Systems & Control Transactions

Research Article - Peer Reviewed Conference Proceeding
ESCAPE 35 - European Symposium on Computer Aided Process Engineering
Ghent, Belgium. 6-9 July 2025
Jan F.M. Van Impe, Grégoire Léonard, Satyajeet S. Bhonsale,
Monika E. Polańska, Filip Logist (Eds.)

PSE PRESS

Balancing modelling complexity and experimental effort for conducting QbD on lipid nanoparticles (LNPs) systems

Daniel V. Batista[a] and Marco S. Reis[a*]

[a] University of Coimbra, CERES, Department of Chemical Engineering, Coimbra, Portugal
*Corresponding Author: marco@eq.uc.pt.

ABSTRACT

The promising properties of lipid nanoparticles (LNPs) as drug carriers have been attracting significant attention in the field of drug delivery. However, further research is still required for a better understanding of their integration in the pharmaceutical industry. The Quality by Design (QbD) approach aims at ensuring the safety and efficiency in the development of new drugs, through an holistic, risk-based approach that gathers all sources of knowledge available about the system under analysis. One key resource of the QbD framework is the rich toolkit of Design of Experiments (DOE), to deepen the understanding of how the synthesis of LNPs by microfluidics can be effectively conducted and controlled. This study aimed to explore and understand the effectiveness of different DOE strategies, through an *in silico* study focused on the impact of factors related to the LNPs synthesis, namely the molar ratio of each lipid component in the lipidic mixture and the N/P ratio, while also considering potential economic constraints without disregarding the need for a statistically valid analysis. A simulation model of the LNP synthesis derived from real experiments was adopted as a basis to assess the potential efficacy of estimated models with different levels of complexity, to extract useful insights in future DOEs in these types of systems, given the high cost of each experimental run. The statistical metrics used were the coefficient of determination (R^2) and the Root Mean Squared Error (RMSE). With the results obtained, it was possible to verify that different responses from the same system could require quite different model structures, namely, the models developed for potency and for size of the LNPs differed significantly in their complexity. Furthermore, a number of experiments of the order of 30 can be anticipated as necessary for a DOE in a real process, involving similar factors.

Keywords: Lipid nanoparticles (LNPs), Quality by Design (QbD), Design of Experiments (DOE).

INTRODUCTION

Lipid nanoparticles (LNPs) have recently attracted considerable interest, particularly for their role in the development of COVID-19 mRNA vaccines by Pfizer-BioNTech and Moderna. Their ability to efficiently encapsulate nucleic acids, while simultaneously ensuring effective intracellular delivery and endosomal escape, has sparked increasing interest from both the industrial and research communities for exploring their unique properties as promising drug carriers.

A widely used technique for the synthesis of these LNPs, is microfluidics. This method allows for the efficient mixing of an organic solution, which contains the necessary lipid components dissolved in ethanol, with an aqueous solution that contains the intended payload (nucleic acid).

To ensure the successful and safe synthesis of these LNPs attributes, the pharmaceutical industry typically recommends following a Quality by Design (QbD) approach. This approach aims to guarantee that manufacturing is able to deliver the expected quality of the drugs, rather than relying solely on testing the final product and reject any lot that is non-conforming. QbD can also reduce costs associated with product optimization and ensure that patients receive higher-quality drugs in a shorter time frame [1]. One of the key aspects of the QbD approach is the use of Design of Experiments (DOE) to establish the Design Space that guarantees the quality requirements of the LNPs are met [1]. However, before

defining a design space, several DOE stages may be necessary for screening the important factors, modelling the system's behaviour accurately, and finding the optimal operational conditions.

GOLDEN STANDARD MODEL

As previously highlighted, one of the key aspects of the QbD approach is the use of the DOE. However, the employment of this statistical method can greatly benefit from a preliminary simulated *in silico* study approach, especially when economic constraints play a significant role. This is especially true in the case of LNPs' synthesis by microfluidics, where each experiment is expensive due to the high cost of the formulation components. Therefore, there is a justified interest in making this process as efficient and informative as possible.

