Supporting Information

Machine Learning Instructed Microfluidic Synthesis of Curcumin-loaded Liposomes

Valentina Di Francesco[1†], Daniela P. Boso[2†*], Thomas L. Moore[1†],

Bernhard A. Schrefler[2,3‡], Paolo Decuzzi[1‡*]

[1]*Laboratory of Nanotechnology for Precision Medicine, Istituto Italiano Di Tecnologia, Via Morego 30, 16163, Genova, Italy*

[2]*Department of Civil, Environmental and Architectural Engineering, University of Padova, Via Marzolo 9, 35131, Padova, Italy*

[3]*Institute for Advanced Studies, Technical University of Munich, Lichtenbergstraße 2 a, 85748, Garching, Germany*

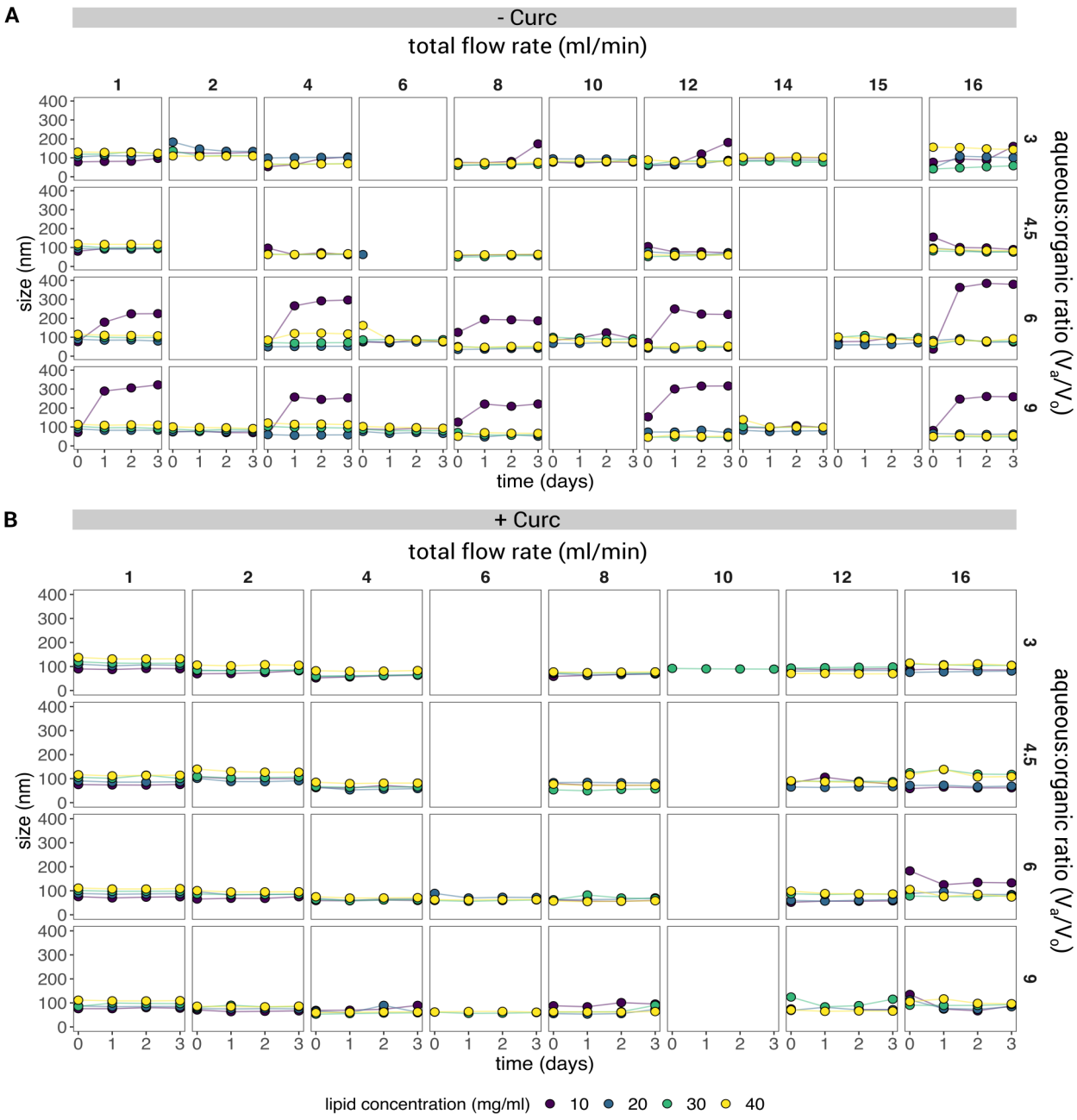
[†] D.P. Boso, V. Di Francesco and T.L. Moore contributed equally to this work

