Asian Journal of Pharmaceutical Sciences 18 (2023) 100811

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/AJPS

Original Research Paper

Predicting liposome formulations by the integrated machine learning and molecular modeling approaches

Run Han[a,1], *Zhuyifan Ye*[a,1], *Yunsen Zhang*[a,1], *Yaxin Cheng*[a], *Ying Zheng*[a,b,*], *Defang Ouyang*[a,b,*]

[a] *State Key Laboratory of Quality Research in Chinese Medicine, Institute of Chinese Medical Sciences (ICMS), University of Macau, Macao 999078, China*

[b] *Faculty of Health Sciences, University of Macau, Macao 999078, China*

ARTICLE INFO

Article history:
Received 13 September 2022
Revised 20 January 2023
Accepted 22 March 2023
Available online 16 April 2023

Keywords:
Liposome
Formulation prediction
Machine learning
Molecular modeling

ABSTRACT

Liposome is one of the most widely used carriers for drug delivery because of the great biocompatibility and biodegradability. Due to the complex formulation components and preparation process, formulation screening mostly relies on trial-and-error process with low efficiency. Here liposome formulation prediction models have been built by machine learning (ML) approaches. The important parameters of liposomes, including size, polydispersity index (PDI), zeta potential and encapsulation, are predicted individually by optimal ML algorithm, while the formulation features are also ranked to provide important guidance for formulation design. The analysis of key parameter reveals that drug molecules with logS [-3, -6], molecular complexity [500, 1000] and XLogP3 (≥2) are priority for preparing liposome with higher encapsulation. In addition, naproxen (NAP) and palmatine HCl (PAL) represented the insoluble and water-soluble molecules are prepared as liposome formulations to validate prediction ability. The consistency between predicted and experimental value verifies the satisfied accuracy of ML models. As the drug properties are critical for liposome particles, the molecular interactions and dynamics of NAP and PAL liposome are further investigated by coarse-grained molecular dynamics simulations. The modeling structure reveals that NAP molecules could distribute into lipid layer, while most PAL molecules aggregate in the inner aqueous phase of liposome. The completely different physical state of NAP and PAL confirms the importance of drug properties for liposome formulations. In summary, the general prediction models are built to predict liposome

* Corresponding authors.
E-mail addresses: yzheng@umac.mo (Y. Zheng), defangouyang@umac.mo (D. Ouyang).

[1] These authors equally contributed to the manuscript.

Peer review under responsibility of Shenyang Pharmaceutical University.

https://doi.org/10.1016/j.ajps.2023.100811

formulations, and the impacts of key factors are analyzed by combing ML with molecular modeling. The availability and rationality of these intelligent prediction systems have been proved in this study, which could be applied for liposome formulation development in the future.

1. Introduction

Liposome is a bilayer vesicle-liked particle composed by various lipids with the diameter ranging from dozen nanometer to micrometer, which is firstly described by Alec Bangham in 1964 and applied as carrier for biological active ingredient delivery by Gregoriadis in 1971 [1,2]. After that, liposome has developed as one of the most widely applied drug delivery systems. Doxil® is the first approved liposome product by the Food and Drug Administration (FDA) in 1995, that realize the process from concept to clinic application [3]. Within several decades, there are dozens of liposome-based products have been come into clinic trials and approved as product, which demonstrates the great application potential of liposomes [4]. Generally, liposome is composed by various lipids with hydrophilic head and non-polar chain that form as membrane-liked bilayer particles [5,6]. Due to the great biocompatibility and biodegradability, liposome has been widely applied for drug delivery, not only small molecules, but also biomolecules. In laboratory research and industrial production, liposomes are mainly prepared by active and passive loading methods, especially passive loading methods including anti-solvent and thin film hydration methods, are widely applied for liposome preparation as the basic techniques [7]. Sonication and extrusion are account for particle downsizing to obtain homogenized liposomes with desired size [8]. There were some key characterization parameters in liposome formulations study that need to be controlled and evaluated, including size, polydispersity index (PDI), zeta potential and encapsulation [9]. Due to the complex formulation components and preparation process, the high-dimensional formulation space of liposomes could reach to 10^{25} – 10^{30} [10]. The formulation development of liposomes faces huge difficulty by the simple trial-and-error experiments that should be improved by more effective ways.

In recent years, there are some computational methods applied for liposome formulation prediction. Design of experiments method (DoE) is a regression model that calculates response factors for each parameter. It has analyzed the impact of physical properties for peptide-coated liposome and built a rapid screen manner for liposomes with microfluidics preparation [11,12]. Langmuir balance study is another mathematic method used for liposome analysis that has revealed the relationship between media ion concentration and drug encapsulation of unilamellar liposomes [13]. Machine learning (ML) is one of the most exciting techniques for pharmaceutical formulation study that could make predictions based on existed data with explicit program [14–17]. There are some trials applying ML for liposome formulation optimization or prediction. Subramanian N et al. have built artificial neural network (ANN) model based on 3^3 factorial variables formulation data to optimize the percentage of cytarabine entrapped in the liposome [18]. Size and PDI of liposomes that prepared by turbulent jet are also predicted by ANN to find the critical technological quality parameters [19]. In addition, Zucker D and co-workers have built a decision tree model on laboratory liposome data that prepared by active loading method to analyze the impact of drug properties and loading parameters for encapsulation [20]. Nevertheless, previous models are based on small amount of data or specific preparation method, which limit their prediction ability and application range. Predicting liposome formulation is still a challenge at present. It is worthy to develop ML models to predict liposome formulations and improve study efficiency.

To study the impacts of key parameters for liposome particles, experimental techniques could not directly observe the microstructure of liposomes, other computational approaches will be used for liposome simulation. Molecular modeling is a powerful technique that could see the structure of liposome at molecular level. In the previous study, the bilayer structure of monopalmitoyl glycerol, cholesterol and dicetyl phosphate composed vesicle has been simulated and the molecular interactions are investigated by molecular dynamic simulations [21]. Hypericin liposome is built by coarse-grained (CG) model, the distribution of drug in vesicles and free energy between two molecules are calculated to analyze the liposome structure quantitatively and qualitatively in a large scale [22]. Tetrandrine, 5-fluorouracil, ligustrazine and osthole are coated with different lipids and constructed by CG model to analyze particle stability and encapsulation [23–25]. So, molecular modeling is suitable for visualizing particle structure and molecular interaction, which provides some novel interpretations for liposome properties.

Current study aims to build the general prediction models for liposomes with passive loading methods by ML approaches. The critical quality attributions (CQA) of liposomes are analyzed by feature importance ranking. Two drugs with different properties, water-insoluble naproxen (NAP) and water-soluble palmatine HCl (PAL), are prepared as a series of liposome formulations to validate the prediction model. To further analyze the influence of different drug properties, liposome structures are investigated by molecular dynamic simulations. The drug distribution and molecular interactions in liposome particle will be revealed.

2. Materials and methods

2.1. Materials

Naproxen (NAP) was obtained from Aladdin (Shanghai, China). Palmatine HCl (PAL) was purchased from Hubei Yuancheng Pharmaceutical Co., Ltd. (Hubei, China). Lecithin, cholesterol, distearoylphosphatidylcholine (DSPC), 1,2-distearoyl-sn–glycero-3-phosphoethanolamine-N-[amino(polyethylene glycol)−2000 (DSPE-PEG2000), 1,2-Dioleoyl-sn–glycero-3-phosphoethanolamine (DOPE) and stearylamine was purchased from Aladdin (Shanghai, China). All other chemical reagents were of analytical or chromatographic grade.

2.2. Building the prediction model

2.2.1. Data collection

The liposome formulation dataset was constituted by published data that collected from Web of Science, Scopus database and internal experimental data. Due to the diverse modification strategies, it was hard to fully cover all kinds of liposomes. Considering thin film hydration and anti-solvent methods were the commonly used and easy-to-operate ways for liposome preparation whether in laboratory or industrial production, the current study focused on the conventional liposomes that prepared by thin film hydration and anti-solvent methods. Although there were various specific operation steps reported for liposome preparation, they were generally grouped into these two categories. The standards of data collection were listed as follows:

(1) The conventional liposomes were defined by components in this study. The novel synthetic lipids, ligands and biological modification in liposomes were not contained in this dataset. All liposome formulations in dataset were composed by 114 drugs and 43 common lipids and stabilizer;
(2) Thin film hydration method: all phospholipids were dissolved in appropriate organic solvent. Removing solvent by rotary evaporator and vacuum drying to form thin films. Then aqueous solutions were added to completely suspend lipids.
(3) Anti-solvent method: briefly, the phospholipids in organic phase were added into large volume of aqueous phase to obtain emulsion with following organic solvent removal.

Finally, 665 formulation data with every component and its percentage were collected as the formulation dataset, including four sub-datasets: 624 data for size, 195 data for PDI, 236 data for zeta potential and 532 data for encapsulation prediction. Moreover, the preparation parameters were also contained, that were preparation time, temperature, the kinds of solvent and whether using sonication or filtration. So, the liposome formulation data included APIs, lipids, solvents, key process indicators and the characteristics of liposome.