In light of this information, we have adopted an *in silico* system, using a model that was developed from real experimental data collected in a closely related experimental and technological scenario. This *in silico* study provides a suitable test bed to analyse and compare the different DOE strategies that may be adopted and collect insights about a reasonable number of experiments to accommodate within a designated budget, while ensuring a statistically valid analysis.

Therefore, we have conducted a systematic study based on the work developed by Karl et al. [2], who provided a simulation model of the LNP synthesis, referred to as the Golden Standard (GS) Model. This model was derived from 23 real experiments and codified in the JMP Pro software using a recently proposed methodology called self-validated ensemble model (SVEM).

The GS Model considers the molar ratio of individual lipid components in the lipid mixture (Ionizable lipid (IL), Structural lipid (SL), Helper lipid (HL) and PEG lipid), the ionizable lipid type, the molar N/P ratio (ionizable amine from ionizable lipids to phosphate from nucleic acid ratio) and the total flow rate (TFR). These factors were established within the ranges, shown in **Table 1**.

Table 1: Ranges of the factors in the GS Model.

Factors	Ranges
IL	[0,1; 0,6]
HL	[0,1; 0,6]
SL	[0,1; 0,6]
PEG	[0,01; 0,05]
Ionizable lipid type	H101 or H102 or H103
N/P ratio	[6,14]
TFR	[1,3] mL/min

It is important to note that the GS model is quite complex in its structure and it was considered, for all practical purposes, unknown throughout this study. Moreover, it has terms that were not included in the DOEs strategies considered. This structural mismatch brings more realism to the *in silico* study carried out, which will be lay down in the next section.

Finally, this model considered as responses the "potency in vivo" and "size of the LNP", with the goal being to maximize potency and minimize size.

METHODOLOGY

The study conducted is based on the GS Model, and followed a methodology that is schematicaly illustrated in **Figure 1**. This figure that depicts the principal steps of the study, organized into two distinct phases: 1) Initial setup; 2) Comparison study.

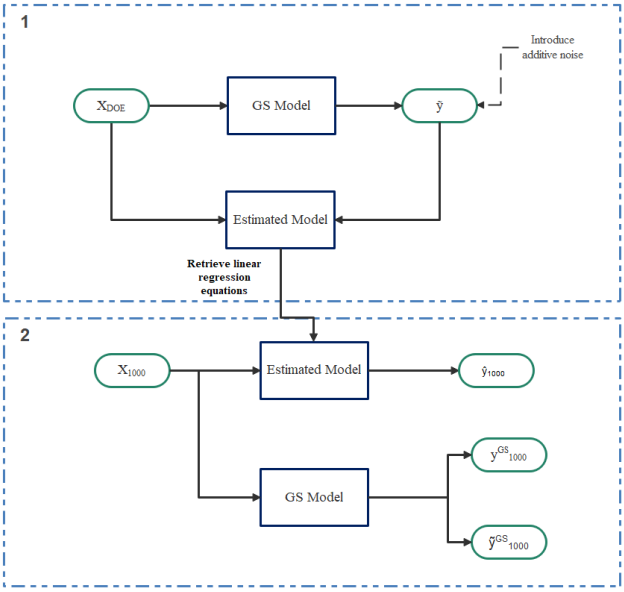

Figure 1: Methodology for the *in silico* study.

Initial Setup

Firstly, the DOE levels (X_{DOE}) for the factors selected, need to be established. For such, an interval of acceptability for each factor was established while respecting the ranges established by the GS model, as shown in **Table 1**. The ranges established and the chosen factors are shown below in **Table 2**.

Table 2: Ranges for the factors considered in X_{DOE}.

