Data-driven development of an oral lipid-based nanoparticle formulation of a hydrophobic drug

Zeqing Bao[1], Fion Yung[1], Riley J. Hickman[2,3,4], Alán Aspuru-Guzik[2,3,4,5,6,7,8,9], Pauric Bannigan[1]*, Christine Allen[1,6,9]*

[1]Leslie Dan Faculty of Pharmacy, University of Toronto, Toronto, ON M5S 3M2, Canada

[2]Department of Chemistry, University of Toronto, Toronto, ON M5S 3H6, Canada

[3]Department of Computer Science, University of Toronto, Toronto, ON M5S 2E4, Canada

[4]Vector Institute for Artificial Intelligence, Toronto, ON M5S 1M1, Canada

[5]Lebovic Fellow, Canadian Institute for Advanced Research (CIFAR), Toronto, ON M5S 1M1, Canada

[6]Department of Chemical Engineering & Applied Chemistry, University of Toronto, Toronto, ON M5S 3E5, Canada

[7]Department of Materials Science & Engineering, University of Toronto, Toronto, ON M5S 3E4, Canada

[8]CIFAR Artificial Intelligence Research Chair, Vector Institute, Toronto, ON M5S 1M1, Canada

[9]Acceleration Consortium, Toronto, ON M5S 3H6, Canada

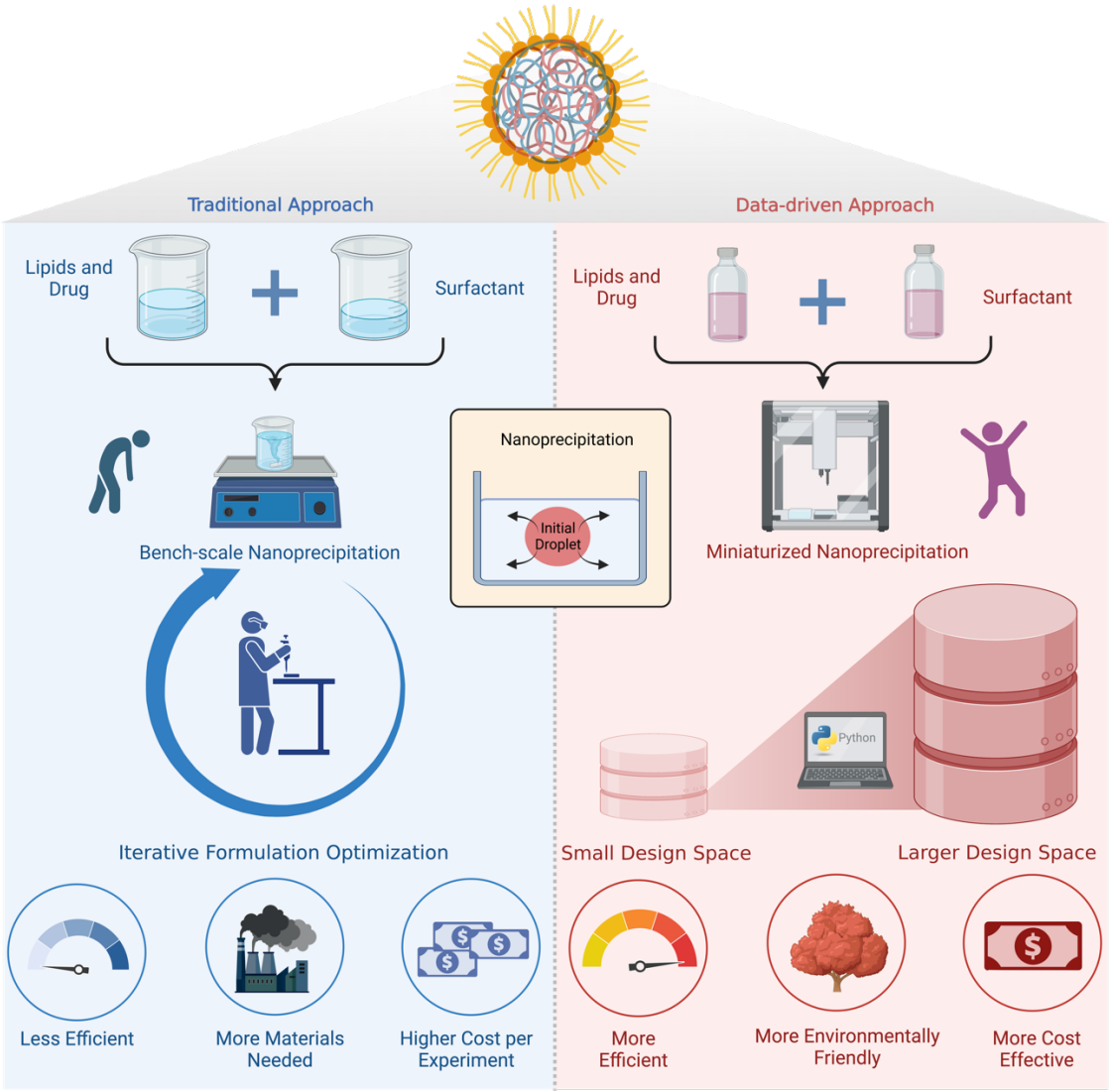
*These authors jointly supervised this work:

Pauric Bannigan (pauric.bannigan@utoronto.ca) and Christine Allen (cj.allen@utoronto.ca)

Leslie Dan Faculty of Pharmacy, University of Toronto, Toronto, ON M5S 3M2, Canada

Abstract

Due to its cost-effectiveness, convenience, and high patient adherence, oral drug administration often remains the preferred approach. Yet, the effective delivery of hydrophobic drugs via the oral route is often hindered by their limited water solubility and first-pass metabolism. To mitigate these challenges, advanced delivery systems such as solid lipid nanoparticles (SLNs) and nanostructured lipid carriers (NLCs) have been developed to encapsulate hydrophobic drugs and enhance their bioavailability. However, traditional design methodologies for these complex formulations often present intricate challenges because they are restricted to a relatively narrow design space. Here, we present a data-driven approach for the accelerated design of SLNs/NLCs encapsulating a model hydrophobic drug, cannabidiol, that combines experimental automation and machine learning. A small subset of formulations, comprising 10% of all formulations in the design space, was prepared in-house, leveraging miniaturized experimental automation to improve throughput and decrease the quantity of drug and materials required. Machine learning models were then trained on the data generated from these formulations and used to predict properties of all SLNs/NLCs within this design space (i.e., estimated to be more than 1200 formulations). Notably, formulations predicted to be high-performers via this approach were confirmed to significantly enhance the solubility of the drug by up to 3000-fold and prevent drug degradation. Moreover, our high-performance formulations significantly enhanced the oral bioavailability of the drug compared to both its free form and an over-the-counter version. Furthermore, this bioavailability matched that of a formulation equivalent in composition to the FDA-approved product, Epidiolex®.



Graphical Abstract

1. Introduction

Oral delivery remains the preferred route of drug administration, accounting for over half of all drug products on the US market [1]. The preference for oral dosage forms is attributed to several advantages including ease of administration, patient compliance, and reduced cost [2]. However, it is estimated that the majority of small-molecule drugs currently in research and development are hydrophobic in nature [3]. These molecules which fall into Biopharmaceutics Classification System (BCS) Classes II and IV are known to have limited bioavailability following administration in conventional oral formulations. As a result, they necessitate the use of advanced oral formulation strategies to fully exploit their therapeutical potential. Such advanced formulations include solid lipid nanoparticles (SLNs) and nanostructured lipid carriers (NLCs), which are gaining increasing attention for the oral delivery of hydrophobic drugs [4–6]. First introduced in the 1990s, SLNs and NLCs [7–9] have been proposed as a means to enhance the oral bioavailability of their cargo. This enhancement in oral bioavailability is attributed in part to a marked improvement in the apparent solubility of the drug. In addition, the small size of the SLN and NLC particles promotes a higher dissolution rate of the drug as well as lymphatic transport which provides a means to avoid first-pass metabolism [10,11].

While SLNs/NLCs offer these advantages, their design process remains complex, partially due to the reliance on an iterative trial-and-error approach. For a given drug, there is a multitude of formulation variables to consider, including the types of excipients, drug-excipient ratios, surfactant concentrations, and various manufacturing parameters. As a result, predicting the formulation outcomes *a priori* becomes a challenge and numerous rounds of drug formulation studies are required before settling on a satisfactory formulation. Thus, a more efficient development strategy is needed to expedite formulation design.

In the past decade, researchers have begun to address this issue by utilizing data-driven methodologies to design SLN/NLCs formulations, with the design of experiment (DoE) approach being the central focus [12–16]. For example, Diwan et al. employed the principles of DoE to design cilnidipine (CND, a BCS Class II drug) loaded SLNs [12]. In comparison to a free aqueous suspension of CND (with 0.5% Tween 80), the SLNs designed via DoE resulted in a twofold increase in oral bioavailability (as measured by $AUC_{0-\infty}$) and a significant improvement in therapeutic efficacy in a rat model of hypertension [12].

Although DoE has been the mainstream data-driven approach for SLN/NLC development, there is a rising interest in the adoption of *in silico* experimentation [17–19]. The rise of self-driving laboratories [20–28]has shown that machine learning (ML) approaches such as Bayesian models are routinely more competitive that traditional DoE to address a variety of chemistry and materials science problems [29–33]. ML approaches can be used to perform experiments virtually, providing a low-cost and effective method for designing a formulation with desirable properties. In our study, we employed ML in conjunction with high-throughput, miniaturized automation to accelerate the formulation development of SLNs and NLCs for a hydrophobic drug, cannabidiol (CBD). CBD, a clinically-approved drug and nutraceutical, was chosen as our BSC class II model compound. Specifically, a subset of CBD-loaded SLN/NLC formulations from the design space were prepared and measured using customized high-throughput automation protocols. ML models were then trained on this subset of SLNs/NLCs to predict the properties of all the formulations within the design space. Subsequent *in vitro* and *in vivo* characterization of the promising formulations identified through this data-driven workflow revealed improvements in the apparent solubility and chemical stability of the drug. As well, the oral bioavailability was improved following administration in healthy rats, in comparison to its free form and an over-the-counter version.

2. Materials and methods

2.1. Materials

Stearic acid (SA, 95%), Carbitol™ (DGME, 2-(2-Ethoxyethoxy)ethanol, 99%), ammonium formate, tetrahydrofuran (THF, HPLC grade), formic acid (FA, 98 - 100%), Pluronic™ F-127 (P407), methanol (LC-MS grade, ≥99.9%), acetonitrile (ACN, LC-MS grade, ≥99.9%), delta-9-tetrahydrocannabinol (THC, 1.0 mg/mL in methanol) were purchased from Sigma Aldrich (ON, CA). Glycerol monostearate (GM), ACN (HPLC grade), and water (LC-MS grade) were purchased from ThermoFisher Scientific (MA, USA). Cannabidiol (CBD, 97.5 - 99%) was purchased from the Ontario Cannabis Store (ON, CA). Compritol® 888 ATO (C888) was a gift from Gattefossé Canada Inc (ON, CA). Tween 80® (Polysorbate 80, reagent grade) was purchased from BioShop (ON, CA). Methanol (HPLC grade) was purchased from Caledon Laboratory Chemicals (ON, CA). Sprague Dawley rat plasma was purchased from Innovative Research (ON, CA).

