Electronic Supplementary Material (ESI) for Digital Discovery.

Supplementary information for:

Machine learning-guided high throughput nanoparticle design

Ana Ortiz-Perez[1,†], Derek van Tilborg[1,2,†], Roy van der Meel[1], Francesca Grisoni[1,2], Lorenzo Albertazzi[1*]*

[1]Institute for Complex Molecular Systems (ICMS), Eindhoven University of technology, PO Box 513, 5600 MB Eindhoven, The Netherlands.

[2]Centre for Living Technologies, Alliance TU/e, WUR, UU, UMC Utrecht, Princetonlaan 6, 3584 CB, Utrecht, The Netherlands.

[†] These authors contributed equally to this work

**Corresponding authors*

Table of contents

Part 1. Formulation variable characteristics

Theoretical space of the design space

The design space consists of one discrete variable (the solvent/antisolvent ratio) of four steps and four continuous formulation variables as shown in Sup. Table 1. Since these formulation variables represent composition fractions and need to sum to one, we sample them from a Dirichlet distribution, which we constrain by variable range and distinguishability according to the experimental errors of each variable.

Sup. Table 1. Characteristics of the formulation variables.

Formulation variable	Range	Experimental error
PLGA	0.06 – 0.70	1.2490%
PP-L	0.06 – 1.00	1.2121%
PP-COOH	0.06 – 1.00	1.2359%
PP-NH2	0.06 – 1.00	1.2398%
S/AS	0.10, 0.15, 0.20, 0.25	

To estimate the number of unique nanoparticle formulations in this design space, we first calculate the number of distinguishable points for each variable $N = \int_a^b \frac{1}{\varepsilon \times x} dx$, where a and b are the lower and upper bounds of the variable range, respectively, and ε is the experimental error. This yields:

$$N_{PLGA} = \int_{0.06}^{0.70} \frac{1}{0.01249\, x} dx$$

$$N_{PP-L} = \int_{0.06}^{1.00} \frac{1}{0.012121\, x} dx$$

$$N_{PP-COOH} = \int_{0.06}^{1.00} \frac{1}{0.012359\, x} dx$$

$$N_{PP-NH2} = \int_{0.06}^{1.00} \frac{1}{0.012398\, x} dx$$

Since solvent/antisolvent ratio can only take on four values, $N_{S/AS} = 4$. This gives us

$$N_{design\ space} = N_{PLGA} \times N_{PP-L} \times N_{PP-COOH} \times N_{PO-NH2} \times N_{S/AS}$$

To factor in the constraint of $x_{PLGA} + x_{PP-L} + x_{PP-COOH} + x_{PO-NH2} = 1$, we can simulate the number of valid formulations by finding all combinations in the discretized vector X for each variable and check if all values sum to one considering an error ε (see Algorithm 1). If we take ε as the mean experimental error of all formulation variables, this yields an $N_{design\ space} \approx 1.85 \times 10^8$.

Algorithm 1 Design space estimation

```
procedure ValidCombinations(X_1, X_2, X_3, X_4, ε)
n←0
for x_{1_i},⋯,x_{4_i} in X_1,⋯,X_4 do
    if 1 − ε < Σ x_{1_i},…,x_{4_i} < 1 + ε) then
        n = n + 1
    end if
```

```
7.  end for
8.  end procedure
```

Part 2. Composition and experimental values for all formulated nanoparticles

Sup. Table 2. Starting dataset: Composition and experimental values.