Factors	Ranges
IL	[0,3;0,6]
HL	[0,1;0,28]
SL	[0,185;0,555]
PEG	[0,01;0,045]
N/P ratio	[6,14]

Taking into account the constraints in **Table 2**, the X_{DOE} was generated by a D-optimal design approach for different model structures, spanning different levels of modeling complexity. These model structures present an increasing number of terms (complexity), by considering more effects to be estimated from data. Consequently, more experiments were necessary to estimate them. The models were designated as Model 1 (M1), Model 2 (M2), and Model 3 (M3), with M3 being the most complex in terms of considered effects. In M1 only the main effects associated with the factors were considered. On the other hand, M2 considered the same effects as M1, except for the main effect associated with the N/P ratio. Additionally, M2 considered the effects related to the non-linear blending of two mixture components and the interactions of mixture components with the process factor (N/P ratio). Finally, M3 incorporated the same effects as M2, along with some third-order terms, considering the Scheffé cubic model effects, such as the non-linear blending associated with three components. The effects considered for each model are presented in **Table 3**.

With regard to the number of experiments, between 10 to 30 experiments were considered, in accordance with the corresponding minimum limits suggested by optimal design software (JMP Pro), that increase according to the increment of their complexity. Therefore, a designation of the number of experiments was explicitly included together with that for the corresponding model. For example, M1-10 represents the model M1 that was generated considering a total of 10 experiments.

Next, the X_{DOE}'s elaborated were introduced in the GS Model to obtain the simulated responses with additive noise ($\tilde{y}$). Note that, when introducing the levels for the selected factors in the model (X_{DOE}), the levels for the other factors not considered in the designs have also to be set to run the simulation; therefore, the following factors were established as constants for the GS Model to run: a TFR of 2 mL/min and H103 as the ionizable lipid type.

Moreover, three levels of noise in the response were considered: the originally estimated noise level (during the development of the GS model) and two additional levels corresponding to an increase of 10% or 20%. This alternative is indicated by a dashed arrow in **Figure 1** and was implemented in an effort to better reflect the reality of having different levels of reproducibility of the measured responses.

Based on the X_{DOE}'s and the corresponding outputs, $\tilde{y}$, with or without additive noise, a linear regression model was estimated for each model structure, as mentioned above, by fitting these models using the least square method. These models will be collectively referred throughout this article, as Estimated Models (EM).

Comparison study

In the second part of the study, a new input space of 1000 points (X_{1000}) was generated at random for the mixture components and the N/P ratio. The new input space was generated to guarantee a proper assessment of the predictive capability of the different models and, therefore, the comparison between the GS Model and the EM Models. In other words, this new dataset is not used for conducting any estimation task, but simply to compare the GS Model responses with and without noise ($\tilde{y}^{GS}_{1000}$ and y^{GS}_{1000}) with the EM Models responses ($\hat{y}_{1000}$).

One of the metrics used for this comparison was the coefficient of determination, R^2, which measures the quality of fit of the EM Model in relation to the GS Model responses (with or without noise). A higher R^2 value indicates a superior fit to the GS data, i.e., a higher proportion of variation that is explained by each EM Model.

Another metric used was the Root Mean Squared Error (RMSE), which describes the average difference between the GS values (with or without noise) and the estimated values, as shown below in Equation (1) and Equation (2).

$$\mathrm{RMSE}_{\mathrm{w/o\ noise}} = \sqrt{\frac{1}{n} \times \sum_{i=1}^{n} \left(y^{GS}_{1000}(i) - \hat{y}_{1000}(i) \right)^2} \quad (1)$$

$$\mathrm{RMSE}_{w/noise} = \sqrt{\frac{1}{n} \times \sum_{i=1}^{n} \left(\tilde{y}^{GS}_{1000}(i) - \hat{y}_{1000}(i) \right)^2} \quad (2)$$

Note that the metrics mentioned above were computed for the test set. Therefore, they are actually measures of quality of prediction (prediction metrics).

ANALYSIS AND DISCUSSION OF THE RESULTS

Figure 2 summarizes the R^2 and RMSE for each response (potency and size) that were computed for each EM Model by comparison with the GS Model with or without noise.