2.2. ML modeling

2.2.1. ML model development

A dataset consisting of 128 SLN/NLC formulations was generated using the automation screening workflow (Supplementary Information). This dataset was modeled using eight ML regression algorithms, including linear regression (LR), LR with least absolute shrinkage and selection operator regularization (Lasso), decision tree (DT), random forest (RF), light gradient boosting machine (LightGBM), extreme gradient boosting (XGB), support vector regressor (SVR), and a neural network (NN), to infer the statistical relationship between the properties of the excipients and the formulation characteristics including particle size, drug loading capacity (DLC, %), and EE (encapsulation efficiency, %). ML models were developed in Python using the ***Scikit-learn*** package [34] (for LR, Lasso, DT, RF, SVR, and NN), the ***XGBoost*** package [35] (for XGBoost), and the ***LightGBM*** package (for LightGBM) [36]. Five input features, including initial drug-to-lipid ratio (mass of drug per total lipid, wt%), solid lipid type (SA=-1; GM=0; C888=1), solid lipid content (mass of solid lipid per total lipid, wt%), liquid lipid content (mass of liquid lipid per total lipid, wt%), and surfactant concentration (aqueous phase surfactant concentration, w/v%), were used to train ML models to predict particle size, DLC, and EE. To enable muti-objective predictions, we employed the ***MultiOutputRegressor*** function from ***Scikit-learn*** library to extend the single-output regression models to handle multiple target variables. Input features were

standardized by removing the mean and scaling to unit variance using the ***StandardScaler*** function in the ***Scikit-learn*** package [34]. All models were initialized with their default hyperparameter settings and trained based on k-fold cross-validation (k=5, random state=0).

2.2.2. ML model deployment

The best-performing model was utilized to predict the properties (particle size, SLC, and EE) of SLNs/NLCs within the design space, creating a synthetic dataset. The Pareto front of this dataset was determined using the ***Olympus*** package [37,38]. This entailed employing the five input features used to train the models with goals set to minimize particle size while maximizing drug loading levels.

Additionally, we construct an overall performance (OP) metric that defines the cumulative merit of formulations by using the arithmetic mean of the normalized target property values. Each target property is transformed to the interval [0, 10], with 0 being the worst property value and 10 being the best.

$$OP = (Normalized\ size + normalized\ DLC + normalized\ EE)/3$$

2.3. Bench-scale SLN/NLC preparation and characterization

2.3.1. SLN/NLC formulation

Drug-loaded SLNs and NLCs were prepared using a modified nanoprecipitation method [39]. In brief, drug and lipids were dissolved in THF to generate an organic phase. A total of 3 mL of the THF organic phase was then rapidly injected into 15 mL of an aqueous phase (+/- surfactant, prepared in a 20 mL scintillation vial) using a pipette. The final mixture (18 mL) was agitated by magnetic stirring for 10 minutes at 400 rpm. The 20 mL scintillation vials containing the final nanosuspensions were then covered by aluminum foil with a few needle holes and continuously stirred at 50 rpm overnight to allow the solvent to evaporate and particles to harden. Prior to analysis, the particles were filtered using a syringe filter membrane (0.45µm, Millex-HV PVDF Syringe Filters, Sigma, CA) to remove drug or excipient precipitate. Tangential flow (50 nm pore size, MicroKros, Repligen Corporation, US) filtration was then performed to further purify and concentrate the NPs. Specifically, one batch (18 mL) or five batches (90 mL, for lead formulations) of particles were concentrated to a final volume of between 2 and 3 mL. Next, the particles were diluted with 10 mL of MilliQ water and then again concentrated to between 2 and 3 mL. This

process was repeated twice. The final purified, concentrated SLNs/NLCs were then characterized in terms of drug loading level and size.

2.3.2. Size measurements

Filtered particles were diluted using MilliQ water prior to particle size and size polydispersity index (PDI) measurements using a dynamic light scattering instrument (DLS, Zetasizer, Malvern Panalytical, UK). Particles were analyzed in disposable cuvettes (Acrylic cuvette, Sarstedt, Germany) at 25 °C. Each measurement consisted of 10 runs (10 seconds/run). Particle size (z-average particle diameter) and PDI of each sample were recorded as the average of three repeat measurements.

2.3.3. Drug loading analysis

For drug loading analysis, filtered SLNs/NLCs were diluted in methanol for drug extraction. The resulting solutions were then filtered via syringe membrane filters (0.45µm, Millex-HV PVDF Syringe Filters, Sigma, CA). Drug analysis was performed using an Agilent Technologies 1260 Infinity II (Agilent Technologies, Santa Clara, CA, USA) high-performance liquid chromatography (HPLC) system with a diode-array detector (DAD). The HPLC assay was modified from a study previously published by our group [40]. In brief, a Restek EXP Direct Connect Holder guard column and a Restek Raptor ARC-18 column (150 x 4.6 mm, 2.7 µm) were used. The mobile phase, containing 30% (v/v) of 5 mM ammonium formate (prepared using MilliQ water with 0.1 wt% formic acid) and 70% (v/v) acetonitrile, was delivered at a flow rate of 1 mL/min, and CBD was detected at a wavelength of 228 nm.

The DLC and EE were calculated as follows:

$$DLC = \frac{Mass\ of\ drug\ loaded}{Mass\ of\ lipid\ added + Mass\ of\ drug\ loaded} \times 100\%$$

$$EE = \frac{Mass\ of\ drug\ loaded}{Mass\ of\ drug\ added} \times 100\%$$

2.3.4. Formulation morphology

The morphology of the SLNs/NLCs prepared at bench scale was examined using a cryo-transmission electron microscope (cryo-EM, Talos L120C) at the Microscopy Imaging Laboratory (MIL, Temerty Faculty of Medicine, University of Toronto).

2.3.5. In vitro formulation stability

To evaluate the physical stability of the SLNs/NLCs suspensions and the chemical stability of drug encapsulated in NPs, batches of SLNs/NLCs were stored in the dark at room temperature for one month. CBD concentrations (along with particle size and PDI) were measured as previously described at predetermined time points (i.e., Days 0, 7, 14, 21, and 28). For comparison, the stability of unencapsulated and molecularly dissolved CBD (i.e., CBD dissolved in aqueous media with 0.5% Tween 80, v/v) was also investigated under identical conditions.

2.4. Animal studies

2.4.1. In vivo pharmacokinetics studies

Animal studies were approved by the Animal Care Committee at the University of Toronto. Female Sprague-Dawley rats (initial body weight ranging from 200 to 225 grams) were purchased from Charles River and housed in the Division of Comparative Medicine. All animals were allowed to acclimatize for a week prior to the study. On the day of the study, groups of rats received either a CBD-loaded SLN/NLC formulation or CBD in an aqueous suspension (in 0.1% Tween 80) via oral gavage at a CBD dose of 20 mg/kg. Serial blood samples were collected from the saphenous vein at 0.25, 0.5, 1, 2, 4, and 6 h post-administration, using Microvette® CB 300 LH. The animals were sacrificed through cardiac puncture at 8 h, and blood was collected using a BD Vacutainer ® Tube (4 mL). Plasma was separated by centrifugation at 3000 rpm at 4 °C (10 min for Microvette® tubes and 15 min for Vacutainer® tubes) and stored at -80 °C until further processing.

2.4.2. CBD extraction from plasma

CBD was extracted from plasma using a method adapted from our previous work [40]. In brief, 1 µL of THC (0.2 mg/mL in methanol, internal standard) was added to a clean vial and then dried via evaporation under nitrogen at 37°C using a Glas-Col ZipVap evaporator (Terre-Haute, IN, USA). A 50 µL aliquot of plasma was added to the vial (containing THC internal standard) and

then mixed with 450 µL of ACN/ethyl acetate (50:50; v/v). The samples were then vortexed for 10 min and centrifuged at 12,000 rpm for 10 min at 4°C. The supernatant (400 µL) of each sample was transferred to a clean vial. These samples were concentrated via evaporation under nitrogen at 37°C using a Glas-Col ZipVap evaporator (Terre-Haute, IN, USA) followed by reconstitution in 150 µL of methanol. The reconstituted samples were then centrifuged at 12,000 rpm for 10 min at 4 °C, following transfer of 100 µL of supernatant into HPLC vials for analysis using LC-mass spectrometry (LC-MS).

2.4.3. LC-MS method

LC-MS analysis was performed using an Agilent 1260 Infinity HPLC system (Agilent Technologies, Santa Clara, CA, USA) consisting of an InfinityLab Poroshell 120 EC-C18 column (150 × 2.1 mm, 1.9 µm, Agilent Technologies, Santa Clara, CA, USA) and a TSQ Endura™ Triple Quadrupole MS (Thermo Fisher Scientific, Mississauga, ON). The column oven was kept at 40 °C. The mobile phase consisted of water (A) and ACN (B), with both A and B phases containing 0.1% (v/v) formic acid. Gradient elution was programmed as indicated in **Table S2**. The flow rate was set at 0.25 mL/min, and the injection volume was 5 µL. Other detailed LC-MS parameters are included in **Tables S3** and **S4**.

2.4.4. Determination of pharmacokinetic parameters

Noncompartmental pharmacokinetic analysis was performed on individual animal data. The terminal elimination rate constant (k_{el}) was estimated using the regression slope of the log-linear terminal elimination phase, and the elimination half-life ($t_{1/2}$) was calculated using the following equation.

$$t_{1/2} = \frac{\ln(2)}{K_{el}}$$

Maximum plasma concentration (C_{max}) and the time at which it was achieved (T_{max}) were determined for each individual animal. The area under the plasma concentration curves ($AUC_{0\text{-}t}$) was calculated using the linear trapezoidal method (the equation below). $AUC_{0\text{-}\infty}$ was calculated by adding C_{last}/k_{el} to $AUC_{0\text{-}t(last)}$.

$$AUC_{0-t} = \sum_{i=1}^{n-1} \frac{(C_{i+1} + C_i)}{2}(t_{i+1} - t_i)$$

2.4.5. Statistical analysis

Statistical significance was assessed using the Analysis ToolPak in Microsoft Excel (Version 16.65). To evaluate the differences in $AUC_{0\text{-}4H}$ across formulations, t-tests (two-sample, assuming equal variances) were performed with the confidence level set at alpha = 0.05.

3. Results

3.1. Formulation design space

In this study, the formulation design space was defined using excipients that are "generally recognized as safe" (GRAS). This includes three solid lipids (i.e., SA [41–43], GM [44–46], C888 [47–49]), one liquid lipid (DGME [50,51]), and one surfactant (P407 [52,53]). By altering the relative ratios of these excipients and drug, SLNs/NLCs were formulated with varied formulation parameters including solid lipid type, liquid lipid content, initial drug-to-liquid lipid ratio, and surfactant concentration. Given that most formulation parameters, such as the liquid lipid content, are continuous numerical values, they inherently result in an infinite design space. We bounded these numerical parameters, setting a range from 0 to their uppermost values commonly investigated in the literature. Additionally, we discretized the continuous parameter ranges into intervals that were expected to produce discernible effects. For example, the liquid lipid content was explored from 0 to 40% at 5% intervals (wt%). The other numerical parameter ranges were discretized similarly, resulting in nine initial drug-to-lipid ratios (ranging from 0 to 40% at 5% intervals, wt%), and five surfactant concentrations (ranging from 0 to 1% at 0.25% intervals, w/v). Ultimately, this approach culminated in 405 possible parameter combinations, or formulations, for each solid lipid type. Considering all three solid lipids, the total number of formulations within this design space totaled 1215.