Firstly, each drug, lipid and solvent were encoded as descriptors that searched from chemical information database and ALOGPS online prediction system. They contained molecular weight, XLogP3, hydrogen bond donor count, hydrogen bond acceptor count, rotatable bond count, topological polar surface area, heavy atom count, and complexity. Melting point as an estimate of dissolution capacity was also used to describe APIs. While boiling point, vapor pressure, electric constant and polar index that related to the preparation requirements and the polarity for solvents were added for solvent describing. The drug/lipid ratio was used to describe the contents of APIs, while a maximum of 3 kinds of lipids were collected and the weight percent was added for lipids. The preparation methods of liposome were encoded to integers. The key process indicators included preparation time, sonication, filter diameter and preparation temperature. The encapsulation efficiency, PDI, particle size, and zeta potential of liposomes were the predicted targets.

2.2.2. Building prediction model

Before training models, each dataset was split into 3 subsets, including the training set (80%), validation set (10%) and test set (10%), using the MD-FIS data splitting algorithm [26]. This three-dataset splitting strategy was extensively adopted in ML, where the training set was for optimizing the model, the validation set was for finding the optimal hyperparameter configuration of model, and the test set was for testing the final generalization performance of the model on unseen data.

In recent years, the advances of ML applications have been driven by the growing computational power, the state-of-the-art algorithms, and the available data. Various ML models have been developed to fit different tasks [27,28]. Thus, it was necessary to choose an appropriate algorithm that was suitable for complex liposome prediction. Five machine learning algorithms, including LightGBM, random forest (RF), support vector machine (SVM), multiple linear regression (MLR), and partial least regression (PLS), were introduced to construct models for the four prediction tasks, consisting of size, PDI, zeta potential and encapsulation of liposome. Based on previous studies, LightGBM and RF algorithms were proved to outperform deep learning on tabular data [29]. LightGBM was a histogram-based decision tree algorithm that used leaf-wise with a depth limit decision tree growth strategy, which could directly support categorical features and ensure high efficiency while prevent overfitting. RF was another tree-based ensemble learning algorithm frequently with superior performance on tabular data for practical applications, and it was well known to be built on a parallel Bagging framework [30]. The RF training was highly parallelized with simultaneous sampling of both sample and feature dimensions, and such randomness ensured strong generalization of the final model. Then the hyperparameters configurations of each algorithm were set before training model that could control the learning process and commonly determine the network architecture to influence accuracies. A random search method was carried out to search the best hyperparameters of each model with higher efficiency rather than grid and manual search [31]. One thousand hyperparameter configurations in the hyperparameter space were searched for each task, and the details of the final settings were listed in Table 1. The hyperparameters for LightGBM were learning rate, the number of trees, the subsample ratio, the subsample ratio of columns, maximum leaves for base learners, and minimum sample in a leaf.

Table 1 – Hyperparameters for size, PDI, zeta potential, and encapsulation models.

	Size	PDI	Zeta potential	Encapsulation
LightGBM	0.01298; 845; 0.9698; 0.6142; 78; 14;	0.003246; 712; 0.9680; 0.9368; 35; 8;	0.01483; 577; 0.2319; 0.8950; 70; 18;	0.002469; 1180; 0.3435;0.6465; 86; 15;
RF	34; 431; 42; 2; 1;	53; 137; 42; 5; 2;	46; 353; 8; 6; 2;	45; 1047; 15; 9; 4;
SVM	radial basis function; 317.18; 0.04588;	radial basis function; 146.7; 0.001109;	radial basis function; 439.4; 0.00409;	radial basis function; 84.4; 0.01022;
PLS	5;	17;	28;	28;
MLR	\	\	\	\

";" separates different hyperparameters.
"," means that the hyperparameter is composed of more than one element.
"\" denotes no hyperparameter.

For RF algorithm, the hyperparameter included maximum features, the number of trees, the maximum depth, the minimum samples for split, and the minimum samples in child leaf. In SVM algorithm, the hyperparameters were the kernel function, the penalty parameter C and γ. And the hyperparameter for PLS was the number of components. The LightGBM models were built by the LightGBM package (3.2.1), the RF, SVM, PLS, and MLR models were established by the Scikit-Learn package (0.24.2) in Python.

After building models, mean absolute error (MAE) between true value and predicted value, root mean error (RMSE) and coefficient of determination (R^2) were calculated to evaluate the accuracy of prediction model, which was defined as follow:

$$MAE = \frac{\sum_{i=1}^{n} |\widehat{y_i} - y_i|}{n}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (\widehat{y_i} - y_i)^2}$$

where n is the number of data, y_i is the i_{th} real label, and $\widehat{y_i}$ is the i_{th} prediction.

2.3. *Experimental validation of liposome prediction model*

To validate the prediction ability of LightGBM models on new data, some liposome formulations were prepared and input into models as unknown data. NAP and PAL with different physicochemical properties were acted as model drugs to prepare liposomes by thin film hydration and anti-solvent methods. NAP was the nonsteroidal anti-inflammatory drug that was almost insoluble (15.9 μg/ml). PAL was an active ingredient of Chinese medicine with slight solubility in aqueous phase. Due to the positively charged nitrogen atom, PAL preferred to interact with negative charged molecules and exist as salt that improved solubility dramatically. These two drugs were encapsulated into liposomes with the variables of three factors, that were different lipid composition, drug/lipid ratio and preparation method. The components of NAP- and PAL-liposomes were listed in Table S1-S2 (in supplementary material). In addition, the size, PDI, zeta potential and encapsulation were measured by experiment and predicted by models.

For thin film hydration method, the drug and lipids were weighted as formulation form listed, and then dissolved in 4 ml ethanol in a round-bottomed flask. The organic solvent was removed under reduced pressure using rotary evaporator at the temperature that above the transition temperature (Tc) of the lipids to form as thin film on the wall of flask. The film was put in vacuum drying at room temperature overnight to complete evaporate solvent. Then 4 ml PBS buffer was added in flask to hydrate lipid film until the homogenous solution was formed. Finally, the solution would be filtered through serious diameter membrane to obtain uniform liposome particles. Meanwhile, anti-solvent method was also applied for liposome preparation. Specifically, each lipid and drug were weighted precisely and co-dissolved in 0.5 ml ethanol. Then the ethanol solution was injected into 4 ml PBS 6.8 buffer at the stirring speed of 1,200 rpm. After stirring with 3 min, liposomes were prepared.

The size, PDI and zeta potential of liposome particles were determined by Zetasizer system (Nano–ZS, Malvern instruments, UK) with triplicate measurements. For the encapsulation analysis, it was necessary to separate free drug and encapsulated drug in liposome particles. The prepared liposome solution was transferred into ultrafiltration tube and centrifuged at 4,000 rpm for 30 min to remove free drug. Methanol was added into liposome solutions to completely destroy particle structure and release encapsulated drugs. Then drug content in liposome was determined by high performance liquid chromatography (HPLC) and encapsulation was calculated as below equation:

$$\text{Encapsulation (\%)} = \frac{\text{weight of drug in liposome}}{\text{weight of total drug}} \times 100\%$$

The experimental values of size, PDI, zeta potential and encapsulation were compared with predicted value to validate the prediction accuracy of LightGBM models.

2.4. *Molecular dynamic simulation*

2.4.1. *Coarse-grained model setup*

Martini, a general forcefield of CG model, was employed to simulate system motion of NAP and PAL–encapsulated liposomes [32]. The CG topology parameters of lipid and cholesterol were obtained from the Martini website (http://cgmartini.nl/index.php/force–field–parameters/lipids). Whereas NAP and PAL molecules must be manually parameterized through the mapping principle by grouping

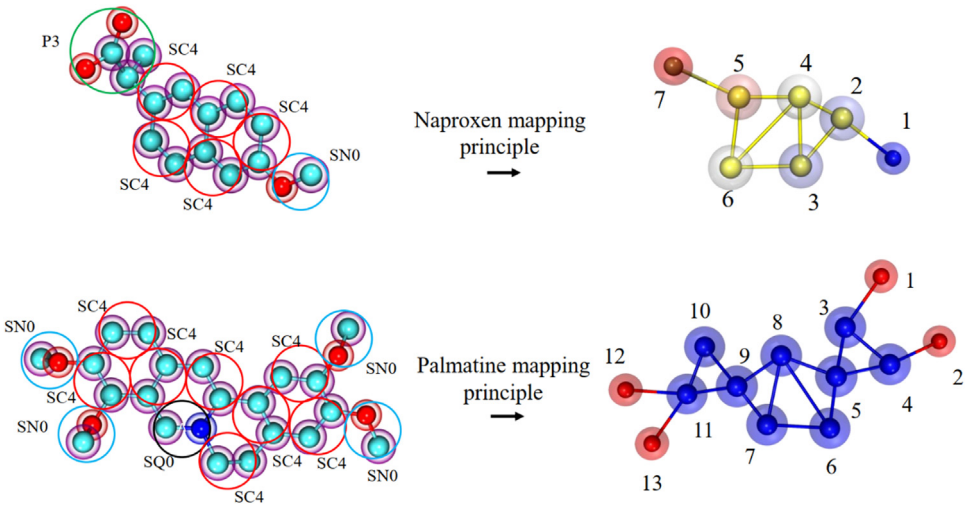

Fig. 1 – Mapping of coarse-grained beads to the NAP and PAL molecules. SN0, SC4, P3 and SQ0 indicate the functional groups of methoxyethane, 2-butyne, choline and acetic acid, respectively.

