Journal of Pharmaceutical Analysis 14 (2024) 100996

Contents lists available at ScienceDirect

Journal of Pharmaceutical Analysis

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

Original article

Machine learning-driven optimization of mRNA-lipid nanoparticle vaccine quality with XGBoost/Bayesian method and ensemble model approaches

Ravi Maharjan [a], Ki Hyun Kim [a, b], Kyeong Lee [a], Hyo-Kyung Han [a], Seong Hoon Jeong [a, *]

[a] BK21 FOUR Team and Integrated Research Institute for Drug Development, College of Pharmacy, Dongguk University, Gyeonggi, 10326, Republic of Korea
[b] College of Pharmacy, Mokpo National University, Jeonnam, 58554, Republic of Korea

ARTICLE INFO

Article history:
Received 3 January 2024
Received in revised form
21 April 2024
Accepted 3 May 2024
Available online 8 May 2024

Keywords:
Vaccine manufacturing
Microfluidic device
XGBoost
Bayesian optimization
Self-validated ensemble model

ABSTRACT

To enhance the efficiency of vaccine manufacturing, this study focuses on optimizing the microfluidic conditions and lipid mix ratios of messenger RNA-lipid nanoparticles (mRNA-LNP). Different mRNA-LNP formulations (n = 24) were developed using an I-optimal design, where machine learning tools (XGBoost/Bayesian optimization and self-validated ensemble (SVEM)) were used to optimize the process and predict lipid mix ratio. The investigation included material attributes, their respective ratios, and process attributes. The critical responses like particle size (PS), polydispersity index (PDI), Zeta potential, pKa, heat trend cycle, encapsulation efficiency (EE), recovery ratio, and encapsulated mRNA were evaluated. Overall prediction of SVEM (>97%) was comparably better than that of XGBoost/Bayesian optimization (>94%). Moreover, in actual experimental outcomes, SVEM prediction is close to the actual data as confirmed by the experimental PS (94–96 nm) is close to the predicted one (95–97 nm). The other parameters including PDI and EE were also close to the actual experimental data.

© 2024 The Authors. Published by Elsevier B.V. on behalf of Xi'an Jiaotong University. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

The coronavirus pandemic has prompted rapid development of nucleic acid-based vaccines, which have been expanded to cancer, chronic disease therapeutics, protein/enzyme replacement therapy, passive immunization, vaccines of influenza, malaria, and herpes simplex, and cellular reprogramming [1–4]. Breakthroughs in the medical application of messenger RNA (mRNA) were made using microfluidic techniques, and lipid nanoparticles (LNPs) have emerged as the leading technology for nucleic acid delivery [5,6]. LNPs are typically composed of ionizable lipids, phospholipids, cholesterol, and lipid-anchored polyethylene glycol (PEG)ylated lipid, with ionizable lipids being the most important for protein expression and delivery [7,8]. The United States Food and Drug Administration (USFDA) states that particle size (PS) and size distribution are "critical quality attributes (CQAs)" for liposomal drug products [9]. Therefore, advanced process control of the critical factors is essential for developing robust continuous biologics manufacturing [10].

The production of LNPs consists of mainly two steps: an upstream process to prepare LNPs and a downstream process to remove an organic solvent (ethyl alcohol) and exchange buffers [11]. Microfluidic devices, which are easy to scale up, have high reproducibility, and can be used on an industrial scale, are mainly used in the upstream process [11]. Various factors, such as pH, ethanol content, and flow rate, must be evaluated to optimize the manufacturing conditions [12]. The downstream process mainly involves solvent removal and buffer exchange via dialysis. It has been reported that the fusion of LNPs increases the PS, which needs to be controlled to minimize the phenomenon. A previous study showed that the PS can be stabilized when dialysis is performed in two steps [13]. Moreover, lyophilized mRNA-LNPs have a simple design, quick manufacturing, enhanced safety, high efficacy, and improved thermostability compared to the previous studies [14–16].

Recently, several studies have reported that LNP ingredient factors and manufacturing conditions can be optimized using tools such as design of experiment (DOE), partial least squares (PLS), principal component analysis (PCA), XGBoost, Bayesian optimization, and the self-validated ensemble model (SVEM) [17–21]. It is still challenging to obtain accurate prediction with machine learning (ML) tools partly due to small DOE experimental size, which may prevent from partitioning data into training and

* Corresponding author.
E-mail address: shjeong@dongguk.edu (S.H. Jeong).

https://doi.org/10.1016/j.jpha.2024.100996
2095-1779/© 2024 The Authors. Published by Elsevier B.V. on behalf of Xi'an Jiaotong University. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

validation sets. A better strategy would be necessary to improve out-of-sample prediction from other ML tools [22]. SVEM model is often used as a predictive approach to DOE with small dataset and to accelerate time-to-market while reducing cost. Moreover, it is possible to virtually screen RNA therapeutics with improved stability [23]. The SVEM model allows the scientist to use the same design for model validation. A prior study created SVEM model with 16-run definitive screening design (DSD) to reliably predict the behavior of chosen response as applied to 28-run central composite design [24].