3.2. Dataset collection

Initially, around 10% of the possible formulations were prepared in-house using a liquid handling robot as described in the Supplementary Information. This dataset encompassed 128 unique SLN/NLC formulations, with each evaluated in triplicate. This relatively small sample of formulations was then used to train the ML models to predict formulation properties for the entire design space. The 128 formulations were strategically spread across the design space to ensure comprehensive coverage, mitigating the risk of out-of-distribution predictions, a prevalent shortcoming of supervised ML models. As shown in **Figure 1**, the SLNs/NLCs in the dataset displayed a range of physico-chemical characteristics, highlighted by differences in particle size, DLC, and EE.

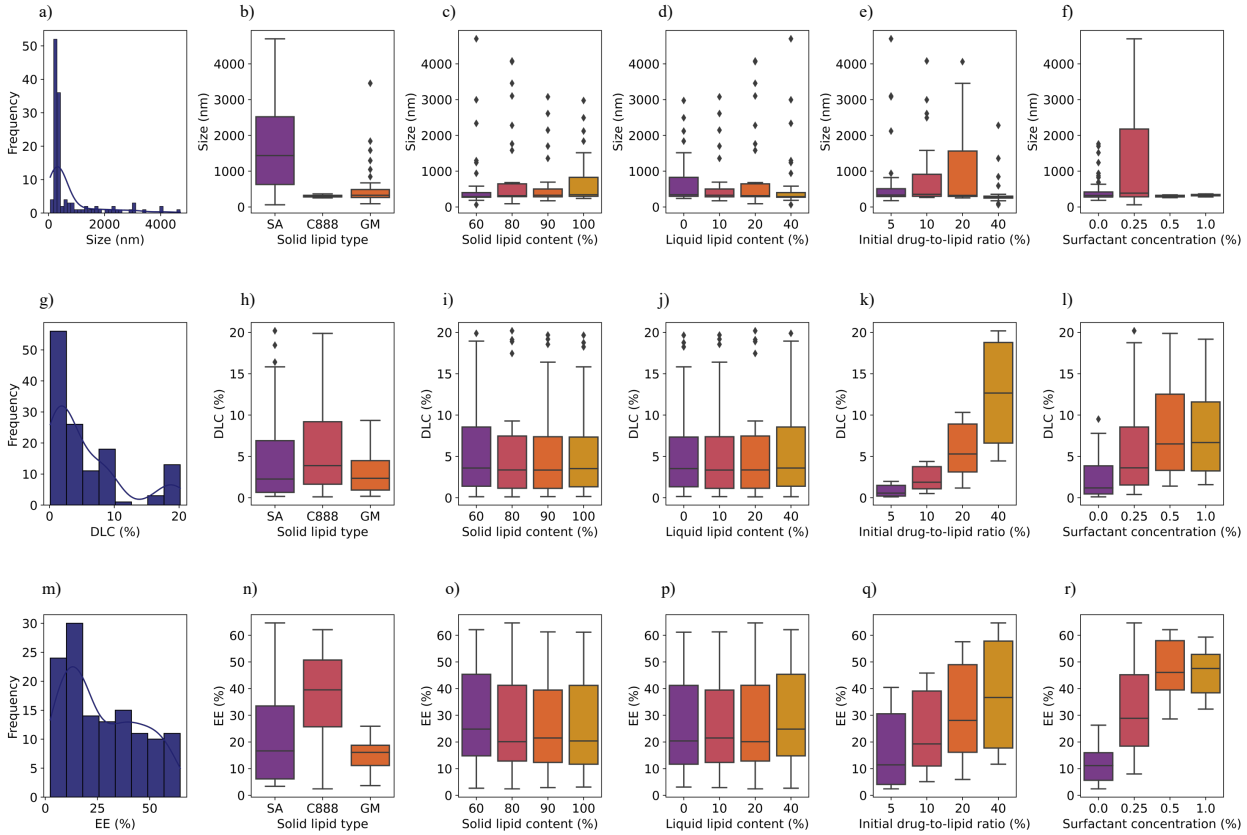


Figure 1. Distributions of a) particle size, g) drug loading capacity (DLC), and m) drug encapsulation efficiency (EE) values for the 128 solid lipid nanoparticle and nanostructured lipid carrier formulations prepared using the automated workflow. The influence of formulation composition parameters such as solid lipid type (b, h, n), solid lipid content (c, i, o), liquid lipid content (d, j, p), initial drug-to-lipid ratio (e, k, q), and surfactant concentration (f, l, r) on particle properties are also summarized.

3.3. ML model development

Using the collected dataset, ML models were trained to predict the formulation properties based on the five formulation composition parameters, including solid lipid type, solid lipid content, liquid lipid content, initial drug-to-lipid ratio, and surfactant concentration. Eight ML models, namely LR, Lasso, DT, RF, LightGBM, XGB, SVR, and NN, were investigated. Each model was wrapped using ***Scikit-learn multioutput*** to enable multi-output prediction (i.e., Size, DLC, and EE), a measure aimed at facilitating training and subsequent utilization. Model performance was evaluated using a five-fold cross-validation technique. During this process, the dataset was randomly split into training (80%) and test (20%) subsets. Models were trained using the training subset and subsequently evaluated with the test subset. This procedure was iterated five times to ensure each sample was used in the evaluation phase once. The performance of the models in predicting each of the outputs using the absolute error metric is summarized in **Figure 2**.

In **Figure 2**, models are organized by decreasing performance from left to right, according to the median absolute error. The DT, XGB, and RF models were found to deliver high and comparable prediction accuracy for all three investigated formulation properties. This finding is further illustrated by **Figure 3a**, which summarizes median absolute error alongside two additional metrics: the Pearson correlation coefficient (PCC) and coefficient of determination (R^2). DT was chosen as the model for the following studies. The predictive performance of the DT is further illustrated in **Figure 3b**, which displays the correlation between the predicted and experimental values. For particle size, a PCC of 0.77 and an R^2 of 0.47 indicate a moderate correlation. For drug loading predictions, the DT model demonstrates high accuracy in predicting both DLC (median absolute error = 0.25%, R^2 = 0.98, PCC = 0.99) and EE (median absolute error = 1.88%, R^2 = 0.95, PCC = 0.97).

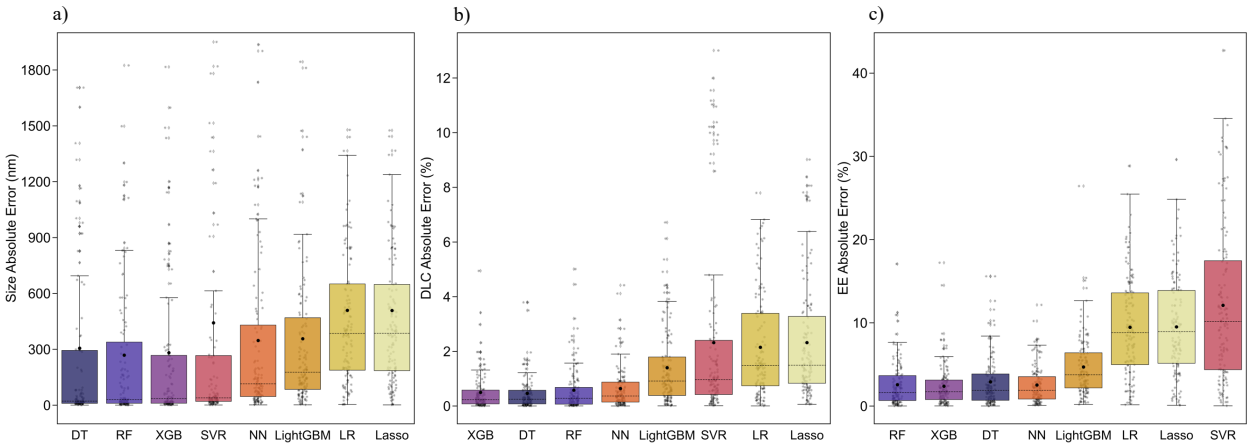


Figure 2. Figures a), b), and c) show the accuracy (absolute errors between the predictions and experimental values) of the eight three-output machine learning regressors in predicting particle size, drug loading capacity (DLC), and encapsulation efficiency (EE), respectively. In each figure, black circles and dashed lines represent the mean and median absolute errors, respectively. Grey dots behind the boxplots show the distribution of all individual absolute errors. The models are arranged in ascending order of median absolute error.

a)

Models	Size			DLC			EE		
	Median absolute error (nm)	R^2	PCC	Median absolute error (%)	R^2	PCC	Median absolute error (%)	R^2	PCC
DT	22	0.47	0.77	0.25	0.98	0.99	1.88	0.95	0.97
RF	30	0.64	0.81	0.28	0.97	0.98	1.62	0.96	0.98
XGB	35	0.58	0.79	0.23	0.98	0.99	1.72	0.96	0.98
SVR	39	-0.18	0.40	0.96	0.52	0.86	10.17	0.28	0.72
NN	115	0.53	0.73	0.37	0.97	0.98	1.89	0.96	0.98
LightGBM	177	0.60	0.77	0.91	0.88	0.94	3.76	0.89	0.94
LR	385	0.38	0.62	1.49	0.75	0.87	8.82	0.62	0.79
Lasso	386	0.38	0.62	1.49	0.70	0.86	8.94	0.62	0.79

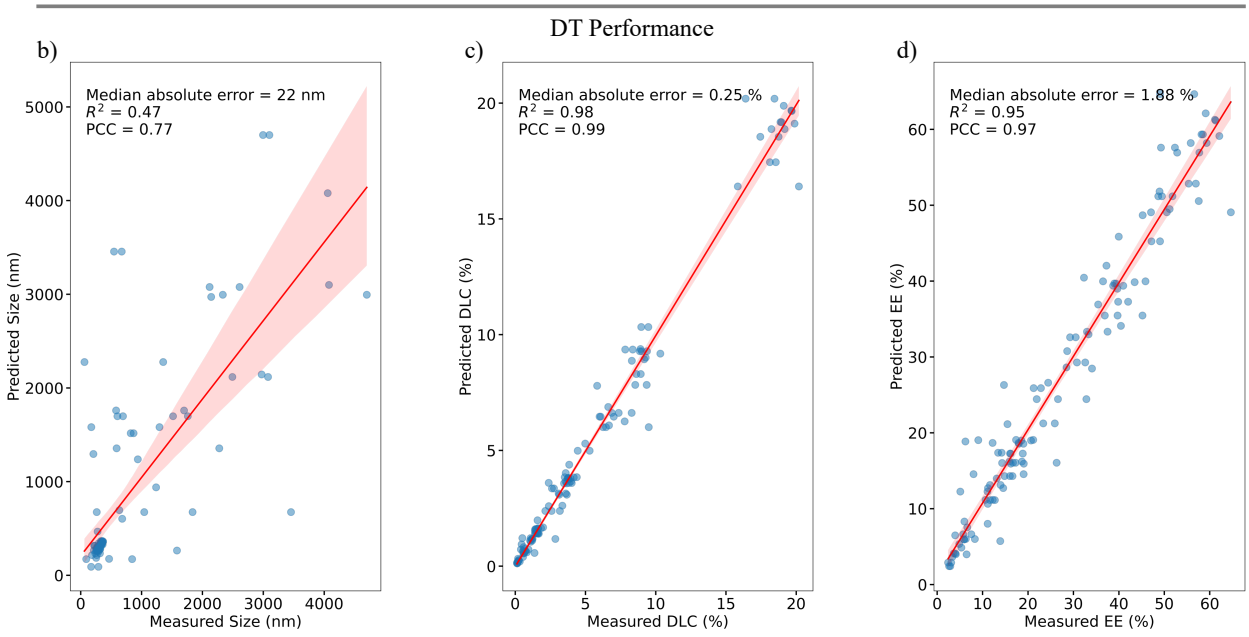


Figure 3. a) summarizes the accuracy (median absolute error, R^2, and Pearson's correlation coefficient) of the eight, three-output machine learning regressors for predicting particle size, drug loading capacity (DLC), and encapsulation efficiency (EE). Three models, including decision tree (DT), random forest (RF), and extreme gradient boosting (XGB), exhibit high and comparable predictive accuracy. For simplicity and efficiency, the decision tree model was selected due to its straightforward structure and high training speed. The accuracy of the decision tree model is visualized in the scatter plots for b) Size, c) DLC, and d) EE predictions.