three and four atoms to one CG bead as Fig. 1 showed. Reference data of all−atom (AA) was fitting with CG model in the bond length, bond angle and dihedral [33]. Four types of CG bead including SN0, SC4, P3 and SQ0 were introduced to represent the structural properties of AA model. While bonds $4-6$ (NAP) and $6-8$ (PAL) were additionally introduced to maintain the ring plane of molecules. The relevant dihedrals analysis of CG and AA model were fitting well, which confirmed the rationality of these drug structures (in supplement Fig. S1). Next, CHARMM−GUI [34] was hired to initialize the CG liposome structures within the cubic box that the diameter was set as 15 nm, water molecules and ions were added to balance systems. More details of two liposome systems were listed in Table S3.

2.4.2. *Molecular dynamic simulation*

In the CG system, each assignment was executed in the isothermal−isobaric (NPT) ensemble with pressure 1.0 bar. Firstly, all molecules were coupled isotropically with 4.5×10^{-5} compressibility with the coupling constant of 12.0 ps under Parrinello−Rahman algorithm [35] and 1.0 ps under $v-$rescale algorithm [36] at the temperature of 298.15 K. The cutoff distance of electrostatic and van der Waals interaction was set to 1.1 nm, and the long−range electrostatic interaction was computed by reaction−field algorithm [37]. Finally, a timestep of 25 fs was applied to integrate Newton's motion equation by the leap−frog integrator. All systems were accelerated by GPU NVIDIA Tesla V100 and Intel(R) Xeon(R) Gold 6148 CPU @ 2.40 GHz using gromacs 2020 package. To better illuminate the interaction and dynamics equilibrium, the motion trajectory on 1 μs were visualized by VMD and Qtgrace (https://plasma-gate.weizmann.ac.il/Grace/).

To validate the reliability and atomistic reproducibility of CG models, dihedrals topological properties of NAP and PAL were computed and compared with atomistic data. Umbrella sampling (US) was a non-boltzmann sampling technique where adding an extra term to the energy for sampling that could explore free energy in collective variable space. It was used to simulate the dimerization process of two drugs in CG and AA model for calculating and comparing the potential of mean force (PMF), which determine whether CG model can reproduce the AA properties in short- and long-range interaction. The reaction coordinate of adjacent samples was defined as at least 28 windows with 0.5 Å spacing. The center of mass (COM) of drug molecule was fixed into the reaction coordinate through a harmonic potential with a force constant of 5,000 kJ· nm^2/mol. Weighted histogram analysis method (WHAM) was employed to describe the PMF with the gmx wham tool in gromacs. In this research, the radial distribution function (RDF) was used to determine the distribution state in the vesicle that was defined in the following way:

$$g_{AB}(r) = \frac{\langle \rho_B(r) \rangle}{\langle \rho_B \rangle_{local}} = \frac{1}{\langle \rho_B \rangle_{local}} \frac{1}{N_A} \sum_{i \in A}^{N_A} \sum_{j \in B}^{N_B} \frac{\delta\left(r_{ij} - r\right)}{4 \ \pi r^2}$$

Where $\langle \rho_B(r) \rangle$ was the particle density of type B at a distance r around particles A, and $\langle \rho_B \rangle_{local}$ was the particle density of type B averaged over all spheres around particles A with radius r_{max}. Type A in our study indicated the COM of vesicles or drugs. Type B was the remaining ingredients including beads NC3, PO4, C4B, CL^-, NA^+.

3. Results and discussion

3.1. *Data distribution*

Before building prediction models, some critical parameters of liposome formulations were statistically analyzed (Fig. 2). There were 114 kinds of drugs with different physicochemical properties and 43 lipids with various chain end groups recorded in dataset, which represented the commonly used ingredients in liposome products and research. All the drugs in liposome dataset were small molecules as the distribution

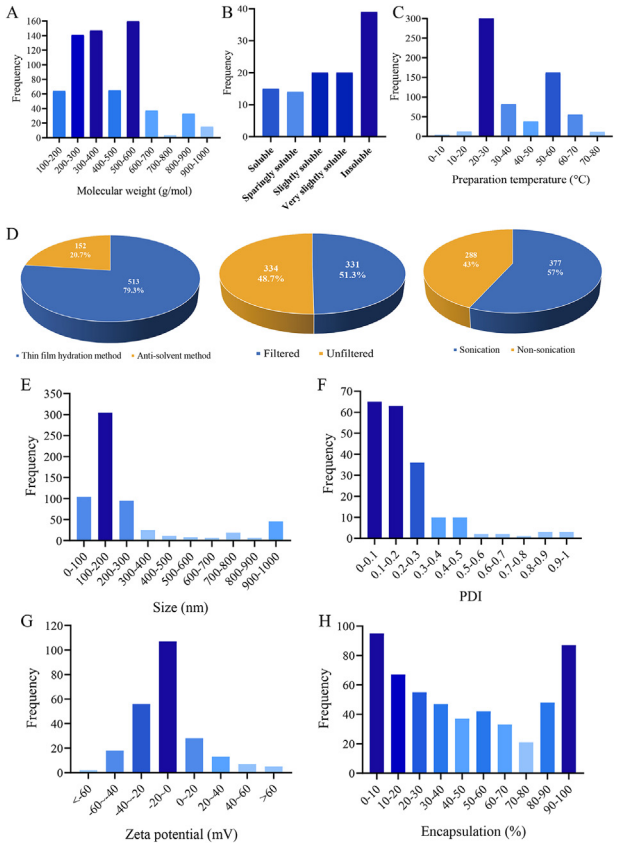

Fig. 2 – The distribution of input parameters (A. drug molecular weight; B. the solubility of drug molecules in dataset; C. preparation temperature; D. preparation method and filtration or sonication) and output parameters in liposome dataset (E. size; F. PDI; G. zeta potential; H. encapsulation).

of molecular weight was from one hundred to a thousand, and the majority was below 600. Referring to the definition of solubility in the pharmacopeia, the solubility of drugs was classified as Fig. 2B with different degrees. Most of drug molecules were insoluble, whereas 30 kinds of drugs were soluble or sparingly soluble. For preparation parameters, 80% liposomes were prepared by thin film hydration, and others (20%) were prepared by anti-solvent method. The temperature in the preparation process was centered on 20 to 60 ºC, which was related to phase transition temperature of lipids. In the post-preparation process, there were about 50% formulations downsized by filtration and 57% by sonication to obtain homogenous liposome particles in this dataset. Looking at the output parameters as prediction targets, particle size was distributed from tens of nanometers to micrometer and the majority were below 300 nm. PDI was less than 0.3 except for the very few large data, which indicated the homogeneity of particles within nanoscale. 80% of liposome was neutral or negative charge as zeta potential was centered in −40 to 20. For the encapsulation, it distributed uniformly from a few percent to over 90 percent, in general, the amount of data that below 50% was more than over 50%. Overall, the comprehensive liposome formulation data were collected including many drugs and lipids with different properties and two kinds of general preparation methods. All the formulation components, preparation techniques, and characterization parameters were recorded and input into LightGBM models to ensure adequate training, which improved the rationality and reliability of these models.

3.2. *The accuracy of lightgbm model*

In order to find the optimal model for liposome formulation prediction, some algorithms that based on boosting (LightGBM and RF), instance (SVM) and linear (MLR and PLS) methods were applied in this study. The MAE, RMSE and R^2 for each task were listed in Tables 2-5. LightGBM model showed the lowest MAE and highest R^2 for size and zeta potential prediction, while RF exhibited the better prediction performance for PDI and encapsulation prediction. The MAE of LightGBM model for size prediction on test set was 42.3 nm and R^2 was 0.83, even considering most size was centered within 300 nm, this absolute error was acceptable. For PDI prediction, MAE of RF model was 0.04 that was a very small error for data distributed between 0 and 1. But R^2 was relatively low with 0.6, which indicated the changes of formulation in liposomes were not strongly linearly correlated with PDI. The key parameters that had impacts for PDI were needed to be analyzed. The MAE of LightGBM model for zeta potential was 9.62 and R^2 was 0.76. The MAE was a bit large that could possibly predict negative potential into positive potential. It was attributed to the less and bias distributed data for zeta potential that LightGBM did not completely learn data features. For encapsulation prediction, MAE of RF model was 15.53 that was 15% absolution error compared with the data distribution in 1%–100% and R^2 was 0.63. The lower R^2 implied the relatively weak correlation between formulation parameters and encapsulation that would be further analyzed by feature importance. Considering the data distribution range for each task, the MAEs and R^2 of these prediction models were reasonable and acceptable. The true value and predicted value for training set, validation set and test set were labeled as scatter plots in Fig. 3. Most data points were distributed along the diagonal within ± 20% error area and just very small amount of data was not well consistent with true value for each prediction task, which indicated the good prediction accuracy of prediction models. LightGBM and RF were the high-level tree-based algorithm that could quickly find the best solutions for feature learning automatically to achieve the better prediction performance for liposome data.