Although recent advances in ML tools have led to the incorporation of predictive modeling on experimental data, subsequently generated models have well-documented issues when implemented in practice [25]. Typically, the predictability of such models improves when the validation datasets are considered [26]. However, in many cases, the size of the experimental datasets is small due to some limitations such as resource constraints. Furthermore, the size will be reduced when a portion of the dataset is retained for conducting model validation. On a number of occasions, the dataset size may lead to a different relationship between predictors and responses, inestimable model terms, and in some cases, even poor predictive performance [26,27]. Specifically, if the objective of the experimenter is to determine which factors affect the responses or simply opt to screen the observed experimental datasets, the high-order polynomial or interaction terms may be insignificant. In addition, a simple and reduced model often generates a high prediction error compared to complex models [28]. In another circumstance where the number of factors exceeds the number of observations, although it is not a standard practice to fit in a supersaturated model, similar cases can be frequently observed while conducting predictive modeling runs.

In the present study, I-optimal design is used as a DOE method to obtain simple operational conditions. ML tools are used to optimize and predict the manufacturing process conditions and lipid mixture ratio of mRNA-LNP formulations at the same time. The tools are beneficial when dealing with a small experimental dataset, which is difficult for grouping [29]. Moreover, SVEM, a ML tool, is simpler compared to XGBoost/Bayesian optimization as it does not need separate training and validation datasets. For the optimization, DOE runs are split into training and validation sets with first 22 runs and later 2 runs, respectively. Other 2 independent experiments are performed to test the SVEM prediction accuracy.

2. Materials and methods

2.1. Materials

The following are obtained from Avanti Polar Lipids, Inc. (Alabaster, AL, USA): 1,2-di-*O*-octadecenyl-3-triMAP(-methylammonium propane) (DOTMA; 104872-42-6), 1,2-dioleoyl-3-triMAP (DOTAP; 132172-61-3), 1,2-dioleoyl-3-diMAP (DODAP; 127512-29-2), 3β-hydroxy-5-cholestene (cholesterol, mentioned as 'chol' hereafter; 57-88-5), 1,2-Distearoyl-*sn*-glycero-3-(PC-phosphocholine) (DSPC; 816-94-4), 1,2-dioleoyl-*sn*-glycero-3-PC (DOPC; 4235-95-4), 1,2-dimyristoyl-*sn*-glycero-3-PC (DMPC; 18194-24-6), 1,2-distearoyl-*sn*-glycero-3-(PE-phosphoethanolamine) (DSPE; 1069-79-0); 1,2-dioleoyl-*sn*-glycero-3-PE (DOPE; 4004-05-1), 1,2-Dilauroyl-*sn*-glycero-3-PE (DLPE; 42436-56-6); 1,2-distearoyl-rac-glycerol-(PEG2K-polyethylene glycol-2K) (DSG; 308805-39-2); 1,2-dimyristoyl-rac-*glycero*-3-methoxyPEG2K (DMG; 160743-62-4), 1,2-dipalmitoyl-*sn*-glycero-3-PE-N-amino(PEG2K (DPPE; 2342575-85-1), and N-palmitoyl-sphingosine-1-succinyl(methoxyPEG2K) (Ceramide; 212116-78-4). The (6Z,9Z,28Z,31Z)-heptatriacont-6,9,28,31-tetraene-19-yl-4-(dimethylamino) butanoate (Dlin-MC3-DMA, mentioned as 'MC3' hereafter; 1224606-06-7) is obtained from Xiamen Sinopeg Biotech Co., Ltd. (Xiamen, China). Citric acid, sodium citrate, and sucrose are obtained from Sigma-Aldrich Co. (St. Louis, MO, USA). The first sample of mRNA (CKD Bio-P1) is supplied by Enzynomics (Daejeon, Korea), and the second sample of mRNA is synthesized using an mMESSAGE mMACHINE™ T7 transcription kit (ThermoFisher Scientific Inc., MA, USA). Phosphate buffer saline (PBS (10×), pH 7.4) is obtained from Gibco Inc. (Waltham, MT, USA); Tris(hydroxymethyl)aminomethane (TRIS Ultrapure) is obtained from Duchefa Biochemie B.V (Haarlem, The Netherlands); ethanol (>99.5%) is obtained from Daejung Chemicals & Metals Co., Ltd. (Siheung, Korea); and water (DEPC-Diethyl pyrocarbonate) is provided by IntronBio Inc. (Burlington, MA, USA).

