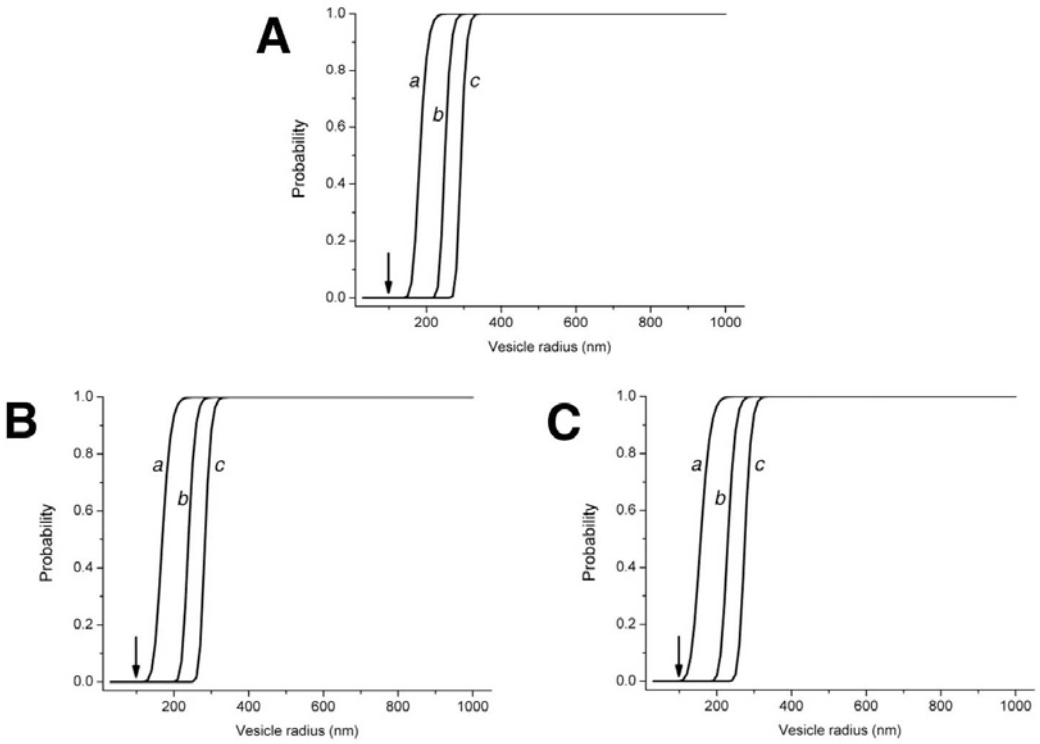
Supplementary Figure 4. Cartoon illustrating the reduction of entrapped volume after extrusion of vesicles. The model refers to spherical unilamellar vesicles, and the bilayer thickness has been neglected for the sake of simplicity in calculations. Extrusion proceeds at constant total membrane surface, so that several small vesicles are formed from a large vesicle. In this simplified analysis, the number of extruded vesicles is calculated as $N = (R_{fin}/R_{in})^2$, and the ratio between the cumulative volume of extruded vesicles and the volume of a single vesicle before extrusion is about R_{fin}/R_{in}, which is < 1. This very simple and idealized model shows that there is a reduction of total internal volume due to the size reduction of vesicles. In real cases, i.e., when the vesicles before extrusion are multilamellar, the volume reduction is higher. The activity in extruded vesicle can only be justified if selective release of solvent (water) is supposed, with little or no solute release.

Supplementary Figure 5



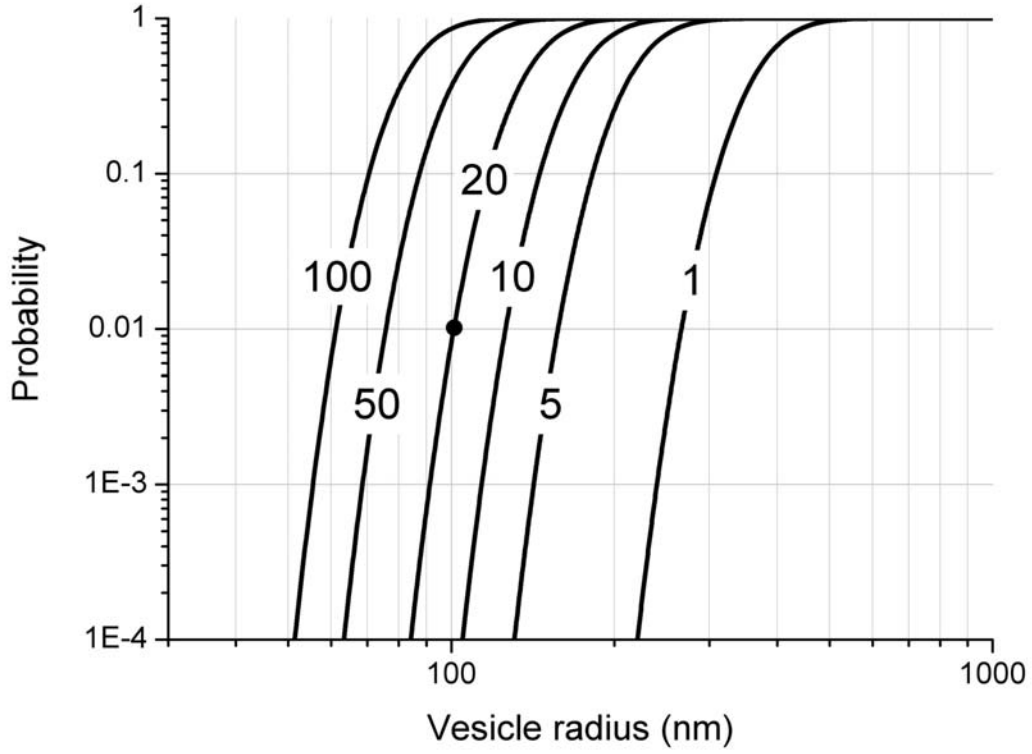
Supplementary Figure 5. DLS (90°) profile of neat PURE system.