	Composition variables (X)					Experimental values					Physicochemical properties							
		Composition (%)				Uptake					Malvern Zetasizer				Wyatt DynaPro *(bonus!)*			
	S/AS	PLGA	PP-L	PP-COOH	PP-NH2	Raw	σ	Coeff	Corrected (Y)	σ	Size	σ	PDI	σ	D (nm)	σ	PDI	σ
F1	0.1	0.33	0.67	0.00	0.00	0.60	0.07	1.01	0.696	0.083	119	1.18	0.069	0.008	116.3	0.07	0.082	0.0212
F2	0.1	0.33	0.33	0.33	0.00	0.85	0.15	1.15	0.830	0.145	121	1.32	0.063	0.006	123.1	0.92	0.089	0.0170
F3	0.1	0.33	0.33	0.00	0.33	0.64	0.11	0.95	0.743	0.122	121	0.58	0.072	0.009	124.1	0.78	0.064	0.0000
F4	0.1	0.33	0.00	0.67	0.00	1.06	0.16	1.14	1.122	0.171	134.4	0.66	0.079	0.017	140.1	0.71	0.091	0.0184
F5	0.1	0.33	0.00	0.33	0.33	1.03	0.04	1.25	0.987	0.037	132.9	0.64	0.069	0.011	136.8	1.34	0.103	0.0212
F6	0.1	0.33	0.00	0.00	0.67	0.93	0.22	1.24	0.888	0.211	145	1.17	0.057	0.018	148.3	0.71	0.098	0.0035
F7	0.1	0.00	1.00	0.00	0.00	0.57	0.05	1.34	0.400	0.034	80.2	0.87	0.059	0.01	76.7	0.21	0.098	0.0028
F8	0.1	0.00	0.67	0.33	0.00	2.18	0.11	1.39	2.368	0.115	103.3	0.95	0.11	0.014	102.5	0.21	0.165	0.0057
F9	0.1	0.00	0.33	0.67	0.00	2.16	0.18	1.56	1.870	0.157	111.3	1.04	0.134	0.015	114.0	0.64	0.183	0.0014
F10	0.1	0.00	0.33	0.33	0.33	4.48	0.25	1.56	3.982	0.225	105.2	0.47	0.142	0.017	106.2	0.14	0.160	0.0212
F11	0.1	0.00	0.33	0.00	0.67	5.88	0.14	1.58	4.770	0.113	89.04	0.96	0.075	0.005	87.0	0.14	0.095	0.0099
F12	0.1	0.00	0.00	1.00	0.00	2.53	0.42	1.52	2.055	0.338	125.3	0.20	0.142	0.013	130.3	1.56	0.177	0.0184
F13	0.1	0.00	0.00	0.67	0.33	4.56	0.12	1.65	3.615	0.097	121.4	0.36	0.14	0.014	124.7	0.71	0.175	0.0071
F14	0.1	0.00	0.00	0.00	1.00	1.79	0.29	1.33	1.673	0.271	94.63	0.95	0.098	0.035	97.5	0.00	0.112	0.0064
F15	0.1	0.25	0.25	0.25	0.25	1.38	0.17	1.85	1.095	0.139	128.2	0.35	0.05	0.015	128.0	0.71	0.085	0.0092
F16	0.25	0.67	0.33	0.00	0.00	1.41	0.22	1.94	0.575	0.091	177.3	2.45	0.116	0.012	189.4	7.21	0.159	0.0141
F17	0.25	0.67	0.00	0.33	0.00	10.77	2.08	1.95	3.662	0.706	246.4	1.03	0.062	0.021	281.4	0.85	0.107	0.0106
F18	0.25	0.67	0.00	0.00	0.33	11.27	1.17	2.70	2.293	0.237	260.9	2.68	0.233	0.006	275.1	9.55	0.317	0.0332
F19	0.25	0.33	0.33	0.33	0.00	1.69	0.32	1.44	1.014	0.191	165.3	1.48	0.131	0.015	175.3	1.13	0.165	0.0219
F20	0.25	0.33	0.33	0.00	0.33	1.93	0.09	1.71	0.705	0.031	144.7	1.30	0.083	0.017	154.1	5.30	0.101	0.0113
F21	0.25	0.33	0.00	0.67	0.00	3.74	0.63	1.47	2.177	0.370	176.7	2.50	0.087	0.007	184.6	1.84	0.110	0.0156
F22	0.25	0.33	0.00	0.33	0.33	3.33	1.27	1.58	1.478	0.563	198.1	1.70	0.108	0.011	206.4	5.16	0.141	0.0318
F23	0.25	0.00	1.00	0.00	0.00	7.04	1.01	1.84	3.126	0.450	86.76	0.46	0.101	0.018	89.1	0.21	0.134	0.0134
F24	0.25	0.00	0.67	0.00	0.33	9.11	1.32	2.09	3.598	0.522	87.65	0.56	0.129	0.006	91.6	0.35	0.138	0.0163
F25	0.25	0.00	0.33	0.67	0.00	4.92	0.37	1.89	1.883	0.141	141.9	1.56	0.134	0.027	142.5	0.14	0.156	0.0042
F26	0.25	0.00	0.33	0.33	0.33						122.6	1.18	0.163	0.014	135.1	0.42	0.192	0.0099
F27	0.25	0.00	0.33	0.00	0.67	7.94	0.41	2.11	3.767	0.195	93.14	1.50	0.131	0.009	94.8	0.14	0.143	0.0028
F28	0.25	0.00	0.00	1.00	0.00	2.76	0.31	1.54	1.790	0.202	161.6	1.02	0.171	0.02				
F29	0.25	0.00	0.00	0.67	0.33	5.19	0.33	3.70	1.403	0.088	149.9	1.65	0.146	0.022	153.4	1.06	0.160	0.0127
F30	0.25	0.00	0.00	0.00	1.00	7.83	1.14	1.84	4.265	0.620	108.4	0.76	0.139	0.018	113.9	0.64	0.177	0.0064
Blank		0.00	1.00	0.00	0.00	0		0.00	0		66.78	0.429	0.082	0.001	66.5	0.1	0.117	0.009
Up		0.00	1.00	0.00	0.00	1		1.00	1		68.43	2.237	0.108	0.053	68.9	0.3	0.173	0.010