2.2. In vitro transcription of mRNA

Based on the manufacturer's protocol, nanoluciferase encoding mRNA (Nluc-4895) is prepared from linearized pDNA by *in vitro* transcription using a mMESSAGE mMACHINE™ T7 toolkit. Nluc-4895 is purified using a MEGAclear™ transcription clean-up kit and quantified using absorbance, as illustrated in the manufacturer's protocol, before storing at −80 °C until use. The reaction rate (r) of *in vitro* transcription of mRNA is explained by Eq. (1) [20].

$$r = \frac{act_{pH}.act_{temp}.act_{MgCl_2}.r_{max}}{1+\sum_{j=1}^{N}\frac{K_{M,NTP,j}}{c_{NTP,j}}.\left(1+\frac{cPP_i}{K_{I,PPi}}+\sum_{i=1}^{N}\frac{c_{NTP,i}}{K_{I,PPi}}\right)+\frac{K_{M,DNA}}{c_{DNA}}.\left[1+\frac{K_G}{c_{GTP}}.\left(1+\frac{cPPi}{K_{I,PPi}}+\sum_{i=1}^{N-1}\frac{c_{NTP,i}}{K_{I,NTP,i}}\right)\right].\frac{c_{cap}}{c_{cap}+K_{M,cap}}.\frac{c_{mRNA}}{c_{mRNA}+K_{M,mRNA}}} \quad (1)$$

where Eq. (1) illustrates the effect of the concentration of nucleotide (c_{NTP}), pyrophosphate (c_{PPi}), promoter (c_{DNA}), cap analog (c_{cap}), mRNA (c_{mRNA}), and inhibition from nucleoside ($K_{I,NTP}$). Eq. (1) includes Monod constant of nucleotide ($K_{M,NTP}$), promoter ($K_{M,DNA}$), cap analog ($K_{M,cap}$), and mRNA ($K_{M,mRNA}$). The denominator terms within the square brackets describe the initiation process of in vitro transcription.

2.3. Experimental design

(a) The following four material attributes are considered as input variables to evaluate eight responses (PS, polydispersity index (PDI), Zeta potential (ZP), pKa, heat trend cycle, encapsulation efficiency (EE), recovery ratio, and encapsulated mRNA (Table 1)): ionizable lipid type (DOTAP, DOTMA, DODAP, MC3), phospholipid type (DSPC, DOPC, DMPC, DSPE, DOPE, DLPE), PEGylated lipid type

Table 1
Types of material attributes, material ratios, and processing conditions considered for the I-optimal experimental design.

Types	Parameters	Factors (independent variables)
Material attributes	Ionizable	DOTAP, DOTMA, DODAP, MC3
	Phospholipid	DSPC, DOPC, DMPC, DSPE, DOPE, DLPE
	PEGylated lipid	DMG, DSG, DPPE, ceramide
	PEGylated lipid amount (%)	1.5, 1.6, 2.5
Material ratio	Phospholipid-to-PEGylated lipid ratio	3.76, 4.00, 5.88, 6.25, 6.27, 6.67
	Ionizable-to-chol ratio	1.0818, 1.0843, 1.1077, 1.2987, 1.3021, 1.3333
	N/P ratio	6, 8, 10
Processing condition	FRR (mL/min)	3, 4, 5
	TFR (mL/min)	12, 15, 20

DOTAP: 1,2-dioleoyl-3-triMAP; DOTMA: 1,2-di-*O*-octadecenyl-3-triMAP(-methylammonium propane); DODAP: 1,2-dioleoyl-3-diMAP; MC3: (6Z,9Z,28Z,31Z)-heptatriacont-6,9,28,31-tetraene-19-yl-4-(dimethylamino) butanoate; DSPC: 1,2-Distearoyl-*sn*-glycero-3-(PC-phosphocholine); DOPC: 1,2-dioleoyl-*sn*-glycero-3-PC; DMPC: 1,2-dimyristoyl-*sn*-glycero-3-PC; DSPE: 1,2-distearoyl-*sn*-glycero-3-(PE-phosphoethanolamine); DOPE: 1,2-dioleoyl-*sn*-glycero-3-PE; DLPE: 1,2-Dilauroyl-*sn*-glycero-3-PE; DSG: 1,2-distearoyl-rac-glycerol-(PEG2K-polyethylene glycol-2K); DMG: 1,2-dimyristoyl-rac-*glycero*-3-methoxyPEG2K; DPPE: 1,2-dipalmitoyl-*sn*-glycero-3-PE-N-amino(PEG2K; Ceramide: N-palmitoyl-sphingosine-1-succinyl(methoxyPEG2K); N/P: nitrogen from lipid and phosphate from nucleic acid; FRR: flow rate ratio; TFR: total flow rate.

(DMG, DSG, DPPE, ceramide), PEGylated lipid amount (1.5%−2.5%), and nitrogen from lipid and phosphate from nucleic acid (N/P) ratio (6−10); (b) two material ratios: phospholipid-to-PEGylated lipid (3.76−6.67) and ionizable-to-chol (1.0818−1.3333); (c) two processing conditions: aqueous-to-organic phase ratio (flow rate ratio (FRR), 3 to 5) and production speed (total flow rate (TFR), 12−20 mL/min); and (d) probable ionizable/phospholipid/chol/PEGylated lipid combination (mol%). The 24 mRNA-LNP experimental runs are designed using an I-optimal DOE (Table 2), in which the main effects are not confounded with two-factor interactions, while non-linear correlations can be detected [21]. The full model includes two- and three-way interactions, quadratic and partial cubic terms for process factors, and Scheffe cubic terms for mixture factors, which add quadratic effects for the continuous process factors and their quadratic effects with other non-mixture factors. The non-linear quadratic model, as illustrated by Eq. (2), is obtained using JMP® Pro version 17 (SAS Institute Inc., Cary, NC, USA) [30].