Supplementary Figure 6



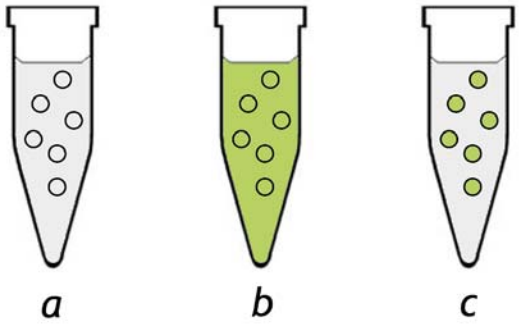
Supplementary Figure 6. Probability of co-entrapment of all macromolecular components of transcription-translation kit inside lipid vesicles of a given radius. In this case, the probability is calculated under the assumption that the components of the kit self-assemble giving rise to molecular clusters. The three panels correspond to the entrapment of (A) 50, (B) 20, (C) 10 molecular clusters. The concentration of each cluster has been fixed to 0.3 μM. The arrow points to the vesicle size used in the experiments. The three curves in each panel indicate to the probability of entrapping at least one (*a*), five (*b*) or ten (*c*) copies of each cluster inside the same vesicle.

Supplementary Figure 7



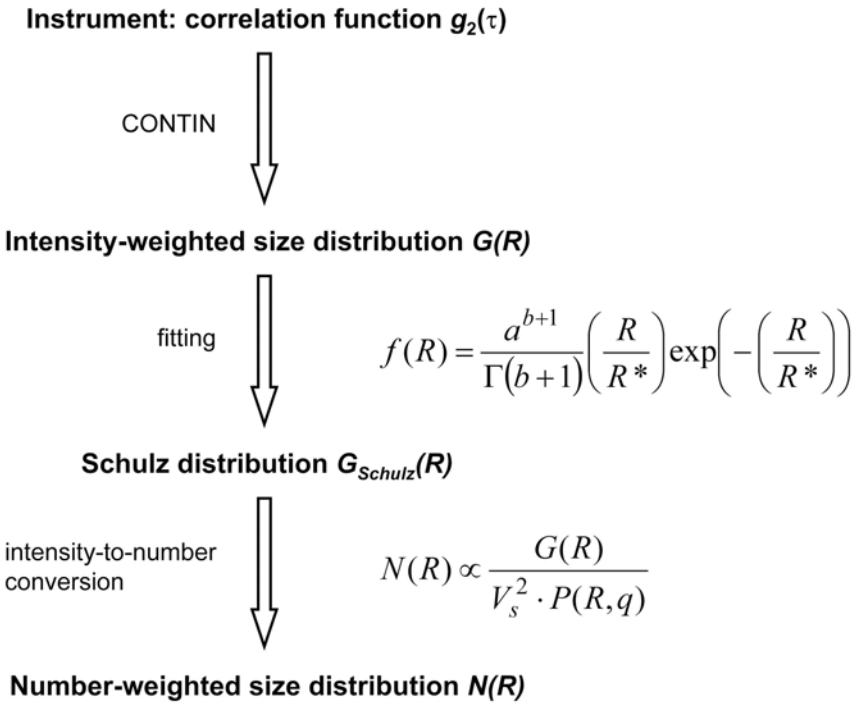
Supplementary Figure 7. Probability of co-entrapment of all macromolecular components (single copy) of transcription-translation kit inside lipid vesicles of a given radius. In this case, the probability is calculated under the assumption that the nominal concentration of the components may be empirically raised of a certain factor (shown in the figure over each curve). In this way, the Poisson model provides the probability of co-entrapment of different macromolecular components as if they were present at higher concentrations. It follows, for example, that there is 1% probability to find a vesicle of radius 100 nm containing at least one macromolecule of each of the 83 species, only if their nominal concentration is empirically raised by a factor 20 (indicated as a black dot). Notice the double logarithmic scale.