3.4. ML model deployment

3.4.1. ML model prediction

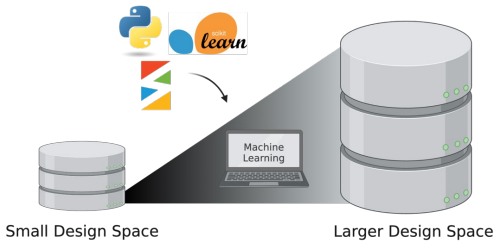
Having been identified as the best multi-output model, the DT model was employed to predict the properties of the 1215 formulations within the design space, thereby generating a larger synthetic dataset (**Figure 4a**). To enable comparison, the particle size (from highest to lowest values in the dataset) and drug loading (DLC and EE, from lowest to highest values in the dataset) were scaled from 0 (worst) to 10 (best). These are the three properties that we consider to be indicative of formulation performance, and the OP of a formulation was determined by averaging the scaled values for size, DLC, and EE, serving as a quantitative measure of formulation performance.

3.4.2. ML model validation

The OP values for the total design space (containing 1215 formulations) are summarized in **Figure S4**, with a breakdown of OP for SLNs/NLCs formulated using distinct types of solid lipids. The median OP for C888-based SLNs/NLCs was found to be higher than that for SA-based and GM-based formulations. Thus, seven C888-based SLN/NLC formulations that were predicted to be high (H1, H2, and H3, OP>=9), medium (M1 and M2, 7<=OP<8), or low (L1 and L2, 5<=OP<6) performers based on their properties were selected for experimental validation of the models (**Figure 4b**). For a visual representation of these formulations within the larger design space, they are plotted on a three-dimensional plot of formulation properties alongside all other formulations, with the Pareto set highlighted in purple. The Pareto front is a well-established concept utilized in multi-objective optimization tasks and defines the constraints of optimization whereby the enhancement of one target unavoidably results in the compromise of another. As shown in **Figure 4c**, formulations predicted to be high performers (H1, H2, H3), based on their properties, are located in close proximity to the Pareto front, contrasting with the medium and low performers situated notably further away. This suggests that these high-performing formulations are likely to be representative of an ensemble of optimal formulations within the dataset. To validate the DT model's predictions, the formulations predicted to be high, medium, and low performers were formulated at bench scale and characterized. The properties of these formulations are shown in **Figures 4d, e and f**. In general, formulations predicted to be high performers exhibited consistent results, surpassing those predicted to have low or medium performance. In particular, high-performing formulations were characterized by DLC values at 15%, which is two-fold greater than

the medium-performance formulations (8%) and over five-fold greater than the low-performance formulations (<3%).

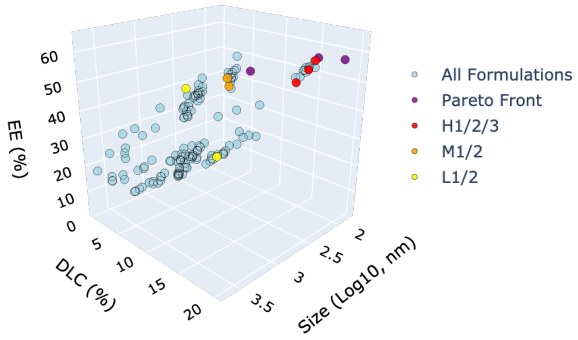
a)



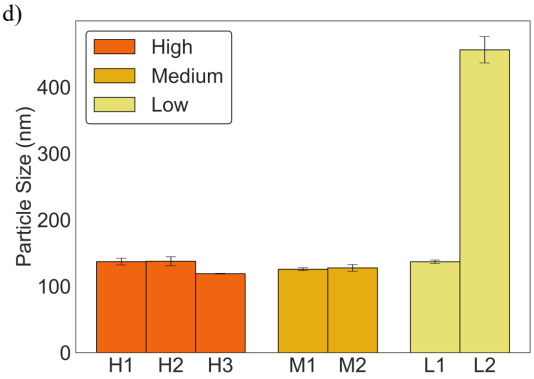
b)

Formulations predicted to be high, medium or low performers		Solid lipid type	Solid lipid content (%)	Liquid lipid content (%)	Initial drug-to-lipid ratio (%)	Surfactant concentration (%)
High [9, 10)	H1	C888	100	0	40	0.25
	H2	C888	95	5	40	0.5
	H3	C888	80	20	40	0.25
Medium [7, 8)	M1	C888	100	0	20	0.75
	M2	C888	80	20	20	0.25
Low [5, 6)	L1	C888	100	0	5	0.50
	L2	C888	65	35	30	0.00

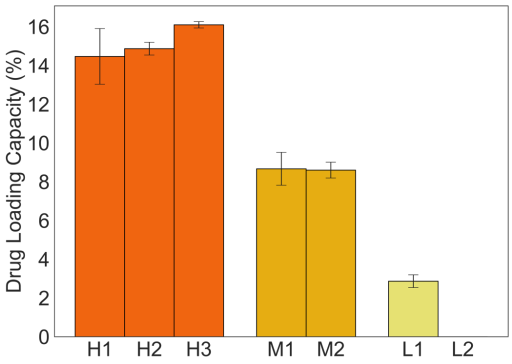
c)



d)



e)



f)

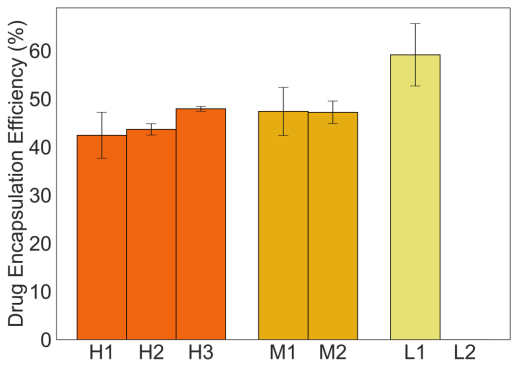


Figure 4. a) Shows the use of machine learning models trained on a subset of the design space to predict the properties of formulations in the unexplored design space. b) Shows the seven formulations, that were predicted to be high (H1/2/3), medium (M1/2), or low (L1/2) performers, and selected for experimental validation of the model. c) Presents a 3D surface plot depicting the predicted properties of the 1215 formulations within the design space: size on the x-axis, drug loading capacity (DLC) on the y-axis, and encapsulation efficiency (EE) on the z-axis. Formulations on the Pareto front are shown in purple. Formulations are color-coded with high, medium, and low performers in red, orange, and yellow, respectively. d-f) Summarize the d) particle size, e) DLC, and f) EE of the seven selected formulations prepared via stirring-based bench-scale nanoprecipitation.

3.5. SLN/NLC characterization

3.5.1. In vitro characterization

Following validation of the models, one SLN (i.e., H1) and one NLC (i.e., H3) were selected for further *in vitro* and *in vivo* characterization. For this additional characterization, the formulations were prepared at "bench-scale" followed by concentration and purification via tangential flow. The concentration of the drug (encapsulated in SLNs/NLCs) following tangential flow of the SLN and NLC formulations was found to be 2.3 ± 0.1 mg/mL and 2.4 ± 0.4 mg/mL, respectively. These concentrations are 3000 times higher than the solubility of CBD in water (i.e., 0.7 μg/mL [54]). Another key advantage of formulating drugs in SLNs/NLCs is to prevent degradation [55,56]. Indeed, CBD encapsulated in the SLN and NLC was more chemically stable than CBD molecularly dissolved in 0.5% (v/v) Tween80. Specifically, only 68±6.1% of the initial CBD added to the CBD control was found to be present after seven days of storage (room temperature in darkness), while the content of CBD encapsulated in the SLN and NLC remained unchanged under the same conditions for up to one month (**Figure S5**).

As measured by DLS, the SLN and NLC particles were found to be nanosized (140±4.2 nm for SLN and 124±3.7 nm for NLC) with a low PDI (<0.2 for both formulations), and their size and PDI were found to remain unchanged during storage over a one-month period (**Figure S6**). As evidenced by cryo-EM analysis, the particles were spherical with a diameter of less than 200 nm (**Figure 5**), which agreed with DLS measurements. Additional *in vitro* characterization of the SLN/NLC formulations (i.e., differential scanning calorimetry and Fourier transform infrared spectroscopy analysis) is included in the Supplementary Information (**Figures S7-S9**).