$$Y = \beta_0 + \sum_{i=1}^{n} \beta_i X_i + \sum_{i=1}^{n-1} \sum_{i<j} \sum_{j=2}^{n} \beta_{ij} X_i X_j + \sum_{i=1}^{n} \beta_{ii} X_i^2 + \varepsilon \quad (2)$$

where X_1, X_2, and X_n are factors; β_0 is constant; β_i, β_{ij}, and β_{ii} are regression coefficients for linear, interaction, and quadratic terms, respectively. Moreover, n, Y, and ε are the number of variables, response, and experimental error, respectively.

2.4. Preparation of mRNA-LNPs using microfluidics

Each lipid is individually dissolved in ethanol. Subsequently, the amount of each lipid is taken and constituted at 6 M ratios (ionizable/phospholipid/chol/PEGylated lipid) (Table 3), which gives ionizable-

Table 2
I-optimal experimental design for the manufacture of messenger RNA-lipid nanoparticles (mRNA-LNP) with three material attributes, a PEGylated lipid amount, three material ratios, and two processing conditions.

S. No.	Factors (independent variables)								
	Ionizable lipid	Phospholipid	PEGylated lipid	PEGylated lipid (%)	Ionizable-to-chol ratio	Phospholipid-to- PEGylated lipid ratio	N/P ratio	FRR (mL/min)	TFR (mL/min)
1	DOTAP	DOPE	DSG	2.5	1.1077	3.76	10	4	15
2	DOTAP	DSPE	DMG	1.6	1.0843	5.88	8	5	12
3	DOTAP	DMPC	DPPE	1.6	1.3021	6.25	10	3	15
4	DOTAP	DOPC	Ceramide	1.5	1.2987	6.67	6	3	12
5	DOTAP	DSPC	Ceramide	1.6	1.3021	6.25	6	5	20
6	DOTMA	DLPE	DMG	1.6	1.3021	6.25	10	4	20
7	DOTMA	DOPE	DPPE	1.5	1.2987	6.67	8	5	20
8	DOTMA	DSPE	DPPE	1.6	1.3021	6.25	6	4	15
9	DOTMA	DMPC	DSG	1.5	1.2987	6.67	6	3	12
10	DOTMA	DOPC	DSG	1.6	1.0843	5.88	10	5	15
11	DOTMA	DSPC	Ceramide	2.5	1.3333	4.00	8	3	15
12	DODAP	DLPE	DPPE	2.5	1.1077	3.76	6	5	12
13	DODAP	DOPE	DMG	1.6	1.0843	5.88	6	3	15
14	DODAP	DSPE	Ceramide	1.5	1.2987	6.67	10	5	15
15	DODAP	DMPC	DMG	2.5	1.3333	4.00	8	5	15
16	DODAP	DOPC	DMG	2.5	1.3333	4.00	10	3	20
17	DODAP	DSPC	DSG	1.6	1.3021	6.25	8	4	12
18	MC3	DLPE	DSG	1.6	1.3021	6.25	8	3	15
19	MC3	DOPE	Ceramide	2.5	1.3333	4.00	10	5	12
20	MC3	DSPE	DSG	2.5	1.3333	4.00	6	3	20
21	MC3	DMPC	Ceramide	1.6	1.0843	5.88	8	4	20
22	MC3	DOPC	DPPE	1.6	1.3021	6.25	8	4	15
23	MC3	DSPC	DPPE	1.5	1.0818	6.27	10	3	12
24	MC3	DSPC	DMG	1.5	1.2987	6.67	6	5	15

DOTAP: 1,2-dioleoyl-3-triMAP; DOTMA: 1,2-di-*O*-octadecenyl-3-triMAP(-methylammonium propane); DODAP: 1,2-dioleoyl-3-diMAP; MC3: (6Z,9Z,28Z,31Z)-heptatriacont-6,9,28,31-tetraene-19-yl-4-(dimethylamino) butanoate; DSPC: 1,2-Distearoyl-*sn*-glycero-3-(PC-phosphocholine); DOPC: 1,2-dioleoyl-*sn*-glycero-3-PC; DMPC: 1,2-dimyristoyl-*sn*-glycero-3-PC; DSPE: 1,2-distearoyl-*sn*-glycero-3-(PE-phosphoethanolamine); DOPE: 1,2-dioleoyl-*sn*-glycero-3-PE; DLPE: 1,2-Dilauroyl-*sn*-glycero-3-PE; DSG: 1,2-distearoyl-rac-glycerol-(PEG2K-polyethylene glycol-2K); DMG: 1,2-dimyristoyl-rac-*glycero*-3-methoxyPEG2K; DPPE: 1,2-dipalmitoyl-*sn*-glycero-3-PE-N-amino(-PEG2K; Ceramide: N-palmitoyl-sphingosine-1-succinyl(methoxyPEG2K); N/P: nitrogen from lipid and phosphate from nucleic acid; FRR: flow rate ratio; TFR: total flow rate.