[‡] B.A. Schrefler and P. Decuzzi share the senior authorship for this work

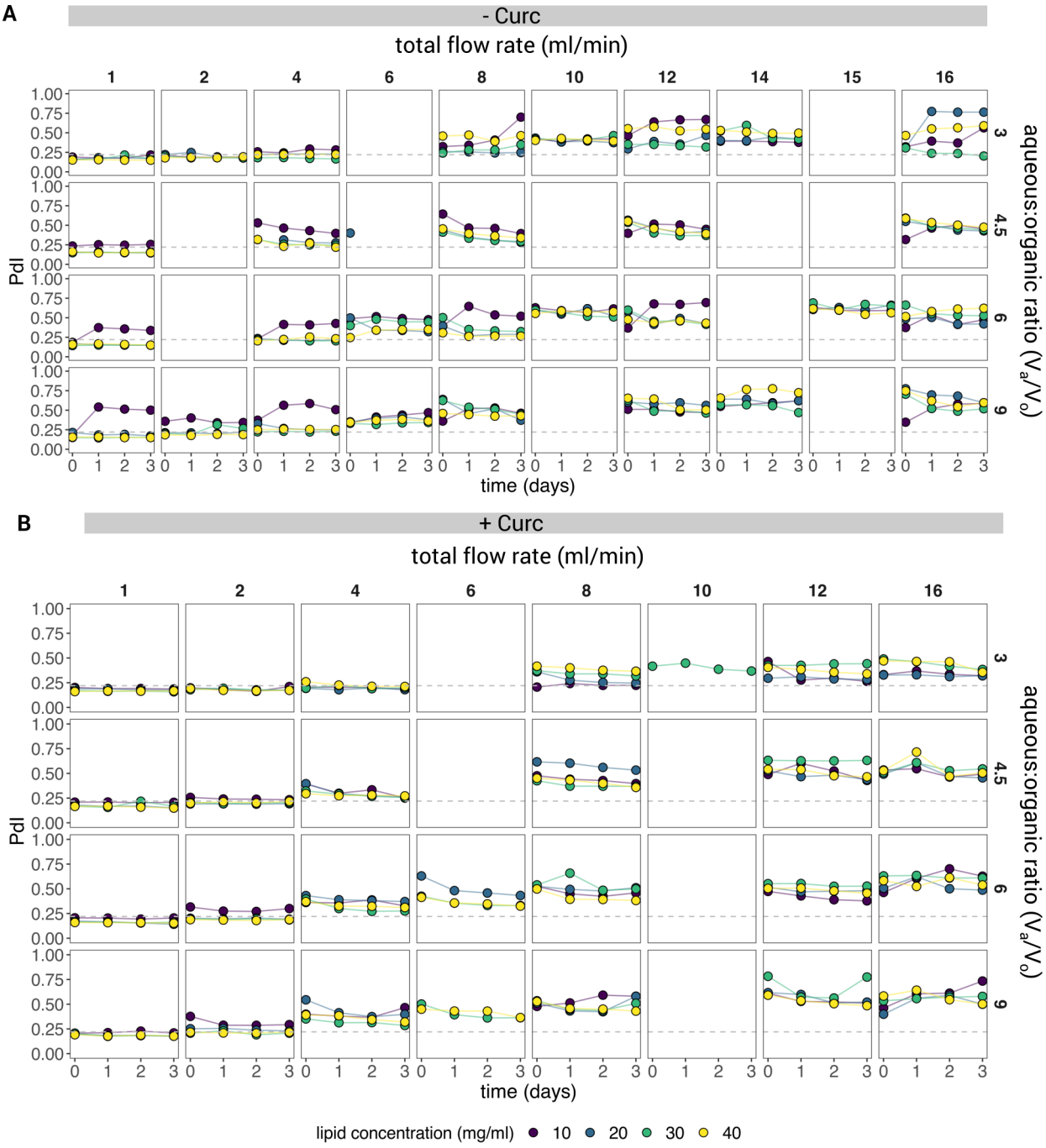
Corresponding author: Daniela P. Boso, PhD – daniela.boso@dicea.unipd.it;