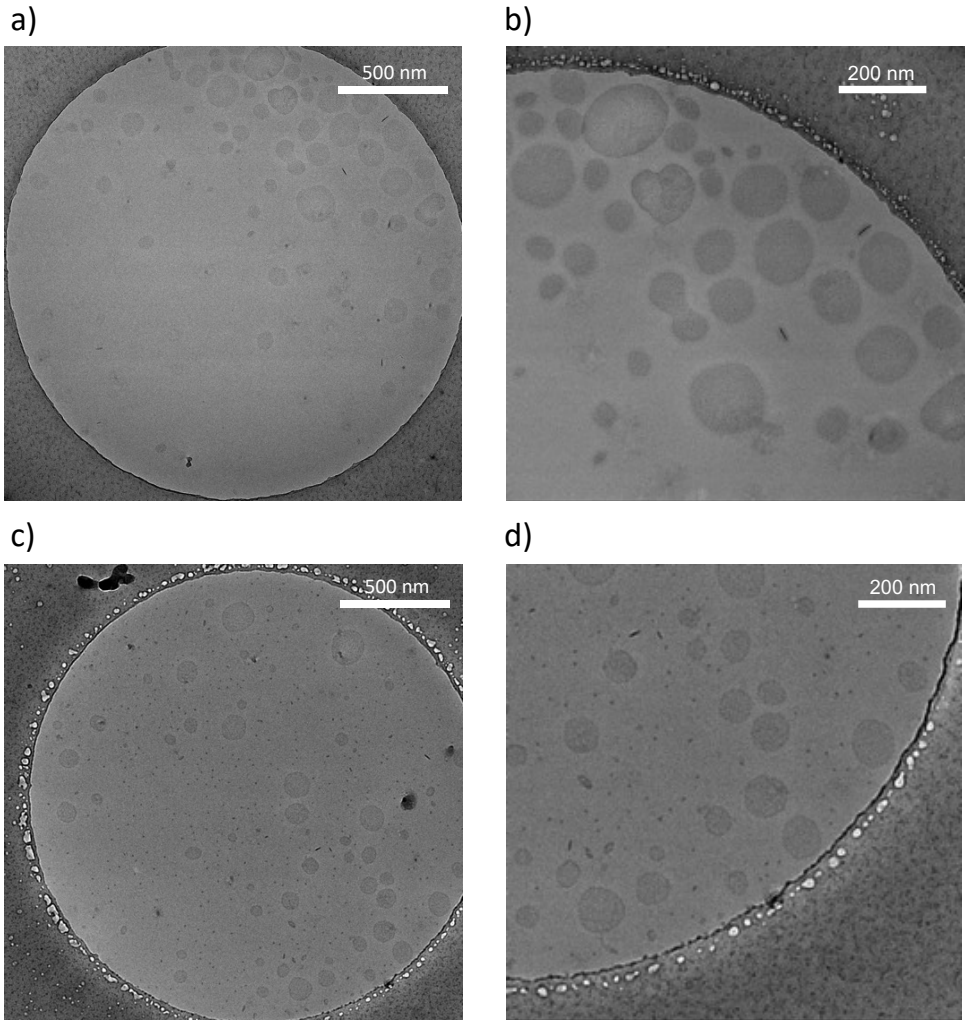


Figure 5. Figures a) and c) include two representative cryo-EM images that display the morphologies of the selected solid lipid nanoparticle and nanostructured lipid carrier formulations, respectively. Figures b) and d) exhibit the same images under increased magnification, providing a more intricate visualization of the solid lipid nanoparticle and nanostructured lipid carrier structures, respectively.

3.5.2. In vivo characterization

The selected SLN and NLC were evaluated *in vivo* following oral administration to normal, female Sprague Dawley rats at a dose level of 20 mg CBD/kg body weight. As shown in **Figure 6a**, administration of CBD in the SLN or NLC formulations results in similar pharmacokinetics with peak plasma concentration of drug reached at one hour. In comparison, CBD administered in 0.1% Tween 80 resulted in plasma concentrations that remained below the detection limit of the LC-MS assay (i.e., 3 ng/mL).

The area under the CBD plasma concentration-time curve during the first four hours of administration (i.e., $AUC_{0\text{-}4H}$) was calculated and is shown in **Figure 6b**. For comparison purposes, the $AUC_{0\text{-}4H}$ of two oil-based CBD formulations previously published by our group, medium-chain triglyceride-based formulation (MCT-CBD) and sesame oil-based formulation (SO-CBD), were also included (**Figure 6b**) [40]. These formulations were used for comparison as MCT oil is the most commonly used oil carrier for CBD, and the SO-CBD formulation is equivalent in composition to the FDA-approved product, Epidiolex®. **Figure 6b** shows that the SLN and NLC, formulations resulted in $AUC_{0\text{-}4H}$ values of 1410±218 and 1226±859 ng·h/mL, respectively. These results suggest a 15-fold improvement in CBD exposure (in the first 4 hours) relative to that achieved with MCT-CBD ($AUC_{0\text{-}4H}$=80±45 ng·h/mL, $p<0.05$). Furthermore, these formulations demonstrated comparable CBD exposure to SO-CBD ($AUC_{0\text{-}4H}$=1497±332 ng·h/mL, $p>0.05$), suggesting their potential as viable alternatives to the FDA-approved drug product, which contains ethanol and the recognized allergen sesame oil.

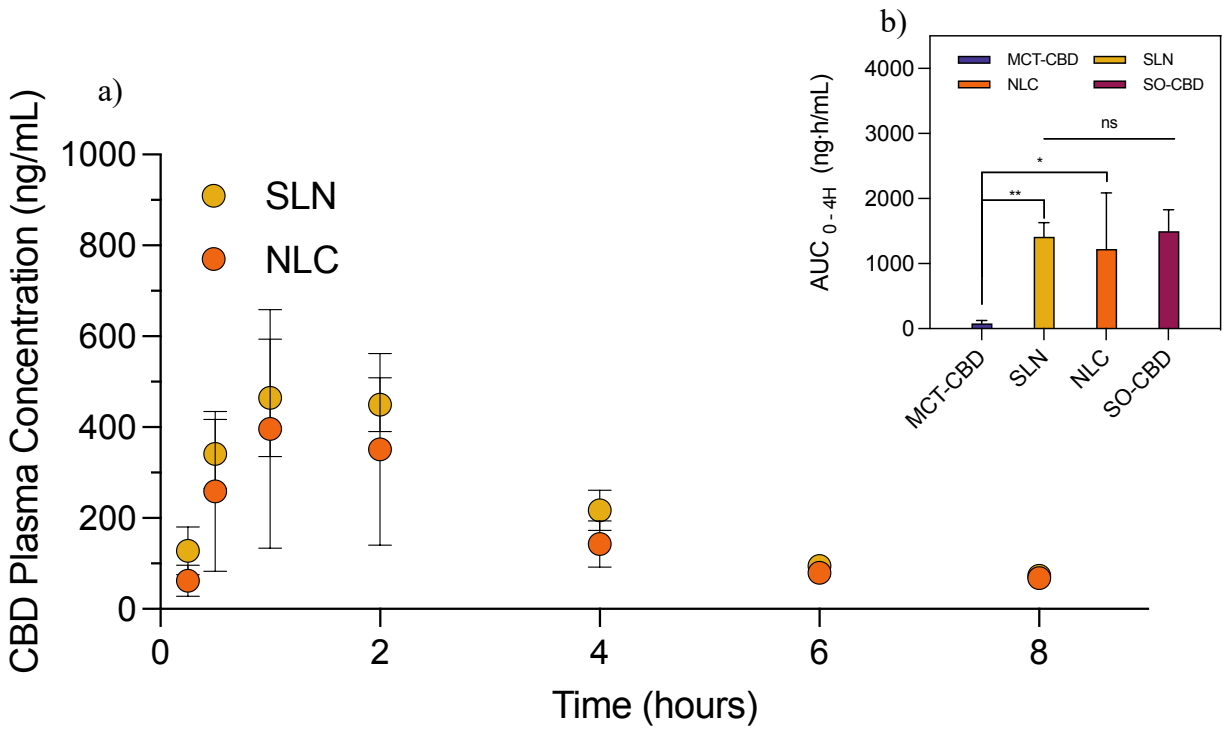


Figure 6. The pharmacokinetic profiles of CBD following oral administration of the selected solid lipid nanoparticle (SLN) and nanostructured lipid carrier formulations (NLC) in Sprague Dawley rats at a dose of 20 mg CBD/kg body weight. Blood samples were taken at predetermined timepoints to analyze the plasma concentrations of CBD as summarized in a), n=4 or 5. Both SLN and NLC formulations result in similar pharmacokinetics for CBD with plasma concentrations peaking at one hour. Exposure to CBD (0 to 4 hours) following administration in the SLNs/NLCs were plotted in b) in comparison to two previously published formulations (MCT-CBD and SO-CBD [40]).

4. Discussion

4.1. Combination of automated experimentation and ML to expedite SLN/NLC design

It is well understood that the physico-chemical properties such as size, DLC, and EE of SLNs/NLCs are influenced by various formulation and processing parameters. Nevertheless, at this time experimental studies are required to confirm and optimize the properties of these formulations. This reduces the overall efficiency of drug product optimization and increases the costs associated with drug development [57,58]. Thus, there is an urgent need to update conventional strategies for formulation development.

For several decades, the DoE approach has been employed to accelerate drug formulation design [59–61]. This data-driven approach maps input-output relationships to save both time and resources during optimization [59,62]. For instance, a recent study by Ramzan al. utilized Taguchi DoE methodology to facilitate the design of drug-loaded SLNs [63]. In that study, the design space of interest consisted of seven variables, and each variable had two levels, resulting in 128 formulations. Employing Taguchi DOE, the number of experiments required to map this design space were reduced from 128 to just eight, significantly decreasing the workload [63]. This DoE method was further paired with linear regression, facilitating the quantification of the influence of variables on the targets.

Compared to DoE, our proposed workflow is particularly advantageous when dealing with a formulation workflow that is compatible with miniaturized automation (e.g., nanoprecipitation). The automation of experiments facilitates performing multiple experiments with reduced batch sizes (e.g., 96 well plate). Therefore, compared to DoE that typically explores selected points within the design space, our approach is able to cover a broader spectrum of the design space resulting in a larger dataset, using equal or even less time and resources [64–66]. This relatively larger dataset enables the use of advanced and more parametrized ML models, as opposed to linear regression, to gain formulation insights. As illustrated in **Figures 2 and 3**, linear regression is not the best model to fit this dataset. In contrast, the more advanced ML models (e.g., tree-based models) that are capable of handling complex input-output relationships deliver higher prediction accuracy. This highlights the limitation of using only linear regression given that the correlation between formulation composition parameters and formulation properties or performance is typically complex and non-linear [57,58].

4.2. Optimal formulations identified through the proposed data-driven workflow

Pareto optimality is a concept in multi-objective optimization strategy that has been widely used in engineering design and many other areas [67,68]. It defines a set of solutions or samples wherein improving one target property invariably necessitates a compromise on others [69]. Optimal performance in terms of a single property does not necessarily suggest the ideal overall performance. In the context of pharmaceutical sciences, formulations that balance all target properties are typically preferred. To overcome this limitation, we combined analysis of Pareto optimal formulations with a customized weighted sum metric, i.e., OP. To compute the OP, the size, DLC, and EE in the dataset were normalized on a scale of 0 (worst) to 10 (best). The OP was then calculated as the mean of the normalized size, DLC, and EE. Formulations with an OP score greater than 9 were categorized as high-performance formulations. As shown in **Figure 4c**, high-performance formulations were indeed closer in proximity to the Pareto front than their medium and low-performance counterparts. To validate the predictions, several formulations were scaled up to bench scale. Overall, the formulations predicted to be high performers were found to have properties that were comparable or superior to the predicted medium and low performers. Within the dataset, around 100 formulations (~8%) could be considered high-performers (OP>=9). It is possible that experienced formulation scientists could achieve or surpass this success rate given their expertise; however, less experienced scientists may not be able to identify this number of promising formulations with the same ease.