Table 3
Different proportions of ionizable, phospholipid, chol, and PEGylated lipids were used to develop I-optimal experimental messenger RNA-lipid nanoparticles (mRNA-LNP) formulations and the corresponding Ionizable-to-chol ratio and phospholipid-to-PEGylated lipid ratio.

S. No.	Ionizable	Phospholipid	Chol	PEGylated lipid (%)	Ionizable-to-chol ratio	Phospholipid-to-PEGylated lipid ratio
1	46.3	9.4	41.8	2.5	1.1077	3.76
2	46.3	9.4	42.7	1.6	1.0843	5.88
3	50	10	38.5	1.5	1.2987	6.67
4	46.3	8.4	42.8	2.5	1.0818	3.36
5	50	10	37.5	2.5	1.3333	4.00
6	50	10	38.4	1.6	1.3021	6.25

Chol: 1,2-dioleoyl-3-diMAP (DODAP, 127512-29-2), 3β-hydroxy-5-cholestene.

to-chol ratios of 1.0818, 1.0843, 1.1077, 1.2987, 1.3021, and 1.3333, and phospholipid-to-PEGylated lipid ratios of 3.36, 3.76, 4.00, 5.88, 6.25, and 6.67. Aqueous samples of mRNA (CKD Bio-P1 and Nluc-4895) are prepared in 100 mM citrate buffer (pH 4.5) at N/P ratios of 6, 8, or 10. The initial and waste amounts are set at 0.45 and 0.05 mL, respectively. The mRNA and lipid mix are infused in a toroidal micromixer cartridge (TrM) of a NanoAssemblr® Benchtop (Precision NanoSystems Inc., Vancouver, BC, Canada) [31]. The reaction rate of the microfluidics process is calculated from Eq. (3), while the pH value of the final mRNA-LNP formulation is calculated using Eq. (4), which is a Henderson-Hasselbalch equation where the pH of buffer systems is obtained with negative logarithm of acid (c_{HA}) and conjugated base concentration ($c_{A\text{-}}$) with acid constant (pKa) [20].

$$r = k.c_{mRNA}.c_{ionizable}.c_{phospholipid}.c_{Chol}.c_{PEG} \tag{3}$$

$$pH = pK_a + \log\left(\frac{c_{A-}}{c_{HA}}\right) \tag{4}$$

LNP formulations are dialyzed (MWCO 10 kDa, Thermo Scientific, Rockford, IL, USA) against 100 times the LNP volume of 10 mM Tris buffer (pH 7.4) for 24 h under magnetic stirring. The buffer is exchanged with a fresh Tris buffer at the 2nd and 4th h. The LNP formulations are concentrated using a centrifuge (Eppendorf Manufacturing Co., Hamburg, Germany) at 5000 *g* for 30 min with Amicon® Ultra (MWCO 100 kD, Sigma-Aldrich) and filtered with a PALL capsule (MWCO 300 kD) in a Tangential Flow Filtration system (TFF; Minimate™ PALL Life Sciences, Ann Arbor, MI, USA). To ensure a nuclease-free environment, the glassware is sterilized at a high temperature (≥300 °C) for at least 12 h. All other equipment are treated with RNaseZap™ (Thermo Scientific) before use.

2.5. Particle size, PDI, and Zeta potential

Dynamic light scattering (DLS; Zetasizer Nano ZS, Malvern Instruments Ltd., Worchestershire, UK) equipped with a 633 nm He−Ne laser and a detection angle of 90° is used to measure the PS, PDI, and ZP. The mRNA-LNPs are initially diluted with water (DEPC-treated) to a concentration of 100 μg/mL and allowed to stabilize for 30 min before measurements [26]. DLS measures Z-average hydrodynamic size (D_z, Eq. (5)) and broadness of the size distribution (PDI) of particles dispersed in a liquid medium based on Stokes-Einstein equation [32].

$$D_z = \frac{k_B T}{3\pi\eta D_{t,avg}} \tag{5}$$

where, D_z, k_B, T, η, and $D_{t,avg}$ are hydrodynamic diameter, Boltzmann's constant, thermodynamic temperature, dynamic viscosity, and translational diffusion coefficient, respectively.

ZP is measured using a laser-Doppler velocimetry method. The dispersant refractive index (RI) and viscosity are 1.33 and 0.8872 cP, respectively, and the material absorbance and refractive index (RI) are 0.01 and 1.49, respectively. Analysis is performed three times for each sample, and mean values are taken (n = 10, (3 × 10) 30 runs in each measurement).