3.3. *Feature importance of prediction models*

Except for the predicting functions, the feature importance of each input parameter could be ranked according to the total gain value of LightGBM and total reduction of RF. The molecular inner properties, drug/lipid ratio, and preparation parameters were comprehensively analyzed to evaluate their impact for size, PDI, zeta potential and encapsulation individually as Fig. 4 showed.

In the LightGBM model for size, the diameter of filtration membrane was the most important factor that ranked top 1.

Table 2 – Performance of various machine learning algorithms for the prediction of size.

Algorithms	Training set			Validation set			Test set		
	MAE	RMSE	R^2	MAE	RMSE	R^2	MAE	RMSE	R^2
LightGBM	**30.37**	**56.58**	**0.95**	**34.22**	**48.25**	**0.84**	**42.30**	**67.37**	**0.83**
RF	28.68	52.01	0.96	29.48	39.06	0.90	46.81	68.11	0.82
SVM	49.98	121.25	0.78	50.39	81.31	0.55	63.27	107.75	0.55
PLS	166.81	230.30	0.22	104.98	145.85	−0.46	156.91	210.38	−0.70
MLR	161.23	214.96	0.32	133.34	176.58	−1.14	179.70	225.05	−0.94

Table 3 – Performance of various machine learning algorithms for the prediction of PDI.

Algorithms	Training set			Validation set			Test set		
	MAE	RMSE	R^2	MAE	RMSE	R^2	MAE	RMSE	R^2
RF	**0.05**	**0.09**	**0.72**	**0.08**	**0.14**	**0.62**	**0.04**	**0.07**	**0.60**
LightGBM	0.06	0.10	0.66	0.08	0.13	0.65	0.04	0.08	0.45
PLS	0.10	0.15	0.31	0.12	0.18	0.36	0.08	0.01	0.34
SVM	0.09	0.13	0.50	0.11	0.16	0.47	0.08	0.10	0.22
MLR	0.10	0.14	0.34	0.13	0.18	0.33	0.11	0.18	−1.75

Table 4 – Performance of various machine learning algorithms for the prediction of zeta potential.

Algorithms	Training set			Validation set			Test set		
	MAE	RMSE	R^2	MAE	RMSE	R^2	MAE	RMSE	R^2
LightGBM	**5.15**	**7.61**	**0.88**	**8.64**	**12.67**	**0.84**	**9.62**	**15.66**	**0.76**
SVF	5.38	11.00	0.74	10.94	14.53	0.79	11.95	16.80	0.72
RF	6.22	9.02	0.83	11.83	16.16	0.74	9.69	17.02	0.71
PLS	10.93	14.46	0.55	13.23	15.94	0.74	13.36	19.12	0.64
MLR	8.22	10.84	0.75	11.10	13.49	0.82	11.42	20.63	0.58

Table 5 – Performance of various machine learning algorithms for the prediction of encapsulation.

Algorithms	Training set			Validation set			Test set		
	MAE	RMSE	R^2	MAE	RMSE	R^2	MAE	RMSE	R^2
RF	**9.36**	**12.84**	**0.85**	**11.88**	**19.22**	**0.64**	**15.53**	**21.22**	**0.63**
LightGBM	9.27	12.42	0.86	13.16	21.10	0.56	15.35	21.46	0.62
MLR	15.30	19.91	0.63	23.53	46.29	−1.12	17.93	24.29	0.52
PLS	15.81	20.52	0.61	16.37	20.62	0.58	17.96	24.70	0.50
SVM	9.73	16.23	0.76	14.96	20.92	0.57	20.68	27.42	0.38

It agreed with the experimental process that filtration was the most common and effective way to minimize particle size. Then drug/lipid ratio was in second place. It claimed that the weight proportion of lipid and drug had a great impact for particle size that could determine the particle structure. XLogP3 of drug, preparation time and rotatable bond count of drug were listed in 3rd to 5th place, which revealed the preparation process and drug properties like lipophilicity and chemical structures of drug molecules were the main influencing factors for liposome size. Unlike with the previous experience of ignoring the nature of the drug, physicochemical properties of drug were related to the interaction way with lipid layers that were emphasized as main influential factors by this model. XLogP3 represented the affinity of drug molecules for lipid bilayer and rotatable bond count reflected the molecular flexibility of drug. Generally, the higher flexibility was not benefit for the interaction between drug molecules with lipids. Complexity and molecular weight of drug that related to molecular structure were ranked in next two place. In the post-preparation, besides filtration, sonication was another effective way to minimize liposome size ranked in 8th. Other drug properties and preparation

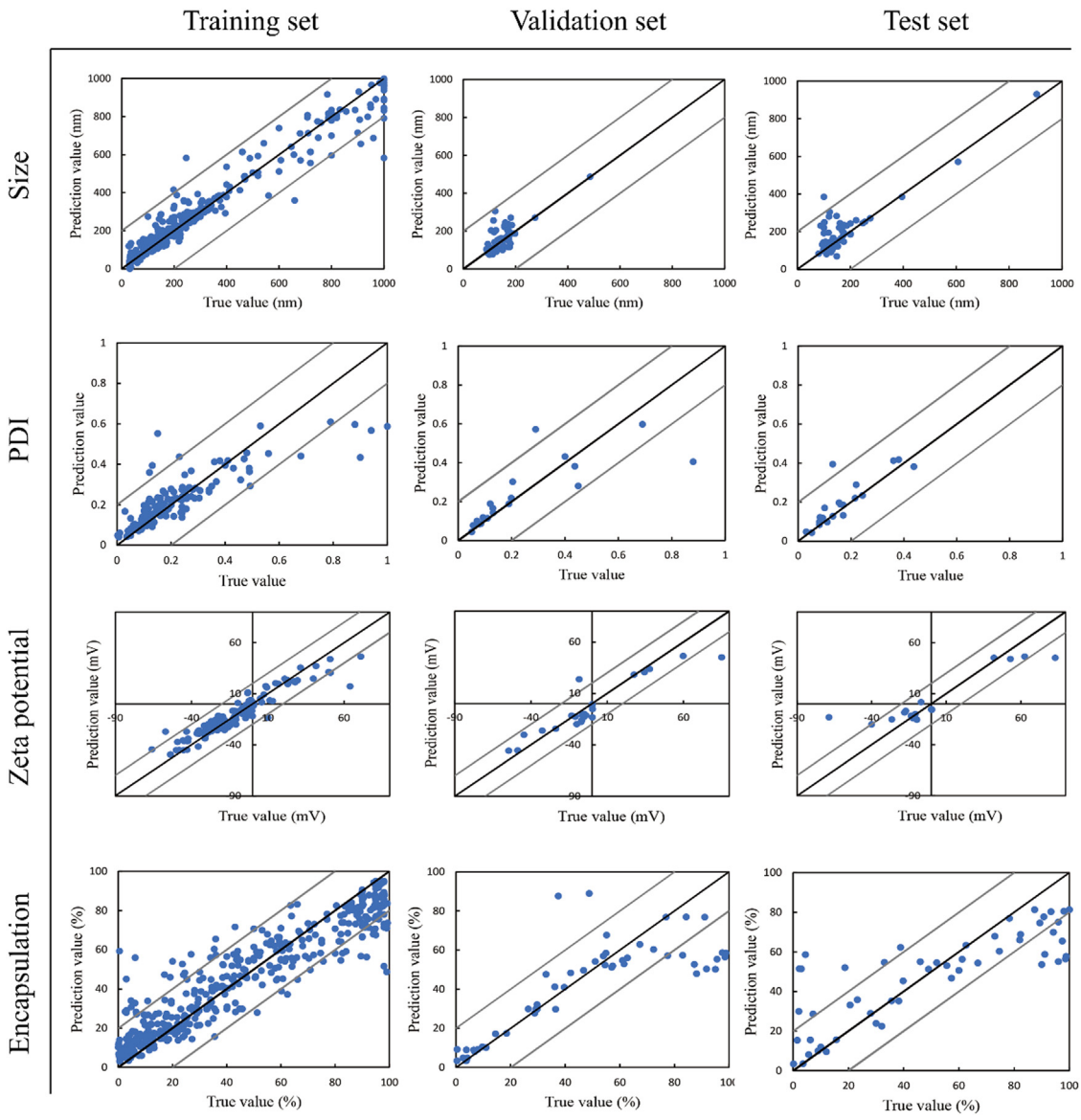

Fig. 3 – Scatter plots of true value and predicted value on training set, validation set and test set for size, encapsulation, PDI and zeta potential prediction.

time were also the contributing factors for liposome size with smaller gain value.