Blank: *PLGA-PEG without dye,* ***Up (Uptake Standard):*** *PLGA-PEG NP with 50 μM (Formulated by bulk nanoprecipitation as described in materials and methods). These are standard particles, not used for modelling purposes. Sample F18 had a polydispersity index > 0.2 (0.233) so it was not considered for the uptake model (its physicochemical data was used to model PDI and size). Sample F26 was also not included in the uptake model, since this sample was lost during handling and the coefficient for measurement could not be measured.*

Sup. Table 3. ML iteration 01 (Exploration): Composition and experimental values.

		Composition variables (X)					Experimental values								
			Composition (%)				Uptake					Physicochemical properties			
	ID	S/AS	PLGA	PP-L	PP-COOH	PP-NH2	Raw	σ	Coeff	Corrected (Y)	σ	PDI	σ	Size	σ
S01.01	screen_37609	0.15	0	0.24	0.06	0.7	11.18	0.24	1.16	9.63	0.21	0.115	0.041	101.9	1.097
S01.02	screen_39525	0.2	0	0.46	0.06	0.48	11.30	0.30	1.16	9.78	0.26	0.099	0.013	93.52	0.9279
S01.03	screen_65935	0.2	0	0.07	0.06	0.87	13.00	0.35	1.15	11.31	0.31	0.099	0.01	109.6	0.4163
S01.04	screen_95443	0.2	0	0.83	0.07	0.1	6.57	0.35	1.15	5.72	0.30	0.117	0.021	87.87	0.9051
S01.05	screen_9364	0.2	0	0.3	0.16	0.55	11.52	0.18	1.10	10.49	0.17	0.127	0.011	105	1.607
S01.06	screen_39930	0.2	0	0.08	0.21	0.71	14.09	0.40	1.22	11.59	0.33	0.135	0.015	120.3	1.217
S01.07	screen_51944	0.2	0	0.62	0.06	0.32	7.29	0.40	1.23	5.93	0.33	0.108	0.007	89.42	0.7111
S01.08	screen_72795	0.2	0	0.67	0.17	0.16	10.58	0.44	0.73	14.40	0.60	0.121	0.007	96.39	0.695
S01.09	screen_24111	0.2	0	0.13	0.37	0.5	11.13	0.77	1.02	10.89	0.75	0.108	0.003	119.1	1.114
S01.10	screen_56633	0.2	0	0.44	0.24	0.32	8.15	0.25	1.10	7.44	0.23	0.13	0.013	105.6	1.778

Physicochemical properties measured with the Malvern Zetasizer as described in the materials and methods.

Sup. Table 4. ML iteration 02 (Exploitation): Composition and experimental values.