Supplementary Figure 8



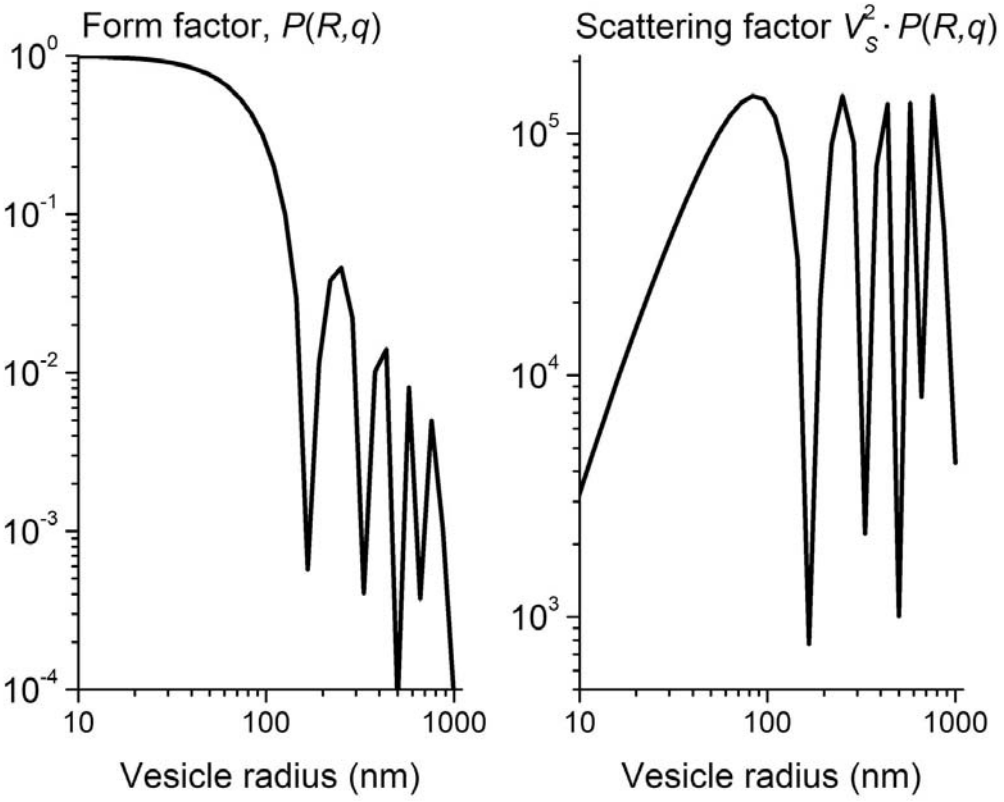
Supplementary Figure 8. Pictorial representation of samples prepared according to the procedure used in this work. (*a*) negative control (inhibitor inside and outside vesicles); (*b*) positive control (no inhibitor); (*c*) compartmentalized EGFP expression (inhibitor only outside). The green colour represents the expression of EGFP.