For PDI prediction by RF, preparation parameters were the major contributing factors, especially filtration diameter was listed in top1 with a large proportion. Then preparation time, XLogP3 of drug, sonication, and drug/lipid ratio were listed in 2nd to 5th with very small reduction value. It suggested that PDI was mostly rely on filtration in post-preparation process rather than the change of formulation parameters, which also explained the relatively low R^2 of RF model. So, size and PDI were mainly dependent on preparation method and technical parameters, while physicochemical properties of drug were also critical that needed to be considered.

In zeta potential prediction model by LightGBM, XLogP3 that reflected the polarity of molecules and molecular weight of lipids were ranked in the top positions. As lipids have constructed the liposome matrix, the surface charge of liposomes was mainly determined by the lipid properties. While melting point that related to the crystal lattice energy of drug molecules and XLogP3 of drugs that represented the affinity between drug-lipid molecules were ranked in 2nd and 3rd position. Overall, zeta potential was major influenced by lipid and drug properties, especially lipids showed the significant impacts.

Encapsulation was essential for liposome characterization that related to the effective drug concentration in liposomes. As Fig. 4D showed, the reduction value of logS_drug was significantly higher than other features that was the prominent influencing feature for encapsulation. Then, complexity and XLogP3 were ranked in 2nd and 3rd. logS was the 10-based logarithm of the aqueous solubility of a compound. Molecular complexity was the

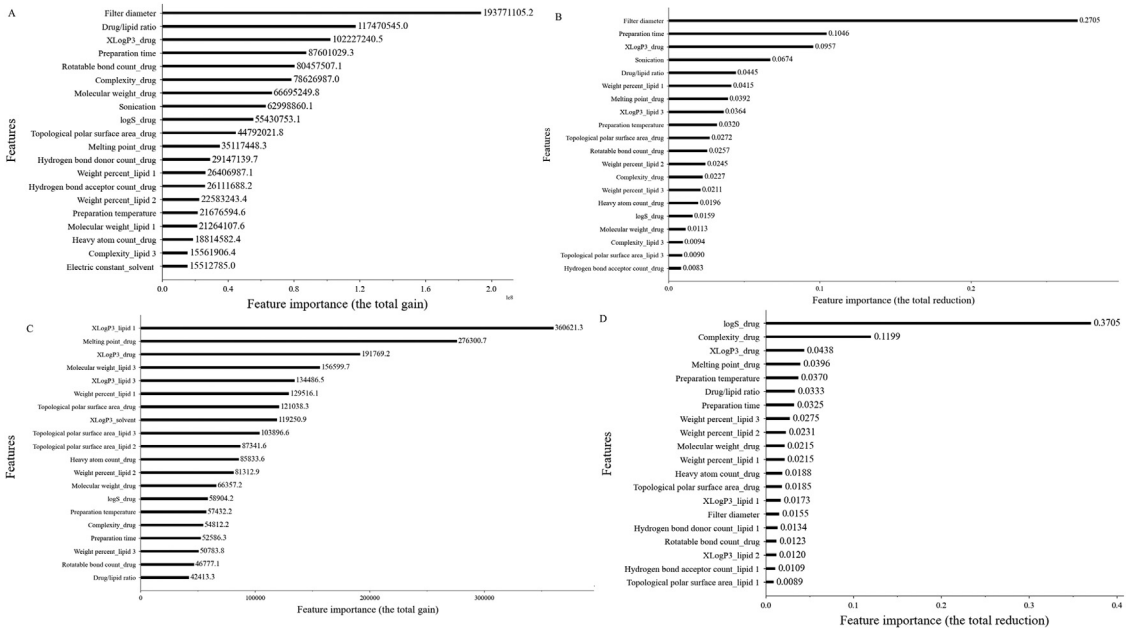

Fig. 4 – The top 20 ranking of feature importance for (A) size, (B) PDI, (C) zeta potential and (D) encapsulation by the optimal prediction model.

estimation of the complexity of a structure and computed by the Cactvs toolkit 3.4.8.18 (https://cactvs.com/) using the Bertz/Hendrickson/Ihlenfeldt formula. XLogP3 was a predicted value of the octanol-water partition coefficient using the algorithm [38]. The technical parameters, like preparation temperature and preparation time, were ranked in 5th and 7th, which reminded the appropriate method should be used for liposome preparation to obtain the superior encapsulation. Looking at the total reduction value, logS_drug and complexity_drug were obviously greater than other parameters, which indicated the higher weights of these two parameters during the process of building prediction model. This uneven distribution of parameter's weights may be the reason for low R^2 in encapsulation prediction. Although drug properties were closely related to encapsulation, it was not to say every drug was appropriate for preparing liposomes with higher drug loading efficiency. What kind of drug was better for liposomes? It was needed to be further analyzed.

Since logS, complexity and XLogP3 of drug were the main influence factors for encapsulation in feature importance ranking, the relationship between these three features with encapsulation were described as heat map that visualized data as color in two dimensions (Fig. 5). Firstly, encapsulation value was divided as 0–10, while logS, complexity, and XLogP3 were separated into several segments. The shades of color in each matrix and its number reflected the frequency of occurrence of each drug property with the corresponding encapsulation value. The legend on the right of Fig. 5 showed the number of occurrences for the color scale, the darker of color indicated the higher of frequency. Firstly, it has defined that greater than 50% was considered as high encapsulation, conversely less than 50% was low encapsulation. When the logS of drug was located in interval [-3, -6] and complexity was in [500, 1000], the color of gride that located in high encapsulation area was darker. For the relationship of XLogP3 of drug with encapsulation, it was worthy to note that the XLogP3 of most drug molecules was 2. Looking at the data frequency in the interval of XLogP3 = 2, the half of corresponding encapsulation was greater than 50% and another half was less than 50%. When it was less than 1, encapsulation was mainly situated in below 50% area. Therefore, when the logS was in [-3, -6], molecular complexity was in [500, 1000], and XLogP3 $\geq$ 2, the drugs were priority for preparing liposomes with higher encapsulation.

3.4. *Experimental validation of prediction model*

To further validate the prediction accuracy on unknown data, some liposomes with various lipid combinations were prepared by thin film hydration and anti-solvent methods in the laboratory. NAP and PAL represented two types of compounds with different molecular complexity and solubility were selected as model drugs and 24 formulations were prepared by different methods. The size, PDI, zeta potential, and encapsulation were measured and compared with prediction values as Fig. 6 listed. With the changes of lipid compositions and preparation process, liposome properties were greatly varied by experimental measurement. In the optimum prediction model for size and PDI, the majority of predicted values were very close to experimental value that data points distributed along the diagonal. For zeta potential prediction, although most of predicted data were consistent with experimental value that distributed within $\pm$ 20% error, LightGBM model was easier to predict unknown data into negative charge. It was attributed to the bias distribution of zeta potential data with negative charge in training set, but the experimental formulations were

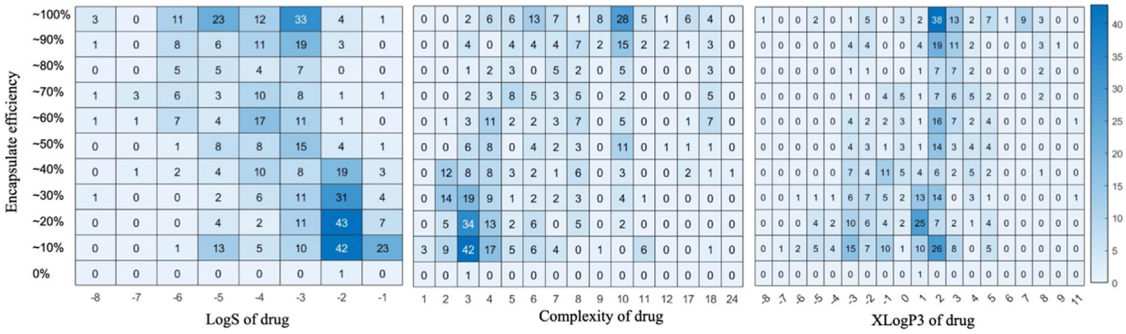

Fig. 5 – Heat map analysis for the relationship between logS, complexity and XLogP3 of drug with liposome encapsulation. The number in each matrix represented the frequency of occurrence of encapsulation corresponding to each property.
Note: −1 in logS of drug represented interval [−1, 0), −2 represented interval [−2, −1)…; 1in complexity of drug represented interval (0,100], 2 represented interval (100,200]…; 1 in XlogP3 of drug represented interval (0,1], −1 represented interval [−1, 0)….