2.6. pKa measurement

The mRNA-LNP sample is diluted to 6.25 ng/μL total mRNA in Tris buffer and transferred to a quartz cuvette (ZEN2112) to measure the size in a Zetasizer Nano ZS with configuration set at absorption of 0.001 at 25 °C, viscosity of 0.888 cP, and RI of 1.335. Previously, the instrument is equilibrated to 25 °C for 30 s, with no delay between measurements. The observed data are fitted to the extended Henderson−Hasselbalch equation, which is represented by Eq. (6), to obtain the pKa [33].

$$\Psi = \Psi_{max} - \frac{\Psi_{max} - \Psi_{min}}{1 + 10^{\frac{pKa - pH}{n}}} \tag{6}$$

where extreme values of pH limits are maximum and minimum Ψ_{max} and Ψ_{min}, respectively. The apparent pKa is calculated as the pH corresponding to 50% LNP protonation, which is fitted using a nonlinear sigmoidal dose−response.

2.7. Heat trend cycle

The heat trend cycle of the mRNA-LNP formulations is determined using a Q-2000 differential scanning calorimetry (DSC) (TA Instruments, New Castle, DE, USA) [34]. Each mRNA-LNP formulation sample (~2 mL) is pipetted into T_{zero} aluminum pan and sealed, with a blank pan employed as a reference. DSC measurements are performed at a scan rate of 10 °C/min from 20 °C to −50 °C under a nitrogen flow of 50 mL/min. The thermal parameters are interpreted using software provided with the instrument. Thermal events during the heat trend cycle were assumed to represent buffer compositions. They often have a broad exothermic hump associated with the eutectic crystallization of salts from the solutions [34].

2.8. Encapsulation efficiency (EE)

A Quant-iT™ RiboGreen® RNA Assay Kit (Invitrogen™, Thermo Fisher Scientific) is used to quantify the amount of encapsulated mRNA and its concentration as a percentage [12]. To determine the un-encapsulated mRNA, the sample is diluted to the expected total mRNA concentration using 1 × Tris-ethylenediaminetetraacetic acid (EDTA) (TE) buffer, and RiboGreen® dye (100 μL) was added and incubated for 5 min. To determine the encapsulated mRNA, the

sample is diluted to the expected total mRNA concentration using 1 × TE buffer containing 0.5% (*V/V*) Triton X-100 (Sigma-Aldrich), incubated for 5 min, measured at fluorescence wavelength $\lambda_{em} = 535$ nm and $\lambda_{ex} = 485$ nm using a UV–Vis plate reader (TECAN Spark, Morrisville, NC, USA), and plotted against a calibration curve, prepared up to 2 mg/mL ($R^2 \geq 0.9999$) [35]. The procedure is based on the supplier's protocol. The limit of detection (LOD) and limit of quantification (LOQ) are calculated for mRNA (8 and 30 μg/mL, respectively). The analyses are performed between the LOQ and 2 mg/mL.

2.9. Recovery ratio

The theoretical weight of the mRNA-LNP sample is noted. After the sample is prepared from the microfluidics, the final weight is calculated. The recovery ratio is calculated by deducting the final weight from the theoretical weight and expressed in percent, as given in Eq. (7).

$$Recovery\ ratio = \frac{theoretical\ weight - observed\ weight}{theoretical\ weight} \times 100\% \tag{7}$$

2.10. Encapsulated mRNA

Encapsulated mRNA is the total amount of mRNA covered by the lipid mixture, which is available in the recovered final weight after microfluidics preparation and expressed as a percentage using Eq. (8).

$$Encapsulated\ mRNA = recovery\ ratio \times encapsulation\ efficiency \tag{8}$$

2.11. XGBoost model

The XGBoost model is an optimized distributed gradient boosting library, which is both efficient and flexible [36]. Eight output responses are used to build an individual model using 13 input factors. Twenty-four formulations are used to construct the model. The learning rate and number of rounds are set within (0.05 and 0.3) and (5 and 30), respectively. The importance of each factor is calculated, and the critical process parameters are determined on the basis of the high importance of the values. Different number of estimators and regularization strengths are evaluated [37].

2.12. Bayesian optimization

The Bayesian optimization model is implemented to optimize the experimental factors based on previous outcomes obtained in terms of probability distributions [36]. A prior study suggested that machine learning approaches, such as artificial neural networks, were capable of tuning the PS [38]. The benefit of the present XGBoost and Bayesian optimization models is that they can promptly optimize the factors with a few experiments. The XGBoost model is used to determine the process factors of the microfluidic instrument, which can be further implemented to obtain the desired output responses with Bayesian optimization.