Supplementary Figure 9



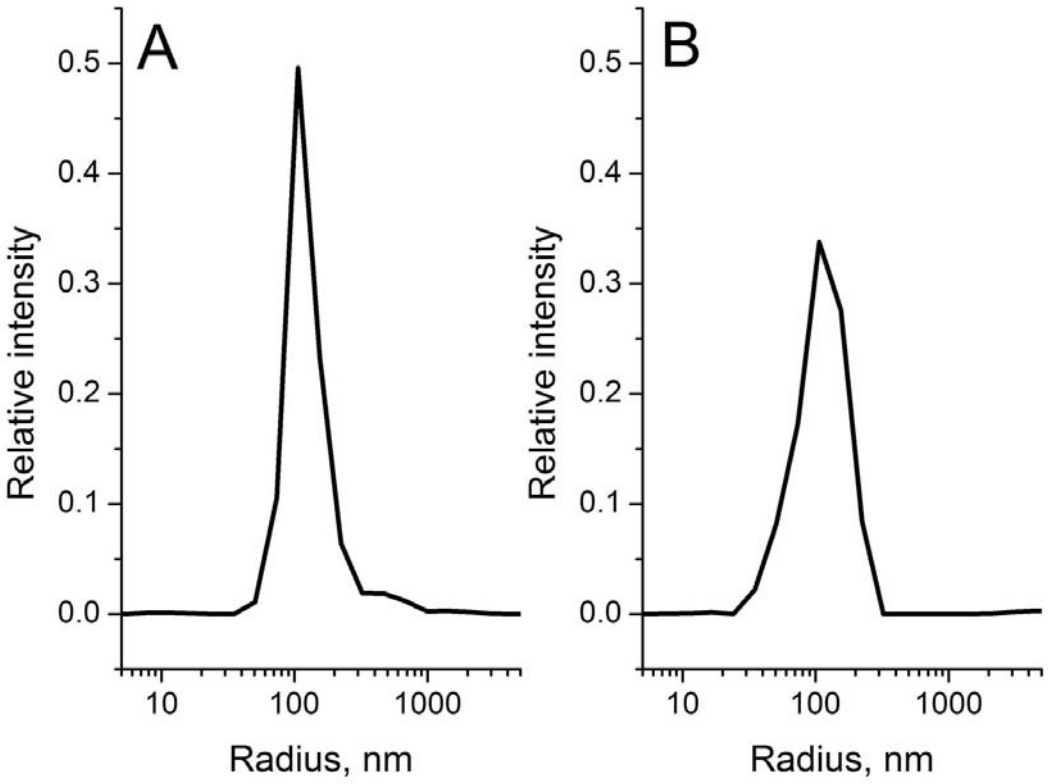
Supplementary Figure 9. Strategy for converting *intensity-weighted* size distributions into *number-weighted* ones. The correlation function is analyzed by the CONTIN algorithm in order to give the intensity-weighted particle size distribution, where large particles are overestimated due to their better ability to scatter light. The data are then fitted in order to obtain the best Schulz distribution function (the parameters *a*, *b* and *R** are obtained by the fitting procedure). The analytical Schulz distribution is then transformed by taking into account the scattering volume V_S and the form factor $P(R,q)$. Γ is the gamma function.

Supplementary Figure 10



Supplementary Figure 10. Dependence of form factor (left) and scattering factor (right) from the vesicle radius. Vesicles are assumed as spherical and unilamellar, the bilayer thickness has beens set to 3.8 nm. The first singular point is located at ca. 156 nm. After transformation (Eq. 12), the corresponding point, which is abnormally high, has been removed and substituted by the average between adjacent values.

Supplementary Figure 11



Supplementary Figure 11. Intensity-weighted size distribution (CONTIN, 90°, before the transformation) of EGFP-expressing vesicles V2 (A) and V3 (B). The corresponding number-weighted size distributions are shown in the main text in Figure 2 and Figure 3, respectively. The two size distributions have been first fitted by the Schulz distribution (V2: a = 551.2; b = 17.4; R^* = 3662.1; V3: a = 209.8; b = 5.8; R^* = 4323.6), then transformed into number-weighted profiles.

3. SUPPLEMENTARY REFERENCES

(1) Boukobza, E., Sonnenfeld, A. & Haran, G. Immobilization in Surface-Tethered Lipid Vesicles as a New Tool for Single Biomolecule Spectroscopy. *J. Phys. Chem. B* **105**, 12165-12170 (2001).

(2) Sunami, T., Sato, K., Matsuura, T., Tsukada, K., Urabe, I. & Yomo, T. Femtoliter compartment in liposomes for in vitro selection of proteins. *An. Biochem.* **357**, 128-136 (2006).

(3) Lonchin, S., Luisi, P. L., Walde, P., Robinson, B. H. A Matrix Effect in Mixed Phospholipid/Fatty Acid Vesicle Formation. *J. Phys. Chem. B* **103**, 10910-10916 (1999).

(4) Shimizu, Y., Inoue, A., Tomari, Y., Suzuki, T., Yokogawa, T., Nishikawa, K. & Ueda, T. Cell-free translation reconstituted with purified components. *Nature Biotechnology* **19**, 751-755 (2001).

(5) Shimizu, Y., Kanamori, T. & Ueda, T. Protein synthesis by PURE translation systems. *Methods* **36**, 299-304 (2005).

(6) Y. Kuruma ("Enrico Fermi" Study and Research Centre, Rome, Italy), personal communication.

(7) Dong, H., Nilsson, L. & Kurland, C. G. Co-variation of tRNA abundance and codon usage in Escherichia coli at different growth rates. *J. Mol. Biol.* **260**, 649-663 (1996).

(8) Schurtenberger, P., & Newman, M. E. Characterization of biological and environmental particles usign static and dynamic light scattering. In "*Environmental Particles*" Vol. 2; J. Buffle & H. P. van Leeuwen (eds.). Environmental Analytical and Physical Chemistry Series, Lewis Publisher, (1993).

(9) Luisi, P. L., Stano, P., Rasi, S., & Mavelli, F. A Possible Route to Prebiotic Vesicle Reproduction. *Artificial Life* **10**, 297-308 (2004).

(10) Hallett, F. R., Watton, J., Krygsman, P. Vesicle sizing. Number distribution by dynamic light scattering. *Biophys. J.* **59**, 357-362 (1991).