Paolo Decuzzi, PhD – paolo.decuzzi@iit.it

Supporting Figures



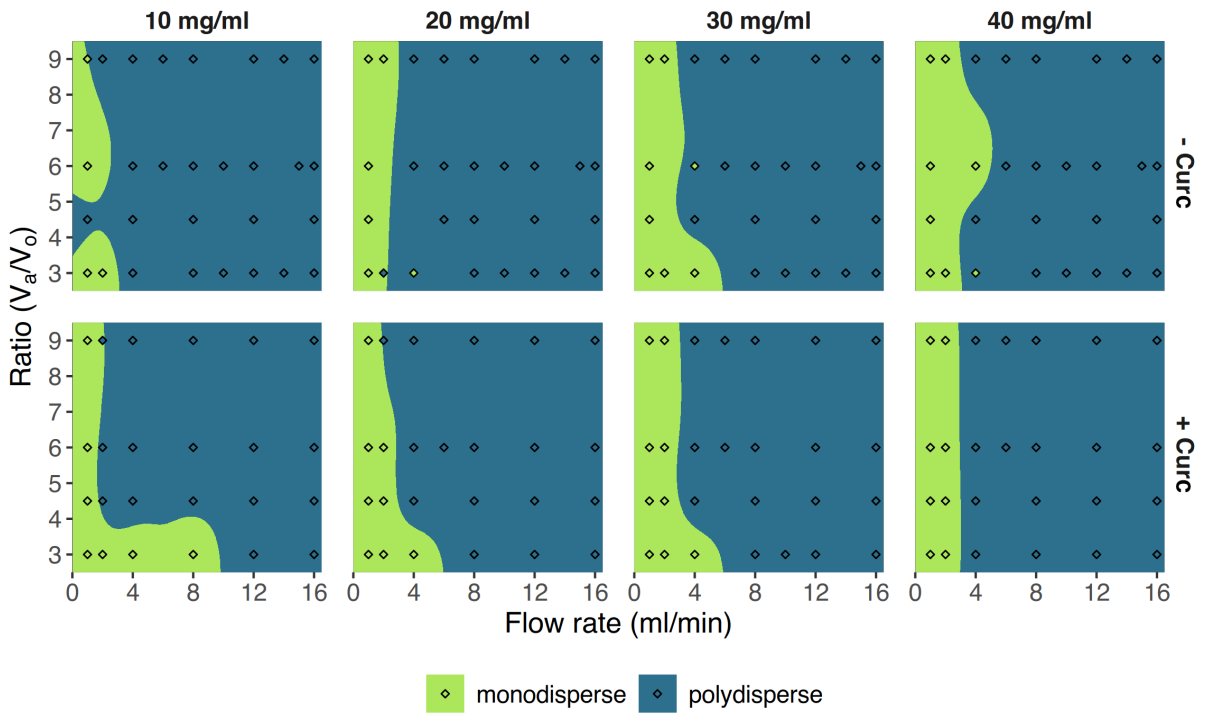
Supporting Figure S1. Characterization of the liposome hydrodynamic diameter over time. Formulations are split by the empty liposomes (- Curc) and those loaded with curcumin (+ Curc), total flow rate (TFR) as well as the aqueous to organic flow rate ratio (FRR). Particles were characterized over three days with dynamic light scattering (DLS).