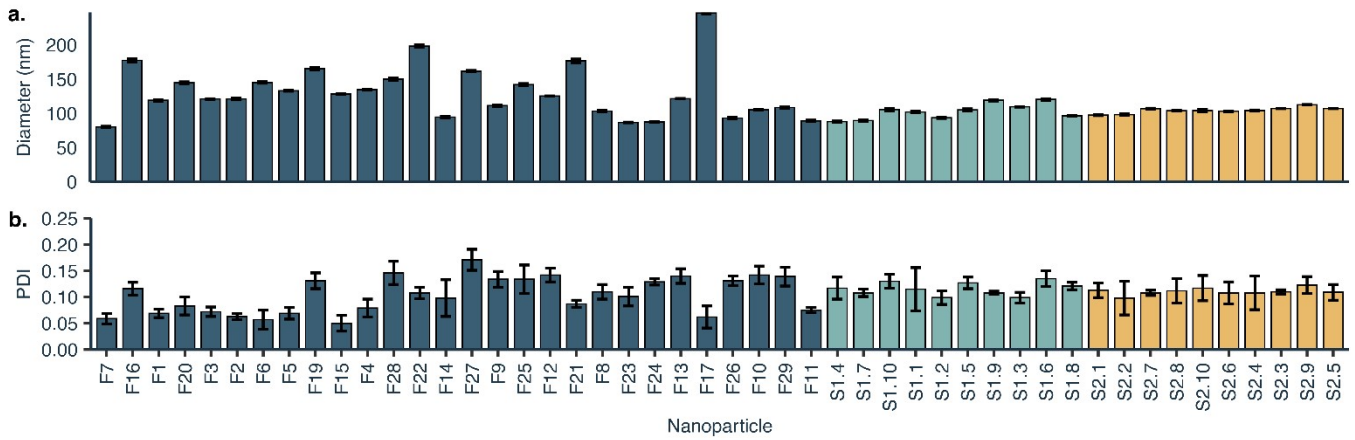
		Composition variables (X)					Experimental values								
			Composition (%)				Uptake					Physicochemical properties			
	ID	S/AS	PLGA	PP-L	PP-COOH	PP-NH2	Raw	σ	Coeff	Corrected (Y)	σ	PDI	σ	Size	σ
S02.01	screen_72872	0.1	0	0.33	0.22	0.45	14.55	0.63	1.69	8.60	0.37	0.113	0.014	97.38	0.81
S02.02	screen_91542	0.1	0	0.34	0.22	0.44	15.08	0.19	1.57	9.62	0.12	0.098	0.032	98.14	1.13
S02.03	screen_52734	0.2	0	0.25	0.3	0.45	17.93	0.44	1.29	13.91	0.34	0.11	0.004	107	0.20
S02.04	screen_38116	0.2	0	0.3	0.28	0.42	17.49	1.23	1.27	13.78	0.97	0.108	0.032	104	0.95
S02.05	screen_96636	0.2	0	0.22	0.29	0.49	18.66	0.93	1.29	14.50	0.72	0.109	0.015	107	0.20
S02.06	screen_72388	0.2	0	0.29	0.27	0.44	17.29	1.10	1.27	13.62	0.87	0.108	0.021	103.2	0.42
S02.07	screen_38000	0.2	0.07	0.26	0.17	0.5	14.42	1.54	1.28	11.27	1.21	0.108	0.005	106.9	0.68
S02.08	screen_20306	0.25	0	0.38	0.28	0.34	14.18	0.78	1.23	11.54	0.64	0.112	0.023	103.8	0.40
S02.09	screen_62254	0.25	0.07	0.17	0.19	0.57	17.55	0.33	1.25	14.02	0.26	0.123	0.016	112.5	0.30
S02.10	screen_82381	0.25	0	0.44	0.29	0.28	14.86	0.98	1.22	12.17	0.81	0.117	0.024	104	1.42

Physicochemical properties measured with the Malvern Zetasizer as described in the materials and methods.

Sup. Table 5. Validation: Composition and experimental values.

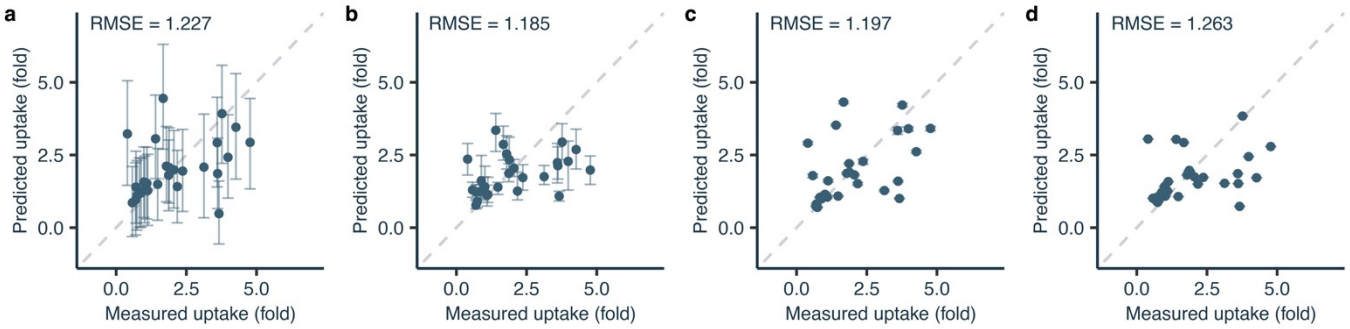
		Composition variables (X)					Experimental values										
			Composition (%)				Uptake					Physicochemical properties					
	ID	S/AS	PLGA	*PP-L*	*PP-COOH*	*PP-NH2*	Raw	σ	Coeff	Corrected (Y)	σ	PDI	σ	Size	σ	Zpot	σ
V1	screen_64729	0.1	0.53	0.37	0.00	0.09	2.70	0.50	1.41	1.92	0.36	0.043	0.033	152.7	2.26	-28.4	0.493
V2	screen_91456	0.1	0.47	0.42	0.00	0.12	2.05	0.25	1.41	1.46	0.18	0.051	0.035	137.9	0.38	-29.1	0.404
V3	screen_79724	0.1	0.52	0.29	0.07	0.12	4.23	0.76	1.35	3.14	0.56	0.066	0.032	151	1.48	-38.9	1.32
V4	screen_19528	0.1	0.46	0.43	0.11	0.00	4.84	0.51	1.32	3.66	0.38	0.076	0.017	139.8	0.56	-30.2	1.04
V5	screen_23350	0.25	0.67	0.18	0.00	0.15	4.51	0.84	1.29	3.50	0.66	0.06	0.018	191	2.16	-31.2	0.379
V6	screen_60153	0.15	0.00	0.35	0.30	0.36	14.14	0.66	1.30	10.91	0.51	0.103	0.013	106.6	0.38	-18.1	0.781
V7	screen_82424	0.15	0.00	0.31	0.30	0.39	14.22	0.78	1.37	10.39	0.57	0.13	0.012	110.6	0.57	-38.8	1.97
V8	screen_3544	0.2	0.06	0.37	0.20	0.37	12.22	1.28	1.28	9.57	1.00	0.109	0.016	116.2	1.60	-21.9	0.551
V9	screen_30061	0.25	0.10	0.30	0.22	0.38	12.49	0.54	1.10	11.33	0.49	0.133	0.003	117.8	1.96	-35.5	1.01
V10	screen_73941	0.25	0.08	0.30	0.21	0.42	12.70	0.60	1.21	10.49	0.50	0.134	0.008	118.8	0.87	-25.5	0.306