2.13. SVEM

The SVEM model is a type of bootstrapping method with fractionally weighted bootstrapping. Every time a different fractional weight is assigned to each data point while maintaining the design structure [19]. During the model fitting, the SVEM assigns each data point to either training or self-validation as 1 or 0, respectively. In the case of validation, the weights are assigned reversely. For self-validation purposes, the original dataset is considered as a training set, while a copy of the original dataset is generated and allocated as a "self-validation set." The fractional weights are assigned in pairs such that a large weight in a training set will balance with a small weight in a self-validation set, as illustrated in Eq. (9). This type of correlation allows the model to function as a de facto validation set. The system requirements recommended to run SVEM model in the JMP® Pro version 17 software (SAS Institute Inc., Cary, NC, USA) are 64-bit CPU, Windows 10 or 11 (except Windows 10S edition), and 4 GB or more RAM [39–41].

$$\begin{aligned} &Generate: u_i \sim U[0,1] \quad for \quad i = 1 \ldots N \\ &Training\ FW: w_{T,i} = F^{-1}(u_i) = \log(u_i) \quad for \quad i = 1 \ldots N \\ &Self-validation\ FW: w_{V,i} = F^{-1}(1-u_i) = \log(1-u_i) \\ &for \quad i = 1 \ldots N \end{aligned} \tag{9}$$

First, N random uniform weights (0,1) are generated, followed by an inverse probability distribution to obtain fractional weights, as illustrated in Eq. (9). U [0,1] represents uniform distribution and F indicates cumulative distribution function.

3. Results and discussion

3.1. Evaluation of the effects of different factors on outcomes

Twenty-four formulations are designed with nine factors (ionizable lipid, phospholipid, PEGylated lipid, ionizable-to-chol, PEGylated lipid amount, phospholipid-to-PEGylated lipid, FRR, TFR, and N/P ratio) using I-optimal DOE (Table 2), and the subsequent eight outcomes (PS, PDI, ZP, pKa, heat trend cycle, EE, recovery ratio, encapsulated mRNA) are presented in Table 4. The dataset distribution of the output responses is illustrated in Fig. 1, where the means of PS, PDI, ZP, pKa, heat trend cycle, EE, recovery ratio, and encapsulated mRNA are 66.84 ± 15.25 nm, 0.17 ± 0.09, 1.18 ± 0.83 mV, 6.28 ± 0.44, −16.59 ± 3.91 °C, 88.66% ± 6.47%, 89.54% ± 3.11%, and 79.39% ± 6.50%, respectively.

The outcomes are further analyzed based on the specific input factors, as illustrated in Fig. 2. It is found that formulation with DOTAP (ionizable lipid) gives a comparably larger PS, ZP (Fig. 2A), and EE (Fig. 2B), whereas PDI and pKa are smaller compared to the formulations using other ionizable lipids. A prior study found that mRNA-LNPs with a larger PS can induce transgene expression in mouse immune cells and demonstrate superior effects as a therapeutic cancer vaccine [42]. Additionally, design of LNP by modulating lipid is particularly essential for organ-specific, cell-specific delivery, biodegradability, and multi-functionality [43]. The outcomes of ZP and EE using DOTAP ionizable lipids are consistent with those of the prior study [26]. The ionizable lipids incorporated in the mRNA-LNPs prevent hydrolysis and colloidal destabilization when lyophilized in the presence of 20% (*w/V*) sucrose [44]. On the contrary, the formulations with MC3 lipids resulted in a smaller PS, ZP, and EE, but higher PDI and pKa values. A prior study found that PS and ZP were crucial and directly impacted *in vivo* behavior and cellular internalization [45]. Particularly, 50–100 nm LNP are approved for mRNA-LNP vaccine delivery in a clinical setting [46]. Among the PEGylated lipid types, PDI is higher with the formulations with DSG (Fig. 2C), whereas the formulations with ceramide have higher ZP. Among the phospholipids, the formulations prepared with that of DMPC and DOPE have better thermostability (Fig. 2D), whereas the formulations with DSPE have comparably lower stability. A lower N/P ratio favors ZP, whereas a higher N/P ratio favors thermal stability and enhances the recovery ratio (Fig. 2E).

Table 4
Different outcome responses were obtained from 24 messenger RNA-lipid nanoparticles (mRNA-LNP) formulations.

	Responses (Dependent variables)							
S. No.	PS (nm)	PDI	ZP (mV)	pKa	Heat trend cycle (°C)	EE (%)	Recovery ratio (%)	Encapsulated mRNA (%)
1	66.55	0.120	1.34	6.11	−7.04	93.176	90.912	84.708
2	78.57	0.107	1.97	5.66	−15.75	92.391	87.143	80.512
3	96.95	0.071	2.33	6.21	−7.62	96.439	94.508	91.143
4	94.42	0.069	3.09	5.87	−17.93	95.558	96.634	92.342
5	92.45	0.099	3.26	5.59	−18.03	91.141	90.233	82.239
6	75.26	0.212	0.43	5.86	−12.04	93.162	92.322	86.009
7	77.76	0.208	0.87	6.19	−17.91	91.161	88.842	80.989
8	78.82	0.178	0.96	6.04	−19.14	92.526	91.103	84.294
9	82.88	0.258	2.09	5.62	−19.65	92.043	87.392	80.438
10	69.44	0.222	0.80	5.57	−12.79	90.146	93.890	84.638
11	82.11	0.233	1.74	5.83	−16.78	92.571	87.758	81.238
12	51.13	0.087	0.47	6.73	−20.96	91.816	85.555	78.553
13	50.99	0.073	0.97	6.10	−21.11	93.363	87.359	81.561
14	51.89	0.067	0.64	6.59	−14.54	91.361	88.761	81.093
15	57.25	0.203	0.61	6.80	−18.07	90.261	91.000	82.138
16	57.00	0.211	0.59	6.62	−14.63	93.491	90.001	84.143
17	57.13	0.198	1.00	6.57	−18.26	90.138	87.358	78.743
18	55.88	0.320	0.84	6.75	−18.35	79.216	85.499	67.729
19	57.44	0.374	0.16	6.81	−14.03	74.532	94.088	70.126
20	56.43	0.331	1.12	6.49	−21.41	84.168	85.525	71.985
21	51.82	0.149	0.86	6.63	−18.06	75.923	90.140	68.437
22	56.13	0.156	0.84	6.70	−18.75	80.542	88.878	71.584
23	51.08	0.146	0.16	6.84	−13.93	77.349	93.135	72.039
24	54.70	0.090	1.12	6.57	−21.44	85.416	92.508	79.017