Firstly, it is important to note that the R^2 and RMSE trends of the different EM Models remained

Table 3: Effects considered for models M1, M2 and M3.

Models	Effects
M1	PEG, HL,IL, SL and N/P ratio
M2	PEG, HL, IL, SL, PEG×IL, PEG×HL, PEG×SL, PEG×N/P ratio, HL×IL, HL×SL, HL×N/P ratio, IL×SL, IL×N/P ratio and SL×N/P ratio
M3	PEG, HL, IL, SL, PEG×IL, PEG×HL, PEG×SL, PEG×N/P ratio, HL×IL, HL×SL, HL×N/P ratio, IL×SL, IL×N/P ratio, SL×N/P ratio, PEG×HL×IL, PEG×HL×SL, PEG×IL×SL, HL×IL×SL, PEG×HL×(PEG-HL),PEG×IL×(PEG-IL), HL×IL×(HL-IL), PEG×SL×(HL-SL),HL×SL×(HL-SL) and IL×SL×(IL-SL)

consistent regardless of the presence of noise in GS Models outputs, as well as considering additional levels of noise. The only difference occurs when considering the various levels of noise, where the values reported were lower and higher for R^2 and RMSE, respectively.

The R^2 for the EM Models differed for potency and size. For size, it was reported a better fit of the EM Models to the GS Model with values typically around between 0,97 and 0,99. Moreover, when compared to potency, the shift in the behaviour of the R^2 values in accordance with the models was slightly less significant; see **Figure 2 (b)**.

For potency, the R^2 values displayed a general increasing trend, **Figure 2 (a)**, as the complexity of EM Models and the number of experiments increased, with the exception of M3, which exhibited a slightly lower R^2 value compared to M2. On the other hand, the RMSE values displayed a general decreasing trend, **Figure 2 (c)**, as the complexity and the number of experiments increased, with the exception of M3 which exhibited a slightly higher RMSE value compared to M2.

According to these findings, M2-30 appears to be the most effective alternative due to its high R^2 and a low RMSE. However, if a reduced number of experiments was necessary, M2-20 would still be a viable option, as it presented a slightly lower R^2, but a lower RMSE in comparison with M2-30.

On the other hand, regarding size, the R^2 and RMSE values showed different trends across the number of experiments and models' complexity. In the case of the R^2 values, an increment in the number of experiments did not necessarily lead to an increase in R^2, **Figure 2 (b)**. In fact, there was a slight decrease in R^2 values from M1-10 to M1-20, and an increase in M1-30. Following the sequence in this figure, there is then a decrease in M2-20 and an increase in M2-30 and M3-30, with M3-30 showing the highest value.

With regard to RMSE, **Figure 2 (d)**, an increase in the number of experiments and complexity from M1-10 to M2-10 resulted in a decrease in values similar to the previous case. However, there was an increase in RMSE at M2-20, and a subsequent decrease until M3-30, which corresponds to the lowest RMSE value.

According to these findings, the M3-30 would likely be a viable option as it reported the highest R^2 and lowest RMSE value. It is important to note, that for size the R^2 values reported generally only showed changes around their 3rd or 4th decimal place. Therefore, M2-30, would also be a viable option, despite being the 3rd highest R^2 value. Moreover, it had the lowest RMSE value.

Based on these findings, in both cases, rather surprisingly, the M2-30 model was considered a good possibility, although it was not the first choice for size. Therefore, using the same model is still a viable option.

Nevertheless. the most effective EM Model for potency and size differed: M2-30 for potency and M3-30 for size. Thus, it is also likely that different responses may not necessarily require the same EM Model. Still, the suggested number of experiments is the same, which is informative for future studies involving similar factors.

It is important to note that, as previously stated, there was a notable disparity between size and potency in R^2 values, with R^2 values for size being consistently higher (around 0,97 to 0,99), in comparison to those for potency (around 0,30 to 0,88). The reason behind these results was also object of analysis.