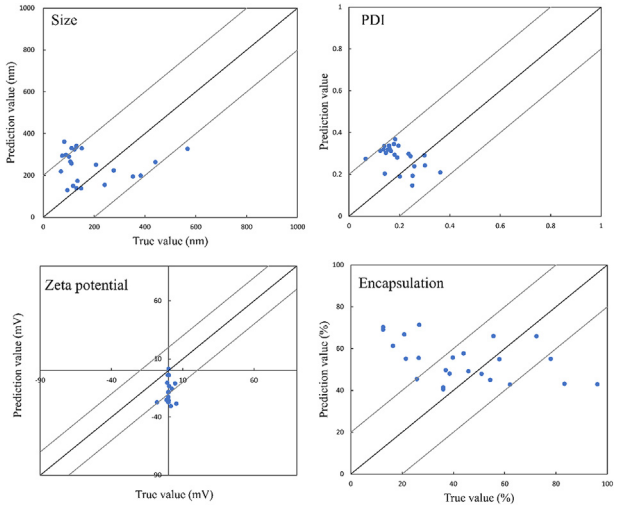

Fig. 6 – The scatter plot of experimental value and predicted value for size, PDI, zeta potential and encapsulation.

almost neutral charged. Fortunately, the feature importance analysis and previous research claimed that zeta potential was decided by lipid composition [39], the appropriate lipids could ensure surface charge of liposomes in formulation design. For the RF model in encapsulation prediction, about 60% predicted value were corelated with experimental value that distributed along diagonal within ± 20% error. According to the conclusion of feature importance, drug properties were critical for encapsulation. There were just two drugs in experiment, main variables were the formulation composition and preparation method. The sensitivity of RF model for formulation change was relatively poor, which would be the reason for inaccurate prediction in experimental data. Due to the limited amount of liposome formulation data, current prediction models just achieved good accuracy for new formulation prediction not excellent, but these also provided some guidance for formulation design to improve efficiency of formulation screening.

As drug properties were critical for liposomes, how the different drugs existed in vesicles would be related to the particle properties. However, it was still not very clear to characterize molecular interactions by experimental skills. Molecular dynamic simulation was an efficient way to mimic particle structure that can be applied for liposome structure building to see the drug distribution at molecular level.

3.5. *The structure building of liposome by molecular dynamic simulations*

AA simulation was an ideal method to reflect the atom motion of biosystem but was time-consuming with large-scale in both space and time. So, CG model was hired to describe the molecular motion of NAP− and PAL−encapsulated liposomes. Rational mapping principle and conformation parameter were critical for CG model to deduce the system properties in AA model. Fig. S1 exhibited the self-designed mapping principle of NAP and PAL from AA to CG, and the relevant dihedrals of CG models were well corresponded to AA properties. Moreover, PMF curve (Fig. S2) and Δ error (Table S4) confirmed that CG model could well reproduce the free energy change of AA model. It was sufficiently illustrated CG mode could reappear the precision of AA model with the apparent enhancement of simulation speed and volume. In the simulation process of CG model, the radius changes of gyration from axes X, Y, and Z revealed the stable state of two liposome systems during 1 μs movement (Fig. S3). Therefore, CG model was reliable for liposome structure building to reflect the distribution and interaction of molecules.

NAP ($C_{14}H_{14}O_3$) was insoluble (0.0159 mg/ml) at 25 ºC, whereas PAL ($C_{21}H_{22}NO_4^+$) existed as salt with distinct dissolution states due to the positive charge exposure after losing hydroxyl group. To balance the ions of PAL system, moderate chloride ions were added into box, which transferred PAL from slightly soluble to soluble salt. The initial inclusion systems of drug−lipid were constructed in Fig. 7A and 7A1, then these two systems were conducted 1 μs dynamic simulations. Fig. 7B and 7B1 were the snapshots during modeling process, and the final structures of two liposomes were shown in Fig. 7C and 7C1. Meanwhile, the size distribution was measured, and the structure of NAP and PAL liposomes were featured by transmission electron microscopy (TEM) in Fig. 7D and 7D1. It claimed that NAP and PAL liposome

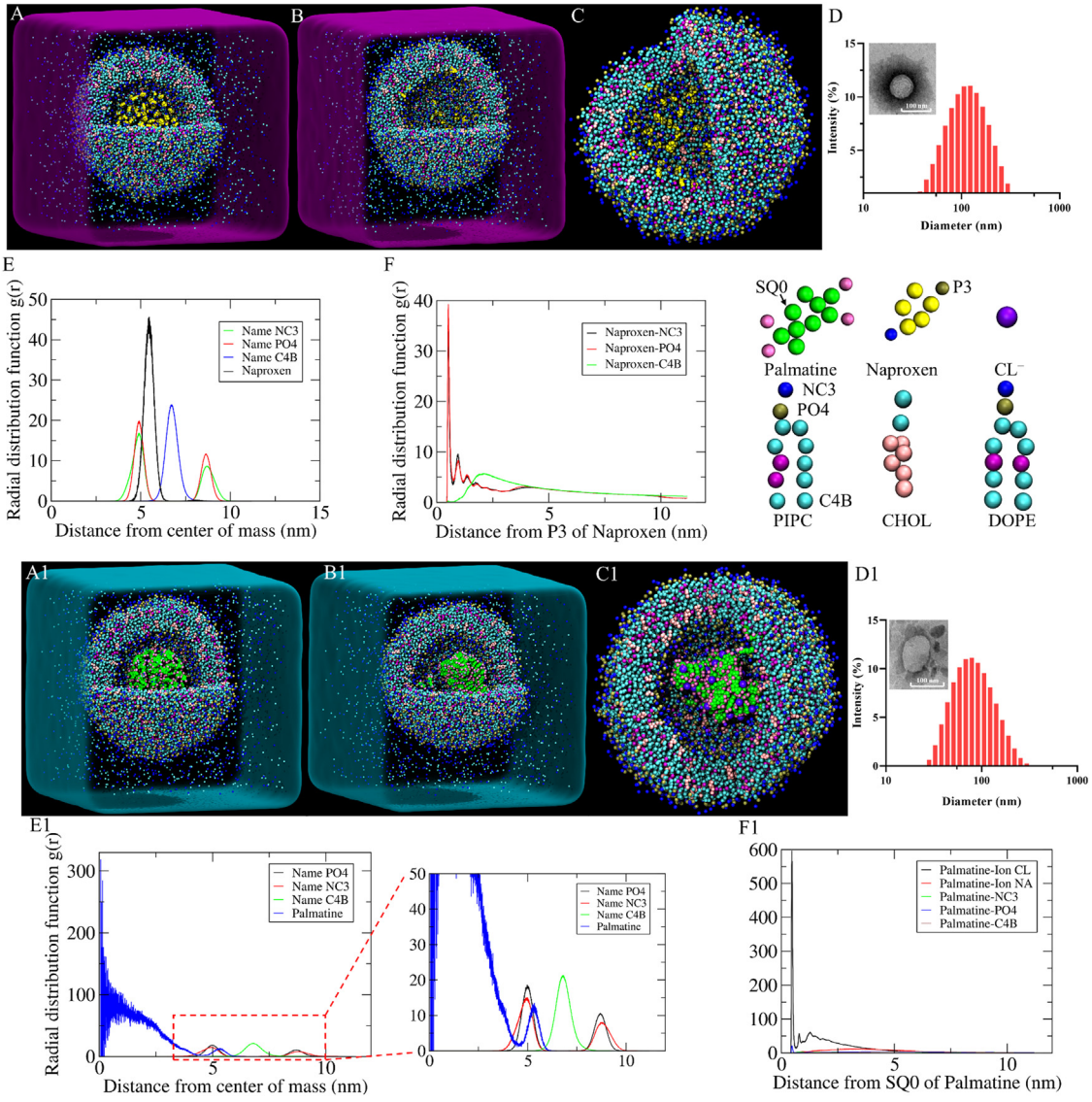

Fig. 7 – The snapshots during 1 μs molecular dynamics simulations and the experimental size distribution of NAP (A, B, C and D) and PAL (A1, B1, C1 and D1) liposomes were demonstrated. The radial distribution function of beads NC3, PO4, C4B, CL^-, NA^+ in NAP (E and F) and PAL (E1 and F1) were analyzed. In detail, A and A1 indicated the initial conformation of two liposome systems with 100 drug molecules inside the inner sphere; B and B1 represented the snapshots of section profile during the equilibrium process; C and C1 showed the final location and distribution of NAP and PAL in the vesicles; D, D1 was the size distribution and TEM image of NAP and PAL liposome; E, E1 and F, F1 calculated the radial distribution function of each ingredient of vesicles.

particles were both vesicle-liked structures with about 100 nm diameter, but the microscopic differences between these two particles were not discovered by TEM images. Thus, molecular modeling would explain the structure of liposomes at molecular level.