Physicochemical properties measured with the Malvern Zetasizer as described in the materials and methods.

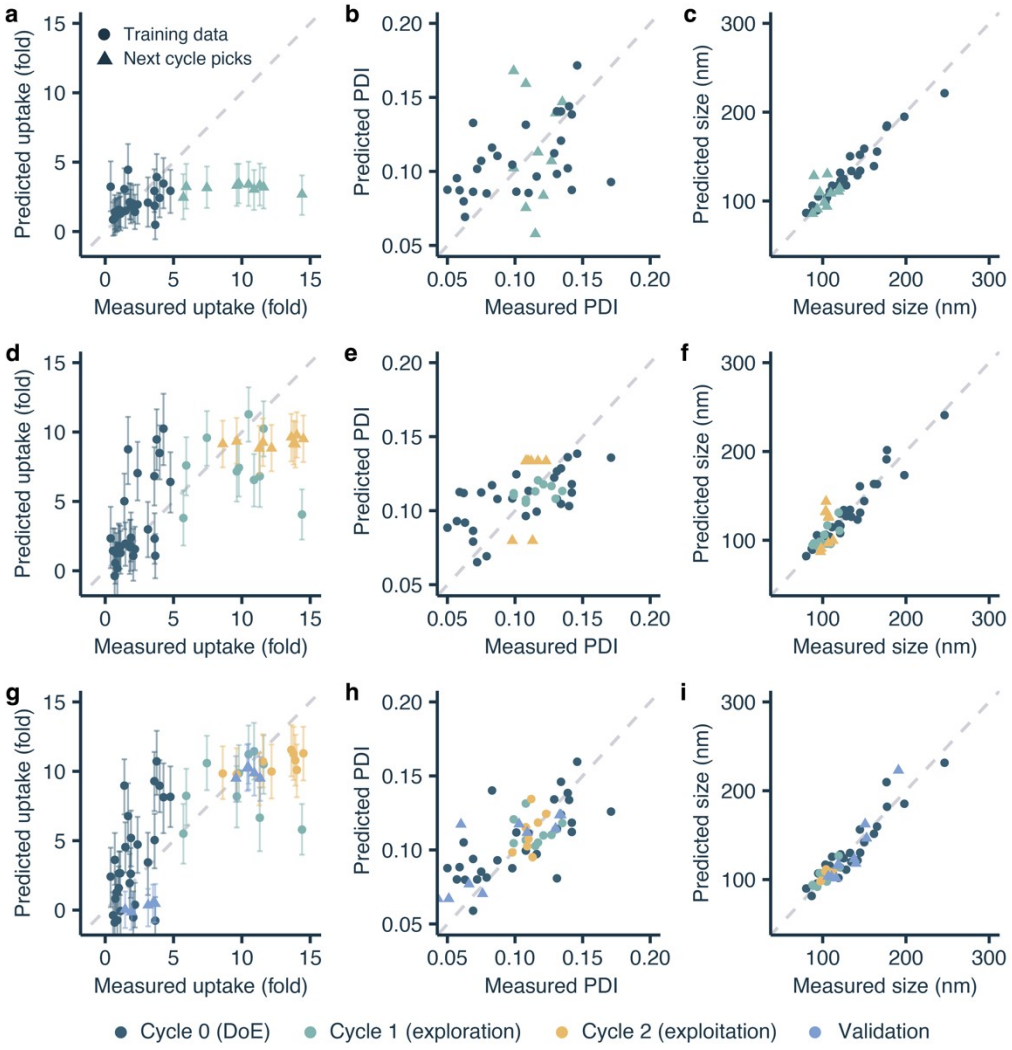


Sup. Figure 1. Size and PDI for all formulated nanoparticles.