This discrepancy could be attributed to the fact that not all the responses, as previously stated, behave necessarily equally, and may follow different mechanisms, which may differ in their inner complexity.

Furthermore, the measurement systems may also be more limited in some cases, inputting more uncertainty in data, which is transferred to the model development stage; this also may affect more extensively one type of model than the other.

Moreover, the fact that the models being

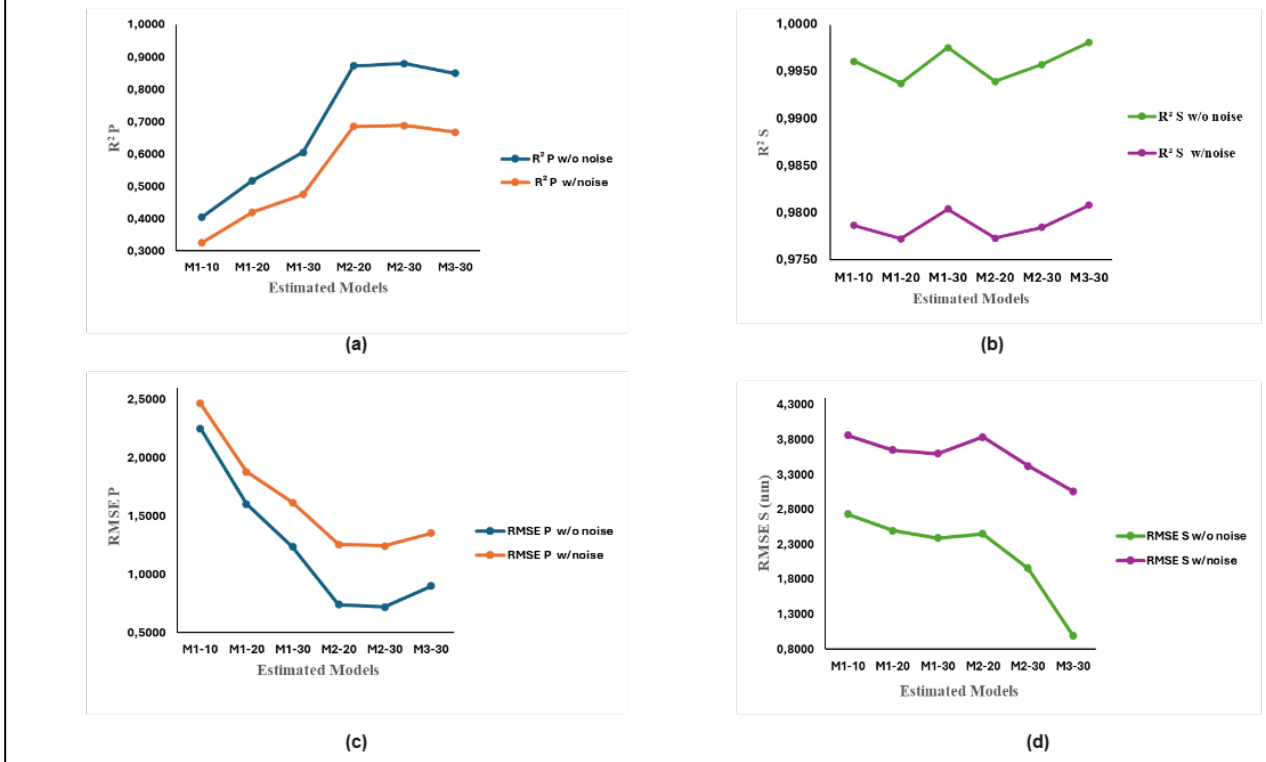

Figure 2: Results of R^2 and RMSE according to the Estimated Models: (a) R^2 of potency (P) with and without noise; (b) R^2 of size (S) with and without noise; (c) RMSE of potency (P) (units not referred in the original article [2]) with and without noise (d); RMSE of size (S) with and without noise.