4.3. Use of SLNs/NLCs to facilitate the oral delivery of small molecule drugs

Oral delivery of hydrophobic drugs, such as those categorized as BCS Class II, face significant limitations due to their low solubility, which can result in poor bioavailability [70,71]. A typical example is the hydrophobic drug CBD used in this study; a BCS Class II drug with a reported aqueous solubility of only 0.7 µg/mL [54] and reported oral bioavailability of 6% [72]. SLNs and NLCs have been shown to enhance the bioavailability of hydrophobic drugs [73–75]. Similarly, our research revealed a marked enhancement in the absorption and pharmacokinetics of orally administered CBD. Our *in vivo* study confirmed that both SLN and NLC formulations that were predicted to be high performers significantly increased the bioavailability of CBD when compared to its unformulated counterpart. This improved bioavailability was comparable (p>0.05) to that achieved with a formulation equivalent in composition to the only FDA-approved CBD drug product, Epidiolex®.

SLN and NLC drug formulations hold significant therapeutic promise. While no product based on these technologies has received approval yet, it is important to consider that SLN/NLC technologies have been in existence for roughly 30 years. This timeline is on par with liposomes, which also took about 30 years to reach the market after their introduction. Although the production of SLNs and NLCs may have higher costs compared to traditional oral dosage forms such as tablets and capsules, their potential benefits could offer substantial advancements in drug delivery. For SLN/NLC-based products to become popular oral formulation alternatives, they need to exhibit significantly enhanced performance compared to these conventional oral drug products. This study aimed to showcase SLNs/NLCs as an example, highlighting how our proposed workflow can effectively streamline complex formulation development processes.

4.4. Limitations

While this study makes a critical step towards data-driven drug formulation development, it is important to recognize its limitations. Standard nanoprecipitation methods typically necessitate stirring and/or sonication as a source of dispersion energy [76]. This enhances the diffusion rate of organic phase to the aqueous phase, typically facilitating the formation of smaller particles. However, the execution of stirring or sonication within each of the 96 wells remains a technical challenge for our current setup. This limitation results in a difference in particle size between the particles generated by automation and those produced through scale-up methods involving stirring and/or sonication. As evidenced in **Figure S3**, the same SLN/NLC formulations exhibit different sizes when prepared using automation (~300 nm) versus the standard bench-top method (~100 nm). Despite the size difference between particles prepared using the automated miniaturized setup and the bench scale process, the predicted particle size trend (small vs. large) from the models remained consistent.

5. Conclusion

We discovered high-performing small molecule-loaded SLNs/NLCs by means of a holistic data-driven approach that involves the combination of experimental automation and ML. A total of 128 formulations, accounting for around 10% of the investigated design space, were prepared in-house using a liquid handling robot. This generated dataset was then modelled using ML approaches to determine the composition-property relationships for this series of formulations. Employing the developed ML models, the properties of all formulations in this design space were predicted,

resulting in a synthetic dataset that is ten-fold larger than the initial dataset. The formulations predicted to be top-performers were nanosized, and able to markedly improve CBD solubility in aqueous media. Regarding their *in vivo* performance, these formulations significantly enhanced the oral bioavailability of the drug compared to CBD administered in conventional surfactant solution and MCT oil. The formulations also demonstrated CBD exposure comparable to SO-CBD, which is equivalent in composition to an FDA-approved CBD product. This study highlights the efficacy of our data-driven approach in rapidly advancing the development of lipid-based NP formulations of a hydrophobic small molecule.

Data availability

To allow reproducibility, the dataset and codes for the current study are available from the corresponding author on reasonable request.

Abbreviations

SLN: Solid lipid nanoparticle; NLC: Nanostructured lipid carrier; BCS: Biopharmaceutics Classification System; DoE: Design of experiment; CND: Cilnidipine; ML: Machine learning; CBD: Cannabidiol; SA: Stearic acid; DGME: 2-(2-Ethoxyethoxy)ethanol; THF: Tetrahydrofuran; P407: Pluronic^{TM} F-127; ACN: Acetonitrile; THC: Delta-9-tetrahydrocannabinol; C888: Compritol® 888 ATO; LR: Linear regression; Lasso: LR with least absolute shrinkage and selection operator regularization; DT: Decision tree; RF: Random forest; LightGBM: Light gradient boosting machine; XGB: Extreme gradient boosting; SVR: Support vector regressor; NN: Neural network; DLC: Drug loading capacity; EE: Encapsulation efficiency; OP: Overall performance; PDI: Polydispersity index; DLS: Dynamic light scattering instrument; HPLC: High-performance liquid chromatography; DAD: Diode-array detector; Cryo-EM: Cryo-transmission electron microscope; K_{el}: Terminal elimination rate constant; $T_{1/2}$: Elimination half-life; C_{max}: Maximum plasma concentration; T_{max}: The time at which C_{max} was achieved; GRAS: Generally recognized as safe; PCC: Pearson correlation coefficient; R^2: Coefficient of determination.

References

[1] H. Zhong, G. Chan, Y. Hu, H. Hu, D. Ouyang, A Comprehensive Map of FDA-Approved Pharmaceutical Products, Pharmaceutics. 10 (2018) 263. https://doi.org/10.3390/pharmaceutics10040263.

[2] A.C. Anselmo, Y. Gokarn, S. Mitragotri, Non-invasive delivery strategies for biologics, Nat Rev Drug Discov. 18 (2019) 19–40. https://doi.org/10.1038/nrd.2018.183.

[3] J. Leganés, A.M. Rodríguez, M.A. Arranz, C.A. Castillo-Sarmiento, I. Ballesteros-Yáñez, A.S. Migallón, S. Merino, E. Vázquez, Magnetically responsive hydrophobic pockets for on–off drug release, Materials Today Chemistry. 23 (2022) 100702. https://doi.org/10.1016/j.mtchem.2021.100702.

[4] M. Azman, A.H. Sabri, Q.K. Anjani, M.F. Mustaffa, K.A. Hamid, Intestinal Absorption Study: Challenges and Absorption Enhancement Strategies in Improving Oral Drug Delivery, Pharmaceuticals. 15 (2022) 975. https://doi.org/10.3390/ph15080975.

[5] S. Sheoran, S. Arora, R. Samsonraj, P. Govindaiah, S. vuree, Lipid-based nanoparticles for treatment of cancer, Heliyon. 8 (2022) e09403. https://doi.org/10.1016/j.heliyon.2022.e09403.

[6] B. Maddiboyina, Ramaiah, R.K. Nakkala, H. Roy, Perspectives on cutting-edge nanoparticulate drug delivery technologies based on lipids and their applications, Chemical Biology & Drug Design. 102 (2023) 377–394. https://doi.org/10.1111/cbdd.14230.

[7] M. Üner, G. Yener, Importance of solid lipid nanoparticles (SLN) in various administration routes and future perspectives, Int J Nanomedicine. 2 (2007) 289–300.

[8] K.L. López, A. Ravasio, J.V. González-Aramundiz, F.C. Zacconi, Solid Lipid Nanoparticles (SLN) and Nanostructured Lipid Carriers (NLC) Prepared by Microwave and Ultrasound-Assisted Synthesis: Promising Green Strategies for the Nanoworld, Pharmaceutics. 15 (2023) 1333. https://doi.org/10.3390/pharmaceutics15051333.

[9] P. Ganesan, D. Narayanasamy, Lipid nanoparticles: Different preparation techniques, characterization, hurdles, and strategies for the production of solid lipid nanoparticles and nanostructured lipid carriers for oral drug delivery, Sustainable Chemistry and Pharmacy. 6 (2017) 37–56. https://doi.org/10.1016/j.scp.2017.07.002.

[10] G. Poovi, N. Damodharan, Lipid nanoparticles: A challenging approach for oral delivery of BCS Class-II drugs, Future Journal of Pharmaceutical Sciences. 4 (2018) 191–205. https://doi.org/10.1016/j.fjps.2018.04.001.

[11] K. Elbrink, S. Van Hees, D. Roelant, T. Loomans, R. Holm, F. Kiekens, The influence on the oral bioavailability of solubilized and suspended drug in a lipid nanoparticle formulation: In vitro and in vivo evaluation, European Journal of Pharmaceutics and Biopharmaceutics. 179 (2022) 1–10. https://doi.org/10.1016/j.ejpb.2022.08.010.

[12] R. Diwan, P.R. Ravi, N.S. Pathare, V. Aggarwal, Pharmacodynamic, pharmacokinetic and physical characterization of cilnidipine loaded solid lipid nanoparticles for oral delivery optimized using the principles of design of experiments, Colloids Surf B Biointerfaces. 193 (2020) 111073. https://doi.org/10.1016/j.colsurfb.2020.111073.

[13] M.K. Shah, P. Madan, S. Lin, Preparation, in vitro evaluation and statistical optimization of carvedilol-loaded solid lipid nanoparticles for lymphatic absorption via oral administration, Pharmaceutical Development and Technology. 19 (2014) 475–485. https://doi.org/10.3109/10837450.2013.795169.

[14] J. Hao, F. Wang, X. Wang, D. Zhang, Y. Bi, Y. Gao, X. Zhao, Q. Zhang, Development and optimization of baicalin-loaded solid lipid nanoparticles prepared by coacervation method using central composite design, Eur J Pharm Sci. 47 (2012) 497–505. https://doi.org/10.1016/j.ejps.2012.07.006.

[15] S.V. Talluri, G. Kuppusamy, V.V.S.R. Karri, K. Yamjala, A. Wadhwani, S.V. Madhunapantula, S.S.S. Pindiprolu, Application of quality-by-design approach to optimize diallyl disulfide-loaded solid lipid