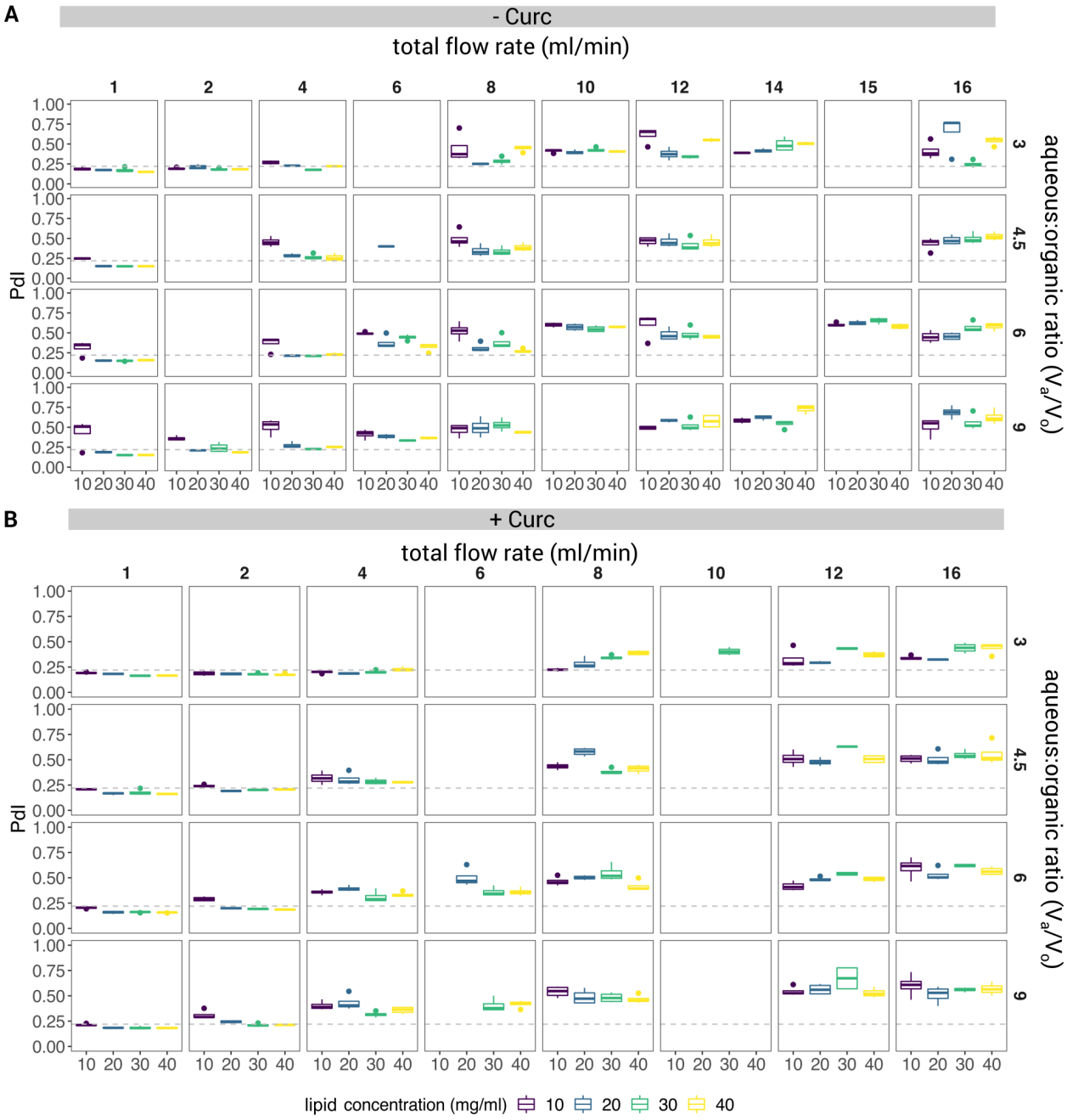
Supporting Figure S2. Characterization of the liposome polydispersity index (PdI) over time. Formulations are split by the empty liposomes (- Curc) and those loaded with curcumin (+ Curc), total flow rate (TFR) as well as the aqueous to organic volume ratio. Particles were characterized over three days with dynamic light scattering (DLS).

Total number of measurements *(including all the different formulations and all the technical replicates)*	**3518**
Total number of measurements (with $n \leq 6$)	1311
Number of unique formulations [Day 0] *(including the different formulations only)*	**218**
Total number of measurements [Day 0, $n \leq 6$]	779
Number of unique monodisperse formulations [Day 0]	**72**
Total number of monodisperse measurements [Day 0, $n \leq 6$]	236

Supporting Table S1. Summary table for the number of liposome formulations and DLS measurements comprising the total data. (n = maximum number of technical replicates considered for each unique formulation)



Supporting Figure S3. Prediction heatmap for a support vector machines (SVM) model trained to predict dispersity. Liposome formulations were rigidly defined as monodisperse (PdI ≤ 0.220) or polydispserse (PdI > 0.220). Heatmaps show the predicted formulation, and filled diamonds show the actual experimental outcome.



Supporting Figure S4. Summary of liposome PdI measurements over time. All measurement timepoints were grouped and the aggregate PdI was compared to a reference PdI of 0.220 in order to determine whether liposomes remained stable over the 72 hr period. Formulations are split by **(A)** the empty liposomes (- Curc) and **(B)** those loaded with curcumin (+ Curc), total flow rate as well as the aqueous to organic volume ratio.

Classification model	Predicted accuracy from stratified k-fold cross-validation (mean % ± SD)	Measured accuracy (%)
LOR	87.8 ± 4.1	91.3
LDA	86.9 ± 4.8	89.7
KNN	87.5 ± 3.7	92.1
CART	88.9 ± 4.7	89.7
GNB	84.9 ± 4.6	89.7
SVM	88.7 ± 3.9	92.1

LOR: logistic regression, LDA: linear discriminant analysis, KNN: k-nearest neighbor, CART: classification and regression trees, GNB: Gaussian Naive Bayes, SVM: support vector machines.

Supporting Table S3. Predicted and measured accuracy of various classification models for predicting particle stability (stable or unstable).