Part 2. Model predictions



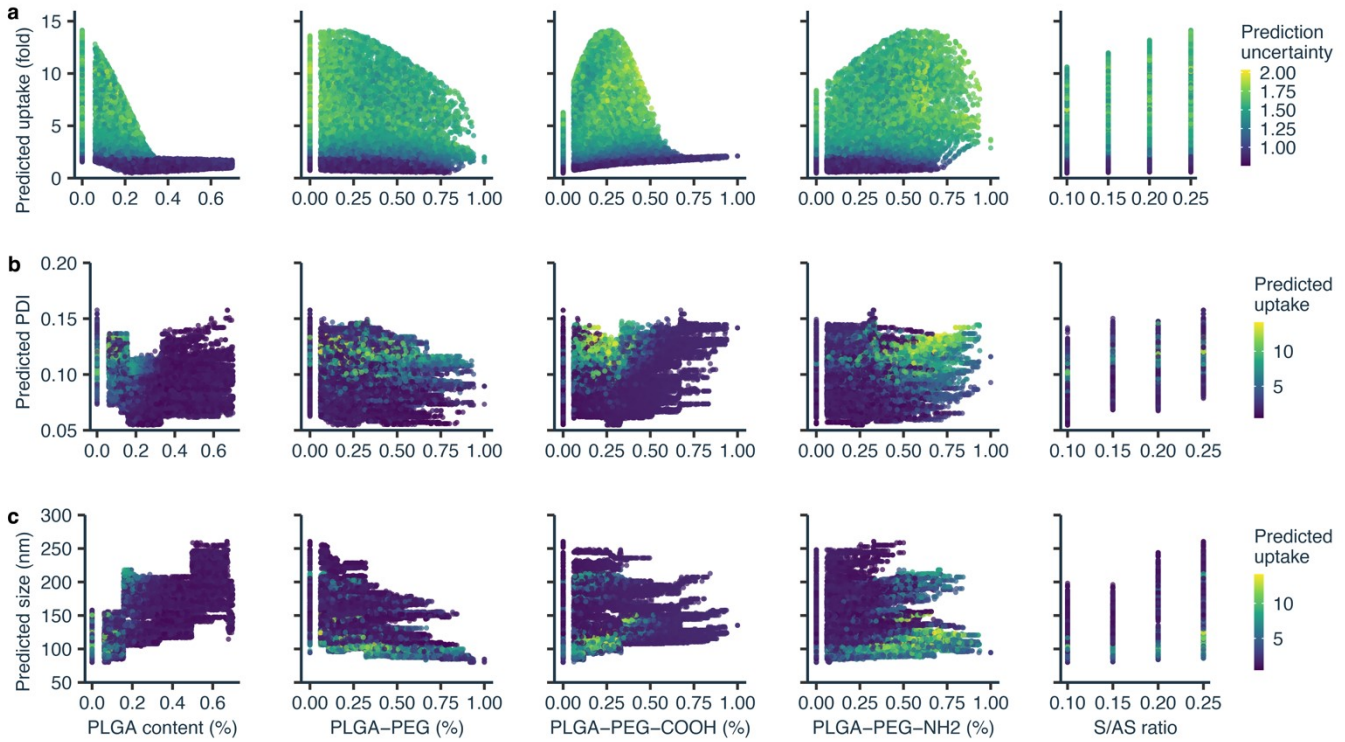
Sup. Figure 2. Predicted vs measured uptake for different models on the initial data (cycle 0). a) Bayesian neural network. **b)** XGBoost. **c)** Random Forest. **d)** Gaussian Process. Error bars represent the predicted 90% confidence interval. Performance is shown as the Root Mean Square Error (RMSE) determined through 5-fold cross validation. Different machine learning models achieve similar predictive errors, while uncertainty estimates differ considerably.



Sup. Figure 3. Predicted vs measured Uptake, PDI, and size. Every plot contains the training data used to fit the model (circles) and the nanoparticles proposed for formulation and screening (triangles). Bayesian neural networks were used for uptake predictions, while XGBoost models were used for PDI and size predictions. Error bars represent the predicted 90% confidence interval. **a-c)** Modelled uptake, PDI, and size using the initial dataset of 28 nanoparticles. **d-e)** Modelled uptake, PDI, and size using data from cycle 0 and 1. **g-h)** Modelled uptake, PDI, and size using data from cycle 0, 1, and 2.

Sup. Table 6. Model performance. Performance is shown as the Root Mean Square Error (RMSE) determined through five-fold-cross validation.