nanoparticles, Artif Cells Nanomed Biotechnol. 45 (2017) 474–488. https://doi.org/10.3109/21691401.2016.1173046.
[16] G.U. Sailor, V.D. Ramani, N. Shah, G.R. Parmar, D. Gohil, R. Balaraman, A. Seth, Design of Experiment Approach Based Formulation Optimization of Berberine Loaded Solid Lipid Nanoparticle for Antihyperlipidemic Activity, Indian Journal of Pharmaceutical Sciences. 83 (2021) 204–218. https://doi.org/10.36468/pharmaceutical-sciences.766.
[17] M.H. Elkomy, M. Elmowafy, K. Shalaby, A.F. Azmy, N. Ahmad, A. Zafar, H.M. Eid, Development and machine-learning optimization of mucoadhesive nanostructured lipid carriers loaded with fluconazole for treatment of oral candidiasis, Drug Dev Ind Pharm. 47 (2021) 246–258. https://doi.org/10.1080/03639045.2020.1871005.
[18] A.A. Ozturk, A.B. Gunduz, O. Ozisik, Supervised Machine Learning Algorithms for Evaluation of Solid Lipid Nanoparticles and Particle Size, Comb. Chem. High Throughput Screen. 21 (2018) 693–699. https://doi.org/10.2174/1386207322666181218160704.
[19] R.M. Hathout, A.A. Metwally, Towards better modelling of drug-loading in solid lipid nanoparticles: Molecular dynamics, docking experiments and Gaussian Processes machine learning, European Journal of Pharmaceutics and Biopharmaceutics. 108 (2016) 262–268. https://doi.org/10.1016/j.ejpb.2016.07.019.
[20] F. Häse, L.M. Roch, A. Aspuru-Guzik, Next-Generation Experimentation with Self-Driving Laboratories, Trends in Chemistry. 1 (2019) 282–291. https://doi.org/10.1016/j.trechm.2019.02.007.
[21] H.S. Stein, J.M. Gregoire, Progress and prospects for accelerating materials science with automated and autonomous workflows, Chem. Sci. 10 (2019) 9640–9649. https://doi.org/10.1039/C9SC03766G.
[22] C.W. Coley, N.S. Eyke, K.F. Jensen, Autonomous Discovery in the Chemical Sciences Part I: Progress, Angewandte Chemie International Edition. 59 (2020) 22858–22893. https://doi.org/10.1002/anie.201909987.
[23] C.W. Coley, N.S. Eyke, K.F. Jensen, Autonomous Discovery in the Chemical Sciences Part II: Outlook, Angewandte Chemie International Edition. 59 (2020) 23414–23436. https://doi.org/10.1002/anie.201909989.
[24] M.M. Flores-Leonar, L.M. Mejía-Mendoza, A. Aguilar-Granda, B. Sanchez-Lengeling, H. Tribukait, C. Amador-Bedolla, A. Aspuru-Guzik, Materials Acceleration Platforms: On the way to autonomous experimentation, Current Opinion in Green and Sustainable Chemistry. 25 (2020) 100370. https://doi.org/10.1016/j.cogsc.2020.100370.
[25] J.A. Bennett, M. Abolhasani, Autonomous chemical science and engineering enabled by self-driving laboratories, Current Opinion in Chemical Engineering. 36 (2022) 100831. https://doi.org/10.1016/j.coche.2022.100831.
[26] E. Stach, B. DeCost, A.G. Kusne, J. Hattrick-Simpers, K.A. Brown, K.G. Reyes, J. Schrier, S. Billinge, T. Buonassisi, I. Foster, C.P. Gomes, J.M. Gregoire, A. Mehta, J. Montoya, E. Olivetti, C. Park, E. Rotenberg, S.K. Saikin, S. Smullin, V. Stanev, B. Maruyama, Autonomous experimentation systems for materials development: A community perspective, Matter. 4 (2021) 2702–2726. https://doi.org/10.1016/j.matt.2021.06.036.
[27] K. Hippalgaonkar, Q. Li, X. Wang, J.W. Fisher, J. Kirkpatrick, T. Buonassisi, Knowledge-integrated machine learning for materials: lessons from gameplaying and robotics, Nat Rev Mater. 8 (2023) 241–260. https://doi.org/10.1038/s41578-022-00513-1.
[28] R.J. Hickman, P. Bannigan, Z. Bao, A. Aspuru-Guzik, C. Allen, Self-driving laboratories: A paradigm shift in nanomedicine development, Matter. 6 (2023) 1071–1081. https://doi.org/10.1016/j.matt.2023.02.007.

[29] F. Häse, L.M. Roch, C. Kreisbeck, A. Aspuru-Guzik, Phoenics: A Bayesian Optimizer for Chemistry, ACS Cent. Sci. 4 (2018) 1134–1145. https://doi.org/10.1021/acscentsci.8b00307.
[30] L.M. Roch, F. Häse, C. Kreisbeck, T. Tamayo-Mendoza, L.P.E. Yunker, J.E. Hein, A. Aspuru-Guzik, ChemOS: An orchestration software to democratize autonomous discovery, PLOS ONE. 15 (2020) e0229862. https://doi.org/10.1371/journal.pone.0229862.
[31] M. Sim, M.G. Vakili, F. Strieth-Kalthoff, H. Hao, R. Hickman, S. Miret, S. Pablo-García, A. Aspuru-Guzik, ChemOS 2.0: an orchestration architecture for chemical self-driving laboratories, (2023). https://doi.org/10.26434/chemrxiv-2023-v2khf.
[32] B.P. MacLeod, F.G.L. Parlane, T.D. Morrissey, F. Häse, L.M. Roch, K.E. Dettelbach, R. Moreira, L.P.E. Yunker, M.B. Rooney, J.R. Deeth, V. Lai, G.J. Ng, H. Situ, R.H. Zhang, M.S. Elliott, T.H. Haley, D.J. Dvorak, A. Aspuru-Guzik, J.E. Hein, C.P. Berlinguette, Self-driving laboratory for accelerated discovery of thin-film materials, Science Advances. 6 (2020) eaaz8867. https://doi.org/10.1126/sciadv.aaz8867.
[33] B.P. MacLeod, F.G.L. Parlane, C.C. Rupnow, K.E. Dettelbach, M.S. Elliott, T.D. Morrissey, T.H. Haley, O. Proskurin, M.B. Rooney, N. Taherimakhsousi, D.J. Dvorak, H.N. Chiu, C.E.B. Waizenegger, K. Ocean, M. Mokhtari, C.P. Berlinguette, A self-driving laboratory advances the Pareto front for material properties, Nat Commun. 13 (2022) 995. https://doi.org/10.1038/s41467-022-28580-6.
[34] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passo, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, Scikit-learn: Machine Learning in Python, Journal of Machine Learning Research. 12 (2011) 2825–2830.
[35] T. Chen, C. Guestrin, XGBoost: A Scalable Tree Boosting System, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, San Francisco California USA, 2016: pp. 785–794. https://doi.org/10.1145/2939672.2939785.
[36] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, T.-Y. Liu, LightGBM: A Highly Efficient Gradient Boosting Decision Tree, in: Advances in Neural Information Processing Systems, Curran Associates, Inc., 2017. https://proceedings.neurips.cc/paper_files/paper/2017/hash/6449f44a102fde848669bdd9eb6b76fa-Abstract.html (accessed April 6, 2023).
[37] F. Häse, M. Aldeghi, R.J. Hickman, L.M. Roch, M. Christensen, E. Liles, J.E. Hein, A. Aspuru-Guzik, Olympus: a benchmarking framework for noisy optimization and experiment planning, Mach. Learn.: Sci. Technol. 2 (2021) 035021. https://doi.org/10.1088/2632-2153/abedc8.
[38] R. Hickman, P. Parakh, A. Cheng, Q. Ai, J. Schrier, M. Aldeghi, A. Aspuru-Guzik, Olympus, enhanced: benchmarking mixed-parameter and multi-objective optimization in chemistry and materials science, (2023). https://doi.org/10.26434/chemrxiv-2023-74w8d.
[39] H. Raina, S. Kaur, A.B. Jindal, Development of efavirenz loaded solid lipid nanoparticles: Risk assessment, quality-by-design (QbD) based optimisation and physicochemical characterisation, Journal of Drug Delivery Science and Technology. 39 (2017) 180–191. https://doi.org/10.1016/j.jddst.2017.02.013.
[40] L.Y. Kok, P. Bannigan, F. Sanaee, J.C. Evans, M. Dunne, M. Regenold, L. Ahmed, D. Dubins, C. Allen, Development and pharmacokinetic evaluation of a self-nanoemulsifying drug delivery system for the oral delivery of cannabidiol, Eur J Pharm Sci. 168 (2022) 106058. https://doi.org/10.1016/j.ejps.2021.106058.
[41] R. Paliwal, S. Rai, B. Vaidya, K. Khatri, A.K. Goyal, N. Mishra, A. Mehta, S.P. Vyas, Effect of lipid core material on characteristics of solid lipid nanoparticles designed for oral lymphatic delivery, Nanomedicine: Nanotechnology, Biology and Medicine. 5 (2009) 184–191. https://doi.org/10.1016/j.nano.2008.08.003.

[42] D. Liu, Z. Liu, L. Wang, C. Zhang, N. Zhang, Nanostructured lipid carriers as novel carrier for parenteral delivery of docetaxel, Colloids and Surfaces B: Biointerfaces. 85 (2011) 262–269. https://doi.org/10.1016/j.colsurfb.2011.02.038.
[43] S. Kumar, J.K. Randhawa, Solid lipid nanoparticles of stearic acid for the drug delivery of paliperidone, RSC Adv. 5 (2015) 68743–68750. https://doi.org/10.1039/C5RA10642G.
[44] F.-Q. Hu, S.-P. Jiang, Y.-Z. Du, H. Yuan, Y.-Q. Ye, S. Zeng, Preparation and characteristics of monostearin nanostructured lipid carriers, International Journal of Pharmaceutics. 314 (2006) 83–89. https://doi.org/10.1016/j.ijpharm.2006.01.040.
[45] H. Yuan, J. Miao, Y.-Z. Du, J. You, F.-Q. Hu, S. Zeng, Cellular uptake of solid lipid nanoparticles and cytotoxicity of encapsulated paclitaxel in A549 cancer cells, International Journal of Pharmaceutics. 348 (2008) 137–145. https://doi.org/10.1016/j.ijpharm.2007.07.012.
[46] M. Shah, K. Pathak, Development and Statistical Optimization of Solid Lipid Nanoparticles of Simvastatin by Using 23 Full-Factorial Design, AAPS PharmSciTech. 11 (2010) 489–496. https://doi.org/10.1208/s12249-010-9414-z.
[47] S. Das, W.K. Ng, R.B.H. Tan, Are nanostructured lipid carriers (NLCs) better than solid lipid nanoparticles (SLNs): Development, characterizations and comparative evaluations of clotrimazole-loaded SLNs and NLCs?, European Journal of Pharmaceutical Sciences. 47 (2012) 139–151. https://doi.org/10.1016/j.ejps.2012.05.010.
[48] C. Puglia, P. Blasi, L. Rizza, A. Schoubben, F. Bonina, C. Rossi, M. Ricci, Lipid nanoparticles for prolonged topical delivery: An in vitro and in vivo investigation, International Journal of Pharmaceutics. 357 (2008) 295–304. https://doi.org/10.1016/j.ijpharm.2008.01.045.
[49] K. Teskač, J. Kristl, The evidence for solid lipid nanoparticles mediated cell uptake of resveratrol, International Journal of Pharmaceutics. 390 (2010) 61–69. https://doi.org/10.1016/j.ijpharm.2009.10.011.
[50] W.A. Kasongo, J. Pardeike, R.H. Müller, R.B. Walker, Selection and Characterization of Suitable Lipid Excipients for use in the Manufacture of Didanosine-Loaded Solid Lipid Nanoparticles and Nanostructured Lipid Carriers, Journal of Pharmaceutical Sciences. 100 (2011) 5185–5196. https://doi.org/10.1002/jps.22711.
[51] X. Li, S. Nie, J. Kong, N. Li, C. Ju, W. Pan, A controlled-release ocular delivery system for ibuprofen based on nanostructured lipid carriers, International Journal of Pharmaceutics. 363 (2008) 177–182. https://doi.org/10.1016/j.ijpharm.2008.07.017.
[52] N. Schöler, C. Olbrich, K. Tabatt, R.H. Müller, H. Hahn, O. Liesenfeld, Surfactant, but not the size of solid lipid nanoparticles (SLN) influences viability and cytokine production of macrophages, International Journal of Pharmaceutics. 221 (2001) 57–67. https://doi.org/10.1016/S0378-5173(01)00660-3.
[53] K. Vivek, H. Reddy, R.S.R. Murthy, Investigations of the effect of the lipid matrix on drug entrapment, in vitro release, and physical stability of olanzapine-loaded solid lipid nanoparticles, AAPS PharmSciTech. 8 (2007) 16–24. https://doi.org/10.1208/pt0804083.
[54] E. Samara, M. Bialer, Pharmacokinetics of the dimethylheptyl homolog of cannabidiol in dogs., Drug Metab Dispos. 16 (1988) 875–879.
[55] S. Pandey, F. Shaikh, A. Gupta, P. Tripathi, J.S. Yadav, A Recent Update: Solid Lipid Nanoparticles for Effective Drug Delivery, Adv Pharm Bull. 12 (2022) 17–33. https://doi.org/10.34172/apb.2022.007.
[56] M. Pinheiro, R. Ribeiro, A. Vieira, F. Andrade, S. Reis, Design of a nanostructured lipid carrier intended to improve the treatment of tuberculosis, DDDT. 10 (2016) 2467–2475. https://doi.org/10.2147/DDDT.S104395.
[57] P. Bannigan, M. Aldeghi, Z. Bao, F. Häse, A. Aspuru-Guzik, C. Allen, Machine learning directed drug formulation development, Advanced Drug Delivery Reviews. 175 (2021) 113806. https://doi.org/10.1016/j.addr.2021.05.016.