considered in the R^2 metric (EM Model and GS Model) are structurally different could also be a possibility for this discrepancy. This mismatch between them is due to the fact, that the GS Model for potency was generated considering additional effects (e.g. PEG×PEG×N/P ratio, HL×HL×PEG, emphasized in bold in Equation 3) along with an intercept term, the same occurring in size for GS Model. A portion of the equation of GS Model for potency is shown below in Equation 3:

$$\text{Potency}_{GS} = -4{,}6532 + 1{,}0608 \times \left(69{,}1343 + 0{,}3518 \times \left(\frac{(\text{PEG}-0{,}01)}{0{,}69}\right) + 0{,}0697 \times \left(\frac{(\text{HL}-0{,}1)}{0{,}69}\right) - 1{,}8941 \times \left(\frac{(\text{IL}-0{,}1)}{0{,}69}\right) + 0{,}1734 \times \left(\frac{(\text{SL}-0{,}1)}{0{,}69}\right) + \text{Ionizable lipid type} + 0{,}0005 \times \left(\frac{(\text{N/P ratio}-10)}{4}\right) - 66{,}0252 \times \left(\frac{(\text{PEG}-0{,}01)}{0{,}69}\right)\left(\frac{(\text{PEG}-0{,}01)}{0{,}69}\right)\left(\frac{(\text{HL}-0{,}1)}{0{,}69}\right) - 270{,}8798 \times \left(\frac{(\text{PEG}-0{,}01)}{0{,}69}\right)\left(\frac{(\text{PEG}-0{,}01)}{0{,}69}\right)\left(\frac{(\text{IL}-0{,}1)}{0{,}69}\right) + 30{,}8712 \times \left(\frac{(\text{PEG}-0{,}01)}{0{,}69}\right)\left(\frac{(\text{PEG}-0{,}01)}{0{,}69}\right)\left(\frac{(\text{SL}-0{,}1)}{0{,}69}\right) + \left(\frac{(\text{PEG}-0{,}01)}{0{,}69}\right)\left(\frac{(\text{PEG}-0{,}01)}{0{,}69}\right) \text{Ionizable lipid type} - 164{,}3110 \times \left(\frac{(\mathbf{PEG}-\mathbf{0{,}01})}{\mathbf{0{,}69}}\right)\left(\frac{(\mathbf{PEG}-\mathbf{0{,}01})}{\mathbf{0{,}69}}\right)\left(\frac{(\mathbf{N/P\ ratio}-\mathbf{10})}{\mathbf{4}}\right) - 103{,}2265 \times \left(\frac{(\text{PEG}-0{,}01)}{0{,}69}\right)\left(\frac{(\text{PEG}-0{,}01)}{0{,}69}\right)(\text{TFR}-2) - 1{,}1640 \times \left(\frac{(\mathbf{HL}-\mathbf{0{,}1})}{\mathbf{0{,}69}}\right)\left(\frac{(\mathbf{HL}-\mathbf{0{,}1})}{\mathbf{0{,}69}}\right)\left(\frac{(\mathbf{PEG}-\mathbf{0{,}01})}{\mathbf{0{,}69}}\right) + 1{,}1584 \times \left(\frac{(\text{HL}-0{,}1)}{0{,}69}\right)\left(\frac{(\text{HL}-0{,}1)}{0{,}69}\right)\left(\frac{(\text{IL}-0{,}1)}{0{,}69}\right)(\ldots) - 1{,}3789 \left(\frac{(\text{IL}-0{,}1)}{0{,}69}\right)\left(\frac{(\text{SL}-0{,}1)}{0{,}69}\right)\left(\frac{(\text{IL}-0{,}1)}{0{,}69} - \frac{(\text{SL}-0{,}1)}{0{,}69}\right)\right) \quad (3)$$

As previously shown in **Table 1** and **Table 2**, the GS Model factors' ranges differed from the ones considered for the EM Model.

These conjectures were further analyzed and corroborated through a comparison of the prediction profiles of the GS Model for potency and size, with the prediction profiles of the EM Models (M2-30 for potency and M3-30 for size).