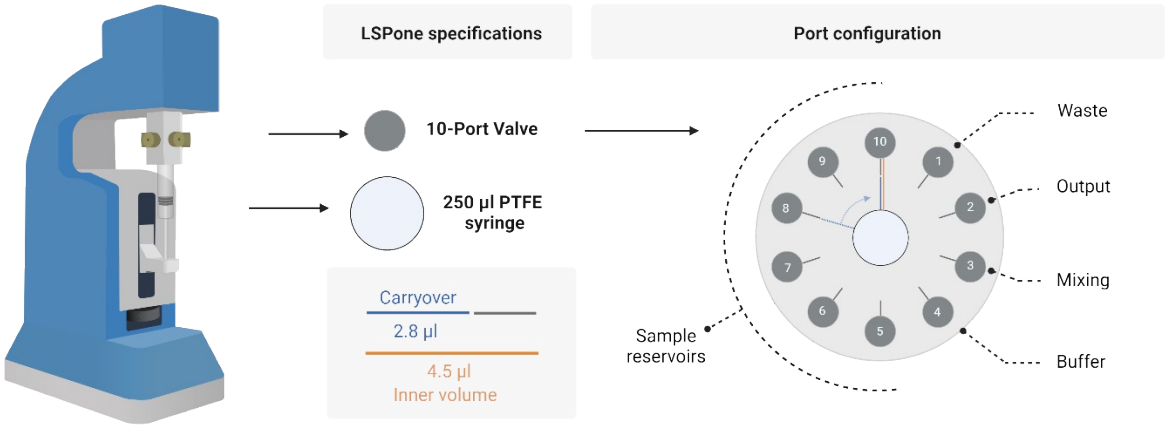
Cycle	Uptake RMSE	PDI RMSE	Size RMSE
0	1.2271	0.0363	12.3527
1	3.1468	0.0320	10.2354
2	3.1295	0.0286	12.3659



Sup. Figure 4. Nanoparticle formulation-property relationships for Uptake, PDI, and size. Every dot is a nanoparticle in the *in-silico* screening library. Predictions are made using the models trained on cycle 0-2. **a)** Relationships between NP formulation variables and predicted uptake using a Bayesian neural network. Colored by prediction uncertainty. **b-c)** Relationships between NP formulation variables and predicted size/PDI using a XGBoost model. Colored by prediction predicted uptake.

Part 3. Experimental set-up and high content imaging optimization

Pump port set-up



Sup. Figure 5. Schematic representation of the valve setup in the LSPOne pump.

Pipetting error

Sup. Table 7. Pipetting errors from manufacturer's manual.

Volume range	Dispensing button color	Systematic error[1]		Random error[1]	
10 µL	medium gray (for epT.I.P.S.® 20 µL)	±1.2 %	±0.12 µL	±0.6 %	±0.06 µL
20 µL	light gray (for epT.I.P.S.® 20 µL L)	±0.8 %	±0.16 µL	±0.3 %	±0.06 µL
10 µL	yellow (for epT.I.P.S.® 200 µL)	±1.2 %	±0.12 µL	±0.6 %	±0.06 µL
20 µL		±1.0 %	±0.2 µL	±0.3 %	±0.06 µL
25 µL		±1.0 %	±0.25 µL	±0.3 %	±0.08 µL
50 µL		±0.7 %	±0.35 µL	±0.3 %	±0.15 µL
100 µL		±0.6 %	±0.6 µL	±0.2 %	±0.2 µL
200 µL		±0.6 %	±1.2 µL	±0.2 %	±0.4 µL
200 µL	blue (for epT.I.P.S.® 1,000 µL)	±0.6 %	±1.2 µL	±0.2 %	±0.4 µL
250 µL		±0.6 %	±1.5 µL	±0.2 %	±0.5 µL
500 µL		±0.6 %	±3.0 µL	±0.2 %	±1.0 µL
1,000 µL		±0.6 %	±6.0 µL	±0.2 %	±2.0 µL