[58] P. Bannigan, Z. Bao, R.J. Hickman, M. Aldeghi, F. Häse, A. Aspuru-Guzik, C. Allen, Machine learning models to accelerate the design of polymeric long-acting injectables, Nat Commun. 14 (2023) 35. https://doi.org/10.1038/s41467-022-35343-w.
[59] M. Elbadawi, L.E. McCoubrey, F.K.H. Gavins, J.J. Ong, A. Goyanes, S. Gaisford, A.W. Basit, Harnessing artificial intelligence for the next generation of 3D printed medicines, Advanced Drug Delivery Reviews. 175 (2021) 113805. https://doi.org/10.1016/j.addr.2021.05.015.
[60] S. N Politis, P. Colombo, G. Colombo, D. M Rekkas, Design of experiments (DoE) in pharmaceutical development, Drug Dev Ind Pharm. 43 (2017) 889–901. https://doi.org/10.1080/03639045.2017.1291672.
[61] Z. Ghaemmaghamian, R. Zarghami, G. Walker, E. O'Reilly, A. Ziaee, Stabilizing vaccines via drying: Quality by design considerations, Advanced Drug Delivery Reviews. 187 (2022) 114313. https://doi.org/10.1016/j.addr.2022.114313.
[62] J.R. Wagner, E.M. Mount, H.F. Giles, 25 - Design of Experiments, in: J.R. Wagner, E.M. Mount, H.F. Giles (Eds.), Extrusion (Second Edition), William Andrew Publishing, Oxford, 2014: pp. 291–308. https://doi.org/10.1016/B978-1-4377-3481-2.00025-9.
[63] M. Ramzan, S. Gourion-Arsiquaud, A. Hussain, J.S. Gulati, Q. Zhang, S. Trehan, V. Puri, B. Michniak-Kohn, I.P. Kaur, In vitro release, ex vivo penetration, and in vivo dermatokinetics of ketoconazole-loaded solid lipid nanoparticles for topical delivery, Drug Deliv. and Transl. Res. 12 (2022) 1659–1683. https://doi.org/10.1007/s13346-021-01058-6.
[64] W. Zeng, L. Guo, S. Xu, J. Chen, J. Zhou, High-Throughput Screening Technology in Industrial Biotechnology, Trends in Biotechnology. 38 (2020) 888–906. https://doi.org/10.1016/j.tibtech.2020.01.001.
[65] E. M. Payne, D. A. Holland-Moritz, S. Sun, R. T. Kennedy, High-throughput screening by droplet microfluidics: perspective into key challenges and future prospects, Lab on a Chip. 20 (2020) 2247–2262. https://doi.org/10.1039/D0LC00347F.
[66] N. Phan, J.J. Hong, B. Tofig, M. Mapua, D. Elashoff, N.A. Moatamed, J. Huang, S. Memarzadeh, R. Damoiseaux, A. Soragni, A simple high-throughput approach identifies actionable drug sensitivities in patient-derived tumor organoids, Commun Biol. 2 (2019) 1–11. https://doi.org/10.1038/s42003-019-0305-x.
[67] L. Cibulski, H. Mitterhofer, T. May, J. Kohlhammer, PAVED: Pareto Front Visualization for Engineering Design, Computer Graphics Forum. 39 (2020) 405–416. https://doi.org/10.1111/cgf.13990.
[68] M. Elarbi, S. Bechikh, C.A. Coello Coello, M. Makhlouf, L. Ben Said, Approximating Complex Pareto Fronts With Predefined Normal-Boundary Intersection Directions, IEEE Transactions on Evolutionary Computation. 24 (2020) 809–823. https://doi.org/10.1109/TEVC.2019.2958921.
[69] S. Petchrompo, D.W. Coit, A. Brintrup, A. Wannakrairot, A.K. Parlikad, A review of Pareto pruning methods for multi-objective optimization, Computers & Industrial Engineering. 167 (2022) 108022. https://doi.org/10.1016/j.cie.2022.108022.
[70] Z. Vinarov, B. Abrahamsson, P. Artursson, H. Batchelor, P. Berben, A. Bernkop-Schnürch, J. Butler, J. Ceulemans, N. Davies, D. Dupont, G.E. Flaten, N. Fotaki, B.T. Griffin, V. Jannin, J. Keemink, F. Kesisoglou, M. Koziolek, M. Kuentz, A. Mackie, A.J. Meléndez-Martínez, M. McAllister, A. Müllertz, C.M. O'Driscoll, N. Parrott, J. Paszkowska, P. Pavek, C.J.H. Porter, C. Reppas, C. Stillhart, K. Sugano, E. Toader, K. Valentová, M. Vertzoni, S.N. De Wildt, C.G. Wilson, P. Augustijns, Current challenges and future perspectives in oral absorption research: An opinion of the UNGAP network, Advanced Drug Delivery Reviews. 171 (2021) 289–331. https://doi.org/10.1016/j.addr.2021.02.001.
[71] M. Sohail Arshad, S. Zafar, B. Yousef, Y. Alyassin, R. Ali, A. AlAsiri, M.-W. Chang, Z. Ahmad, A. Ali Elkordy, A. Faheem, K. Pitt, A review of emerging technologies enabling improved solid oral dosage

form manufacturing and processing, Advanced Drug Delivery Reviews. 178 (2021) 113840. https://doi.org/10.1016/j.addr.2021.113840.
[72] S.A. Millar, R.F. Maguire, A.S. Yates, S.E. O'Sullivan, Towards Better Delivery of Cannabidiol (CBD), Pharmaceuticals (Basel). 13 (2020) 219. https://doi.org/10.3390/ph13090219.
[73] M. Bibi, F. ud Din, Y. Anwar, N.A. Alkenani, A.T. Zari, M. Mukhtiar, I.M. Abu Zeid, E.H. Althubaiti, H. Nazish, A. Zeb, I. Ullah, G.M. Khan, H.-G. Choi, Cilostazol-loaded solid lipid nanoparticles: Bioavailability and safety evaluation in an animal model, Journal of Drug Delivery Science and Technology. 74 (2022) 103581. https://doi.org/10.1016/j.jddst.2022.103581.
[74] C. Ban, M. Jo, Y.H. Park, J.H. Kim, J.Y. Han, K.W. Lee, D.-H. Kweon, Y.J. Choi, Enhancing the oral bioavailability of curcumin using solid lipid nanoparticles, Food Chemistry. 302 (2020) 125328. https://doi.org/10.1016/j.foodchem.2019.125328.
[75] T. Gupta, J. Singh, S. Kaur, S. Sandhu, G. Singh, I.P. Kaur, Enhancing Bioavailability and Stability of Curcumin Using Solid Lipid Nanoparticles (CLEN): A Covenant for Its Effectiveness, Frontiers in Bioengineering and Biotechnology. 8 (2020). https://www.frontiersin.org/articles/10.3389/fbioe.2020.00879 (accessed July 28, 2023).
[76] V.-A. Duong, T.-T.-L. Nguyen, H.-J. Maeng, Preparation of Solid Lipid Nanoparticles and Nanostructured Lipid Carriers for Drug Delivery and the Effects of Preparation Parameters of Solvent Injection Method, Molecules. 25 (2020) E4781. https://doi.org/10.3390/molecules25204781.

Acknowledgments

Figures were created with BioRender.com and Prism.

Funding

NSERC Discovery grant (RGPIN-2022-04910) to C.A. A.A.G. acknowledges support from the Defense Advanced Research Projects Agency under the Accelerated Molecular Discovery Program under Cooperative Agreement No. HR00111920027 dated August 1, 2019. Z.B. acknowledges a scholarship from the Centre for Pharmaceutical Oncology, Leslie Dan Faculty of Pharmacy, University of Toronto. R.J.H. acknowledges NSERC for a postgraduate scholarship (PGSD3-534584-2019), as well as support from the Vector Institute. A.A.G. would like to thank Dr. Anders Frøseth for his support. This research relates to the Acceleration Consortium at the University of Toronto, which receives funding from the Canada First Research Excellence Fund (CFREF).

Author information

Authors and affiliations

Leslie Dan Faculty of Pharmacy, University of Toronto, Toronto, ON M5S 3M2, Canada
Zeqing Bao, Fion Yung, Pauric Bannigan, and Christine Allen

Department of Chemistry, University of Toronto, Toronto, ON M5S 3H6, Canada
Riley J. Hickman and Alán Aspuru-Guzik

Department of Computer Science, University of Toronto, Toronto, ON M5S 2E4, Canada
Riley J. Hickman and Alán Aspuru-Guzik

Vector Institute for Artificial Intelligence, Toronto, ON M5S 1M1, Canada
Riley J. Hickman and Alán Aspuru-Guzik

Lebovic Fellow, Canadian Institute for Advanced Research (CIFAR), Toronto, ON M5S 1M1, Canada
Alán Aspuru-Guzik

Department of Chemical Engineering & Applied Chemistry, University of Toronto, Toronto, ON M5S 3E5, Canada
Alán Aspuru-Guzik and Christine Allen

Department of Materials Science & Engineering, University of Toronto, Toronto, ON M5S 3E4, Canada
Alán Aspuru-Guzik

CIFAR Artificial Intelligence Research Chair, Vector Institute, Toronto, ON M5S 1M1, Canada
Alán Aspuru-Guzik

Acceleration Consortium, Toronto, ON M5S 3H6, Canada
Alán Aspuru-Guzik and Christine Allen

Author contributions

Zeqing Bao: Investigation, Visualization, Writing – original draft. **Fion Yung**: Investigation, Visualization, Writing – review & editing. **Riley J. Hickman**: Investigation, Writing – review & editing. **Alán Aspuru-Guzik**: Project administration, Writing – review & editing, Supervision, Funding acquisition. **Pauric Bannigan**: Conceptualization, Project administration, Writing – review & editing, Supervision. **Christine Allen**: Conceptualization, Project administration, Writing – review & editing, Supervision, Funding acquisition.

Corresponding author

Correspondence to Pauric Bannigan (pauric.bannigan@utoronto.ca) and Christine Allen (cj.allen@utoronto.ca)

Ethics declarations

Ethics approval

Animal studies were approved by the Animal Care Committee at the University of Toronto.

Consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

A.A.-G is a founding member of Kebotix, Inc. Z.B., R.J.H, P.B., C.A. and A.A.-G. are founding members of a new company, 15073383 Canada Inc.