3.5.1. *NAP molecules were evenly distributed into lipid layer*

The molecular motion and distribution of NAP liposome system from initial state to equilibrium conformation were exhibited in Fig. 7A-7C, and supplementary motion video. Primarily, a drug sphere with 100 NAP molecules (distance between each molecule was 8.0 Å, and corresponding interaction potential was 0.44 kJ/mol) (Fig. S2) was inserted into the inner aqueous phase by replacing redundant water. Before touching the inner layer of vesicle, NAP aggregated autonomously and formed as a layer of fragment through weak interaction of intermolecular (dimerization energy: −6.5 kJ/mol) (supplementary motion video). Limited to the small volume of aqueous phase in liposome, the NAP−layer eventually contacted to the hydrophilic head of inner lipid layer, and then swiftly fused together as a fluid−mosaic state. The carboxyl head of NAP (bead P3) was arranged towards to aqueous phase, and methoxy group towards to glycerol region (Fig. 7B and 7C). The final location of each functional group

Table 6 – Average interaction energy among drugs (NAP and PAL), lipid bilayer, and water during whole simulation process using MARTINI coarse-grained model.

	Average interaction energy (kJ/mol)
NAP-NAP	−3703.6
NAP-lipid bilayer	−16,695.1
NAP-water	−2642.6
PAL-PAL	−18,337.9
PAL-lipid bilayer	−7404.9
PAL-water	−11,934.9

including NC3 (choline), PO4 (phosphate), C4B (hydrocarbon chain) were calculated by RDF from COM in Fig. 7E. It was obvious that NAP was embedded in the inner layer between the bead NC3 (green line), PO4 (red line) and C4B (blue line) with the 5.4 nm semidiameter. Interestingly, the RDF from NAP (Fig. 7F) shown the density of lipid beads NC3 and PO4 near the NAP bead PC3 was much higher than their average density, which indicated carboxyl head of NAP was attracted by the choline and phosphate function (space distance: ∼ 4.8 Å), whereas relatively far away from the hydrocarbon chain (space distance: ∼ 20 Å). Thus, NAP molecules could integrate into lipid layer to form fluidly mosaiced status by the interaction between polar molecules (carboxyl function, choline and phosphate function).

3.5.2. *PAL majorly aggregated in the inner aqueous phase of liposome*

PAL molecules (distance between each molecule was 8.0 Å, and corresponding interaction potential was −8.8 kJ/mol as Fig. S2 showed) were firstly placed into aqueous phase. Being distinct from NAP, PAL existed fully different states in dynamic simulation process as showed in Fig. 7A1, 7B1 and 7C1) and supplementary motion video. Firstly, most PAL integrated with nearby molecules including another PAL molecules and CL^- to form as aggregate cluster in aqueous core (Fig. 7B1). After 1 μs simulations, the complex of PAL^+−CL^- aggregated in the water phase and dissociative PAL hydrophobically bound to the inner lipid layer (Fig. 7C1). Relevant RDF revealed that the density of PAL in box center was higher than average density of whole system (seen in Fig. 7E1), while the lower density was also viewed in 5 nm distance and overlap with lipid beads, which indicated a few PAL molecules have embedded in the inner lipid layer (Fig. 7F1). On the other hand, the distance between bead SQ0 ($NC3^+$) of PAL molecules and CL^- (space distance: ∼ 4.0 Å) was most closely than other components in aqueous phase, which confirmed the salt formation of PAL. The radial distribution analysis has proved that the interaction force between positive and negative charge was stronger than van der Waals force between PAL and lipids.

Finally, average interaction energy including coulomb and van del Waals interaction among drugs (NAP and PAL), lipid bilayer, and water during whole simulation process was analyzed by MARTINI coarse-grained model in Table 6. It could be observed that the NAP preferred to interact with lipid bilayer (−16,695.1 kJ/mol) but not fascinated by itself (3,703.6 kJ/mol) and water (−2,642.6 kJ/mol). Whereas PAL was willing to interact with itself and water molecules rather than lipid bilayer, which demonstrated the higher solubility and the aggregation phenomenon of PAL in the aqueous core of liposome. The high correspondence of molecular modeling with experimental phenomenon indicated the reliability of this approach. So, molecular modeling was a promising way that could explain the behaviors of drug in liposome vesicle with visible images and quantitative energy analysis.

Therefore, how the drug molecules distributed in liposome was not only decided by lipid bilayer, but also by the properties of drugs. As the modeling results showed, there were not only single form for drug molecules, the drug-drug aggregation and drug-lipid interaction were also existed simultaneously in liposomes. It was hypothesized that if the distance of drug molecule and lipid membrane within 5 Å, it regarded as drug-lipid interaction, otherwise if the distance larger than 5 Å, it was drug-drug aggregation (The PMF results indicated the molecules of 5 Å showed strong attraction force to other molecules). The proportion of the existence of drug-lipid and drug-drug interaction for each liposome was calculated in Fig. S4. After 1 μs simulation, NAP molecules were all located into lipid bilayer, whereas 27% PAL molecules were distributed into lipid membrane, another 73% PAL molecules were aggregated in aqueous core.

3.6. *ML and molecular modeling synergistically constructed liposome prediction system*

Compared with previous liposome prediction models, this study was an improvement that building reasonable and widely applicable prediction models by ML and combining with molecular modeling to investigate key factors of liposomes. Previous prediction models were the initial trials that based on small datasets or internal data prepared by specific methods like turbulent jet co-flow. This study developed the more general prediction models for higher complexity liposome formulations data with passive loading method by LightGBM and RF algorithms. Although active loading was an effective way for liposome preparation with sufficient drug content, it required the drug molecules was ionizable and encapsulated agent was amphipathic, which was the technical barrier for the application of active loading. So, current study focused on liposome formulations with passive loading to investigate CQAs. The formulation dataset was not limited by specific drug or preparation method, various liposome formulations were collected from all kinds of resources, which promoted the comprehensive feature learning and extended the application of these models. LightGBM achieved the highest accuracy for size and zeta potential prediction, while RF was best for PDI and encapsulation prediction. Liposome formulation dataset covered most common lipids that could be applied for new liposome prediction. For the liposomes composed by novel synthetic lipids, it could be predicted by these prediction models if providing molecular descriptors to expand the feasibility of machine learning applying in liposome research. The quality and quantity of data were the most significant limiting factors for ML, which lead to the insufficient learning in the specific area with few data distribution. Currently, the liposome prediction platform was freely available in FormulationAI platform (https://formulationai.computpharm.org/). With more and more new formulation

data input, it can be believed that prediction ability of these models would be improved in the future.

Except predictions, LightGBM and RF algorithms could output the importance of all parameters without extra analysis. The total gain value in LightGBM and reduction value in RF as the evidence were calculated during learning process, which distinctly improved the credibility of feature importance ranking. Particle size and PDI mostly depended on preparation methods to downsize and homogenize particles and were also related to drug properties. Zeta potential was mainly determined by lipids that construct bilayer matrix. Encapsulation was closely related with drug properties. LogS, molecular complexity, and XLogP3 of drugs that reflected the solubility, spatial structure and affinity between drug and lipids were the key factors for encapsulation. Then melting point represented intermolecular energy between drug-drug molecules also influenced the encapsulation. Feature importance analysis by LightGBM and RF models were mostly in conformity with experimental experience, moreover, many deeper factors were explored to provide some novel understanding of liposomes. Therefore, the properties of each ingredient and preparation parameters comprehensively affected final liposome particles that needed to be considered from all sides.

As drug properties were critical for liposome, two kinds of drugs with different molecular structure and solubility were prepared as liposomes. NAP and PAL liposomes were characterized by experiments and simulated by molecular modeling to verify the impacts of drug properties. Although TEM was the common technique for structure characterization in experiment, it only exhibited the whole particle morphology without molecular details. Molecular dynamic simulations presented how the different drug distributed in liposome vesicles and revealed the impact of drug properties from an atomic perspective. NAP and PAL exhibited totally different molecular motion in dynamic modeling trajectory. Whether drug molecules aggregated into aqueous core or diffused in lipid bilayer was not only related with lipid bilayer vesicle, but also determined by the charge, functional group, and spatial structure of drug molecules. So, molecular dynamic simulation provided a platform for structure visualization at molecular level, which helped us to understand liposome with novel perspective.

With the mutual validation of ML models, experiments and molecular modeling, the prediction ability and key factor exploration have proved to be reliable and reasonable. Appropriate drug, lipid and preparation method should be considered to suit for liposome. Instead of traditional trial-and-error method, a series of formulations can be firstly predicted by ML models to find the best one. Then molecular modeling will build structures to exhibit drug-lipid interactions. The collaboration of ML and molecular modeling could provide some guidance for liposome formulation design.

4. Conclusion

This study has built the prediction model for liposome size, PDI, zeta potential and encapsulation by LightGBM and RF algorithms. Whether in training set, validation set, test set or un-known experimental data, the prediction models have obtained satisfied accuracy. However, there were few experimental formulations were not predicted very well, which due to the limitation of the small data volume and biased data distribution. Then, feature importance for size, PDI, zeta potential and encapsulation were evaluated to find the CQAs for liposome formulation design. Generally, drug molecules with logS [-3, -6], molecular complexity [500, 1000] and XLogP3$\geq$2 were preferred to be prepared as liposomes with higher encapsulation. Then NAP and PAL with different solubility were designed with various liposome formulations to validate the prediction accuracy of models on un-known data, while particle structures were described by CG modeling. The distribution and interaction way of drug molecules in lipid bilayers were exhibited at molecular level and the differences between these two liposomes were compared, which explained the impacts of drug properties for liposomes. Therefore, it was a successful trial to build general liposome formulation prediction models and analyze particle structure by CG modeling. In the future study, it will be an effective tool for liposome formulation design and optimization.