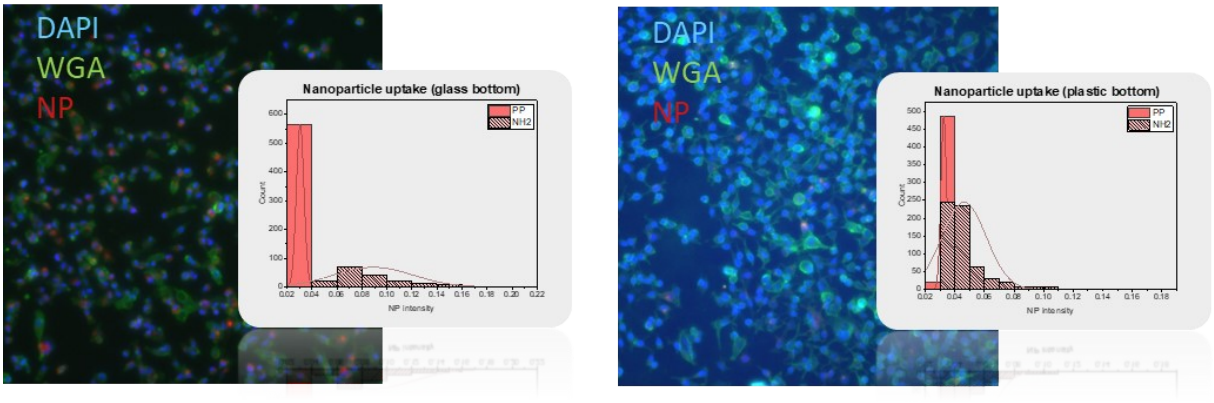
[1] The error data, according to EN ISO 8655, only apply if original Eppendorf tips are used. Technical specifications are subject to change. Errors and omissions excepted.

Manual available online: https://www.eppendorf.com/product-media/doc/en/672922/Liquid-Handling_Technical-data_Research-plus_Eppendorf-Research-plus.pdf

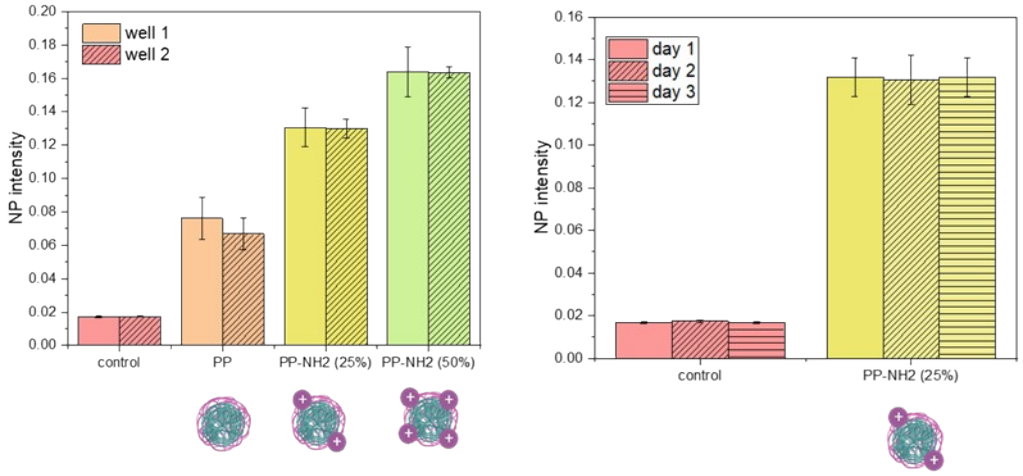
Optimization High content imaging

Optimization of (1) sample preparation (*i.e.*, seeding densities, nanoparticle concentration, incubation time, etc.) and (2) analysis pipeline was done with MDA-MB-231 cells in ibidi 8-µWell with optical bottom (glass and/or polymer).

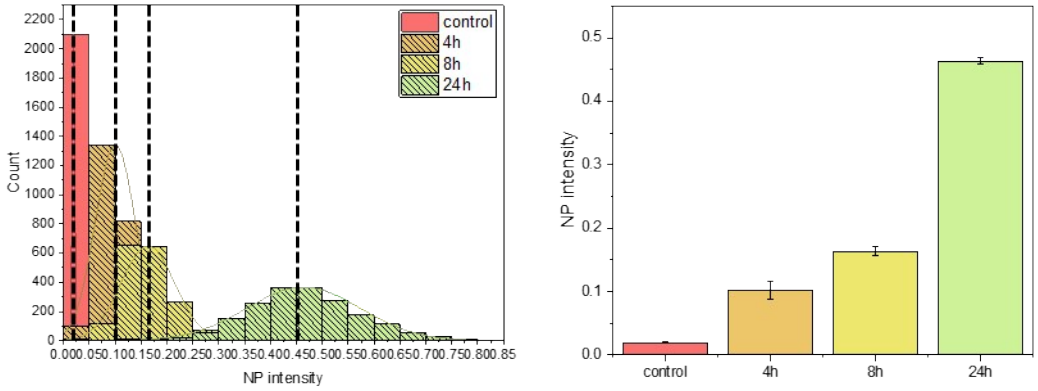
HT screening was then performed in ibidi 96-µWells with optical bottom (polymer). Polymer (over glass) optical bottom was chosen for its superior cell attachment. We highly recommend using a plate with optical bottom to minimize scattering and be able to quantify uptake differences using microscopy:



Sup. Figure 6. Importance of optical bottom for Nanoparticle signal.



Sup. Figure 7. Reproducibility between wells and days.



Sup. Figure 8. Nanoparticle uptake over time.