The predicted behaviour for size in the profiles for the GS Model did not differ much when compared to the M3-30. However, when comparing for potency between the GS Model and M2-30, the M2-30 was not capable of capturing completely the impact that certain factors held

on the response behaviour (e.g. PEG and N/P ratio). Such profiles for potency are shown below in **Figure 3**.

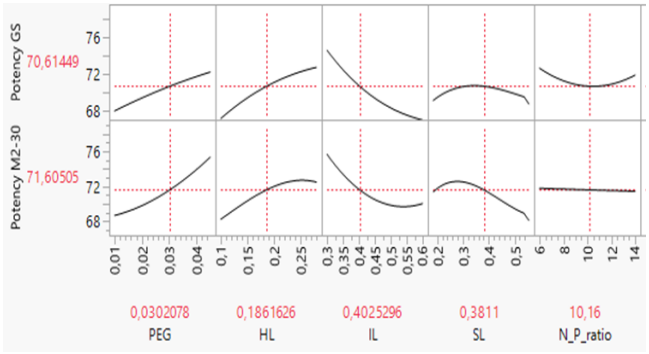

Figure 3: Prediction profiles for potency in GS Model and M2-30.

According to the findings, the proposed methodology provides a better understanding of the number of experiments and complexity of the models required for each response, thereby preventing unnecessary expenses. Furthermore, given that distinct results for each response were obtained, our approach also points out to the need to carefully tailor the DOE model case-by-case. In summary, this approach has demonstrated potential in addressing upfront challenges associated with the design space through *in silico* experiments to streamline the development of the design space in real systems while balancing cost-effectiveness and quality effectiveness.

CONCLUSIONS

In the preliminary study conducted to test the effectiveness of different DOE strategies, considering the selected metrics (R^2 and RMSE), it was found that the responses (potency and size) were modelled with different levels of accuracy according to the EM Models obtained. From our findings, the main viable options considered were M2-30 for potency and M3-30 for size. Therefore, it is expected that for exploring the formulation space with a similar number of factors, a number of experiments of the order of 30 may be anticipated.

In the future, further research using different models with varying factors, will be conducted, such as incorporating only the lipidic composition of three components as factors, as it could substantially contribute to advancing the understanding of the models' effectiveness, thereby facilitating the future experimental application of the DOE, and consequently pave the way for a more comprehensive QbD approach. Furthermore, the DOE methodologies adopted in this study (optimal DOE) will be compared with other Active Learning methods, such as Bayesian Optimization, on their ability to efficiently reach optimal values for the properties, or optimal compromises between them.

ACKNOWLEDGEMENTS

This work was supported by the CInTech project - Technological Hub for Innovation, Translation and Industrialisation of Complex Injectable Drugs -, under reference no. C644865576-00000005, co-financed by Componente C5 - Capitalização e Inovação Empresarial integrada na Dimensão Resiliência do Plano de Recuperação e Resiliência (PRR), through the NextGenerationEU fund.

The authors also acknowledge the support from the Fundação para a Ciência e Tecnologia, I.P., through the projects with DOI: 10.54499/UIDB/ 00102/2020 and DOI :10.54499/UIDP/00102/2020.

REFERENCES

1. Gurba-Bryśkiewicz L, Maruszak W, Smuga DA, Dubiel K, Wieczorek M. Quality by Design (QbD) and Design of Experiments (DOE) as a Strategy for Tuning Lipid Nanoparticle Formulations for RNA Delivery. Biomedicines. 11(10):2752 (2023) https://doi.org/10.3390/biomedicines11102752
2. Karl AT, Essex S, Wisnowski J, Rushing H. A Workflow for Lipid Nanoparticle (LNP) Formulation Optimization using Designed Mixture-Process Experiments and Self-Validated Ensemble Models (SVEM). JoVE. 198:65200 (2023) https://doi.org/10.3791/65200