Conflicts of interest

The authors report no conflicts of interest. The authors alone are responsible for the content and writing of this article.

Acknowledgements

This work was supported by the Multi-Year Research Grants from the University of Macau (MYRG2019–00032-ICMS and MYRG2020–00113-ICMS) and the Macau FDCT research grant (0108/2021/A). Molecular modeling was performed at the High-Performance Computing Cluster (HPCC), which is supported by the Information and Communication Technology Office (ICTO) of the University of Macau.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ajps.2023.100811.

References

[1] Bangham A, Standish MM, Watkins JC. Diffusion of univalent ions across the lamellae of swollen phospholipids. J Mol Bio 1965;13(1) 238-IN27.
[2] Gregoriadis G, Leathwood P, BE Ryman. Enzyme entrapment in liposomes. FEBS Lett. 1971;14(2):95–9.
[3] Barenholz Y. Doxil (R) - The first FDA-approved nano-drug: lessons learned. J Control Release 2012;160(2):117–34.
[4] Wang Y, Grainger DW. Lyophilized liposome-based parenteral drug development: reviewing complex product design strategies and current regulatory environments. Adv Drug Deliver Rev 2019;151:56–71.

[5] Papahadjopoulos D, Kimelberg HK. Phospholipid vesicles (liposomes) as models for biological membranes: their properties and interactions with cholesterol and proteins. Prog Surf Sci 1974;4:141–232.
[6] Frolov VA, Shnyrova AV, Zimmerberg J. Lipid polymorphisms and membrane shape. Cold Spring Harb Perspect Biol 2011;3(11):a004747.
[7] Wagner A, Vorauer-Uhl K. Liposome technology for industrial purposes. J Drug Deliver 2011;2011:591325.
[8] Woodbury DJ, Richardson ES, Grigg AW, Welling RD, BH Knudson. Reducing liposome size with ultrasound: bimodal size distributions. J Liposome Res 2006;16(1):57–80.
[9] Allen TM, Cullis PR. Liposomal drug delivery systems: from concept to clinical applications. Adv Drug Deliver Rev 2013;65(1):36–48.
[10] Wang W, Ye ZYF, Gao HL, Ouyang DF. Computational pharmaceutics-a new paradigm of drug delivery. J Control Release 2021;338:119–36.
[11] Soema PC, Willems GJ, Jiskoot W, Amorij JP, Kersten GF. Predicting the influence of liposomal lipid composition on liposome size, zeta potential and liposome-induced dendritic cell maturation using a design of experiments approach. Eur J Pharm Biopharm 2015;94:427–35.
[12] Sedighi M, Sieber S, Rahimi F, Shahbazi MA, Rezayan AH, Huwyler J, et al. Rapid optimization of liposome characteristics using a combined microfluidics and design-of-experiment approach. Drug Deliv Transl *Re* 2019;9(1):404–13.
[13] Xu X, Khan MA, Burgess DJ. Predicting hydrophilic drug encapsulation inside unilamellar liposomes. Int J Pharm 2012;423(2):410–18.
[14] Han R, Xiong H, Ye Z, Yang Y, Huang T, Jing Q, et al. Predicting physical stability of solid dispersions by machine learning techniques. J Control Release 2019;311-312:16–25.
[15] Gao HL, Wang W, Dong J, Ye ZYF, Ouyang DF. An integrated computational methodology with data-driven machine learning, molecular modeling and PBPK modeling to accelerate solid dispersion formulation design. Eur J Pharm Biopharm 2021;158:336–46.
[16] He Y, Ye ZYF, Liu XY, Wei ZJ, Qiu F, Li HF, et al. Can machine learning predict drug nanocrystals? J Control Release 2020;322:274–85.
[17] Gao H, Jia H, Dong J, Yang X, Li H, Ouyang D. Integrated in silico formulation design of self-emulsifying drug delivery systems. Acta Pharm Sin B 2021;11(11):3585–94.
[18] Subramanian N, Yajnik A, Murthy RSR. Artificial neural network as an alternative to multiple regression analysis in optimizing formulation parmaeters of cytarabine liposomes. AAPS PharmSciTech 2004;5(1):11–19.
[19] Sansare S, Duran T, Mohammadiarani H, Goyal M, Yenduri G, Costa A, et al. Artificial neural networks in tandem with molecular descriptors as predictive tools for continuous liposome manufacturing. Int J Pharm 2021;603:120713.
[20] Zucker D, Marcus D, Barenholz Y, Goldblum A. Liposome drugs' loading efficiency: a working model based on loading conditions and drug's physicochemical properties. J Control Release 2009;139(1):73–80.
[21] Wilkhu JS, Ouyang DF, Kirchmeier MJ, Anderson DE, Perrie Y. Investigating the role of cholesterol in the formation of non-ionic surfactant based bilayer vesicles: thermal analysis and molecular dynamics. Int J Pharm 2014;461(1–2):331–41.
[22] Jambeck JPM, Eriksson ESE, Laaksonen A, Lyubartsev AP, Eriksson LA. Molecular dynamics studies of liposomes as carriers for photosensitizing drugs: development, validation, and simulations with a coarse-grained model. J Chem Theory Comput 2014;10(1):5–13.
[23] Li JJ, Jin X, Zhang LL, Yang Y, Liu R, Li Z. Comparison of different chitosan lipid nanoparticles for improved ophthalmic tetrandrine delivery: formulation, characterization, pharmacokinetic and molecular dynamics simulation. J Pharm Sci-Us 2020;109(12):3625–35.
[24] Wu XW, Dai XX, Liao YY, Sheng MK, Shi XY. Investigation on drug entrapment location in liposomes and transfersomes based on molecular dynamics simulation. J Mol Model 2021;27(4):111.
[25] Hashemzadeh H, Javadi H, Darvishi MH. Study of structural stability and formation mechanisms in DSPC and DPSM liposomes: a coarse-grained molecular dynamics simulation. Sci Rep-Uk 2020;10(1):1837.
[26] Yang Y, Ye Z, Su Y, Zhao Q, Li X, Ouyang D. Deep learning for *in vitro* prediction of pharmaceutical formulations. Acta Pharm Sin B 2019;9(1):177–85.
[27] Abhimanyu T, Ambika Prasad M, Bishnupriya P, Kumari S, Babita M. Detection of disease-specific parent cells via distinct population of nano-vesicles by machine learning. Curr Pharm Des 2020;26(32):3985–96.
[28] Bishnupriya P, Babita M, Abhimanyu T. An Integrated-OFFT model for the prediction of protein secondary structure class. Curr Computer-Aided Drug Des 2019;15(1):45–54.
[29] Grinsztajn L., Oyallon E. and Varoquaux G. Why do tree-based models still outperform deep learning on tabular data? 2022; arXiv:2207.08815.
[30] Breiman L. Random forests. Mach Learn 2001;45(1):5–32.
[31] Bergstra J, Bengio Y. Random search for hyper-parameter optimization. J Mach Learn Res 2012;13:281–305.
[32] Souza PCT, Alessandri R, Barnoud J, Thallmair S, Faustino I, Grunewald F, et al. Martini 3: a general purpose force field for coarse-grained molecular dynamics. Nat Methods 2021;18(4):382–8.
[33] Marrink SJ, Risselada HJ, Yefimov S, Tieleman DP, de Vries AH. The MARTINI force field: coarse grained model for biomolecular simulations. J Phys Chem B 2007;111(27):7812–24.
[34] Jo S, Kim T, Iyer VG, Im W. CHARMM-GUI: a web-based graphical user interface for CHARMM. J Comput Chem 2008;29(11):1859–65.
[35] Parrinello M, Rahman A. Polymorphic transitions in single-crystals - a new molecular-dynamics method. J Appl Phys 1981;52(12):7182–90.
[36] Bussi G, Donadio D, Parrinello M. Canonical sampling through velocity rescaling. J Chem Phys 2007;126(1):014101.
[37] Mathias G, Egwolf B, Nonella M, Tavan P. A fast multipole method combined with a reaction field for long-range electrostatics in molecular dynamics simulations: the effects of truncation on the properties of water. J chem phys 2003;118(24):10847–60.
[38] Cheng T, Zhao Y, Li X, Lin F, Xu Y, Zhang X, et al. Computation of octanol-water partition coefficients by guiding an additive model with knowledge. J Chem Inf Model 2007;47(6):2140–8.
[39] Filipczak N, Pan JY, Yalamarty SSK, Torchilin VP. Recent advancements in liposome technology. Adv Drug Deliver Rev 2020;156:4–22.