PS: particle size; PDI: polydispersity index; ZP: Zeta potential; EE: encapsulation efficiency.

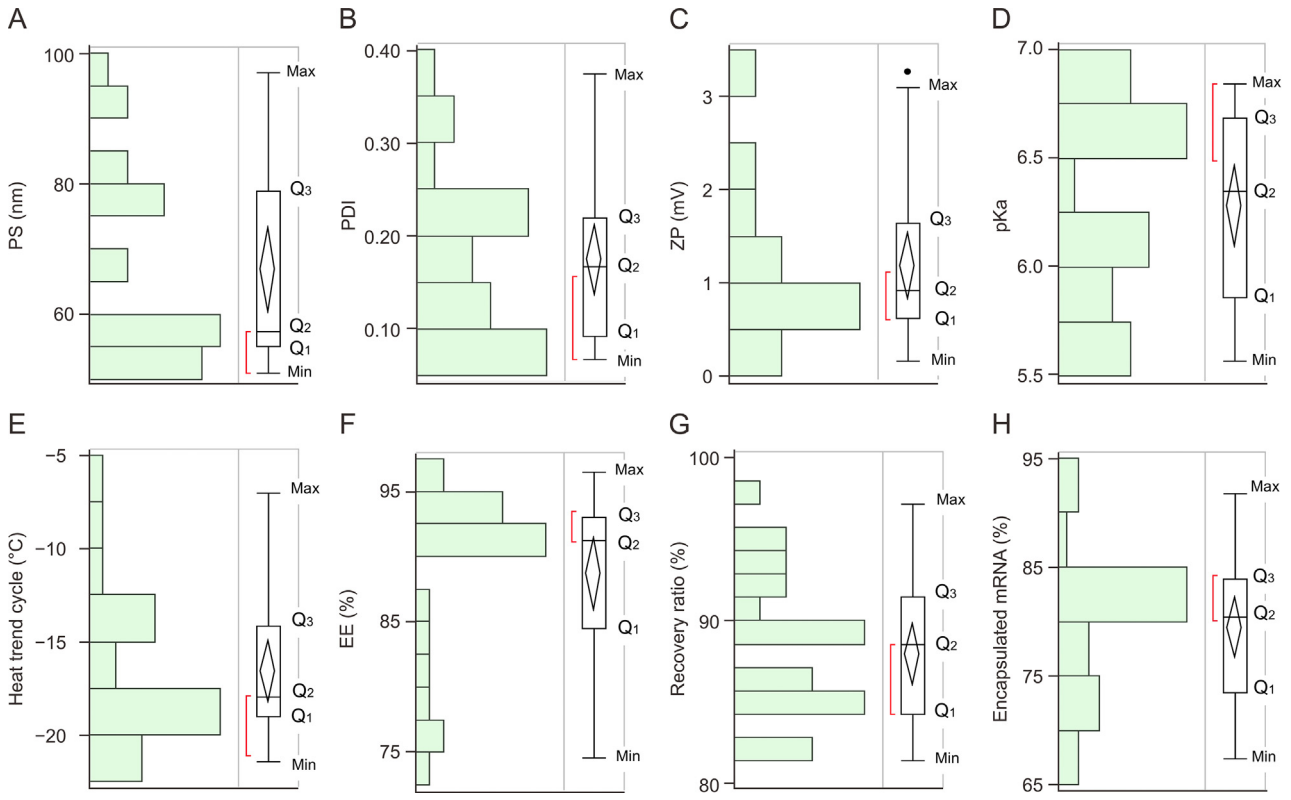

Fig. 1. Distribution of the outcome responses. (A) Particle size (PS), (B) polydispersity index (PDI), (C) Zeta potential (ZP), (D) pKa, (E) heat trend cycle, (F) encapsulation efficiency (EE), (G) recovery ratio, (H) encapsulated mRNA obtained from the experimental runs. To the right side of the histogram, quantile plots are drawn where the Q_1 (25% quartile), median (50% quartile), Q_3 (75% quartile), maximum value, and minimum value are indicated. The red line indicates the mean along with its standard deviation.

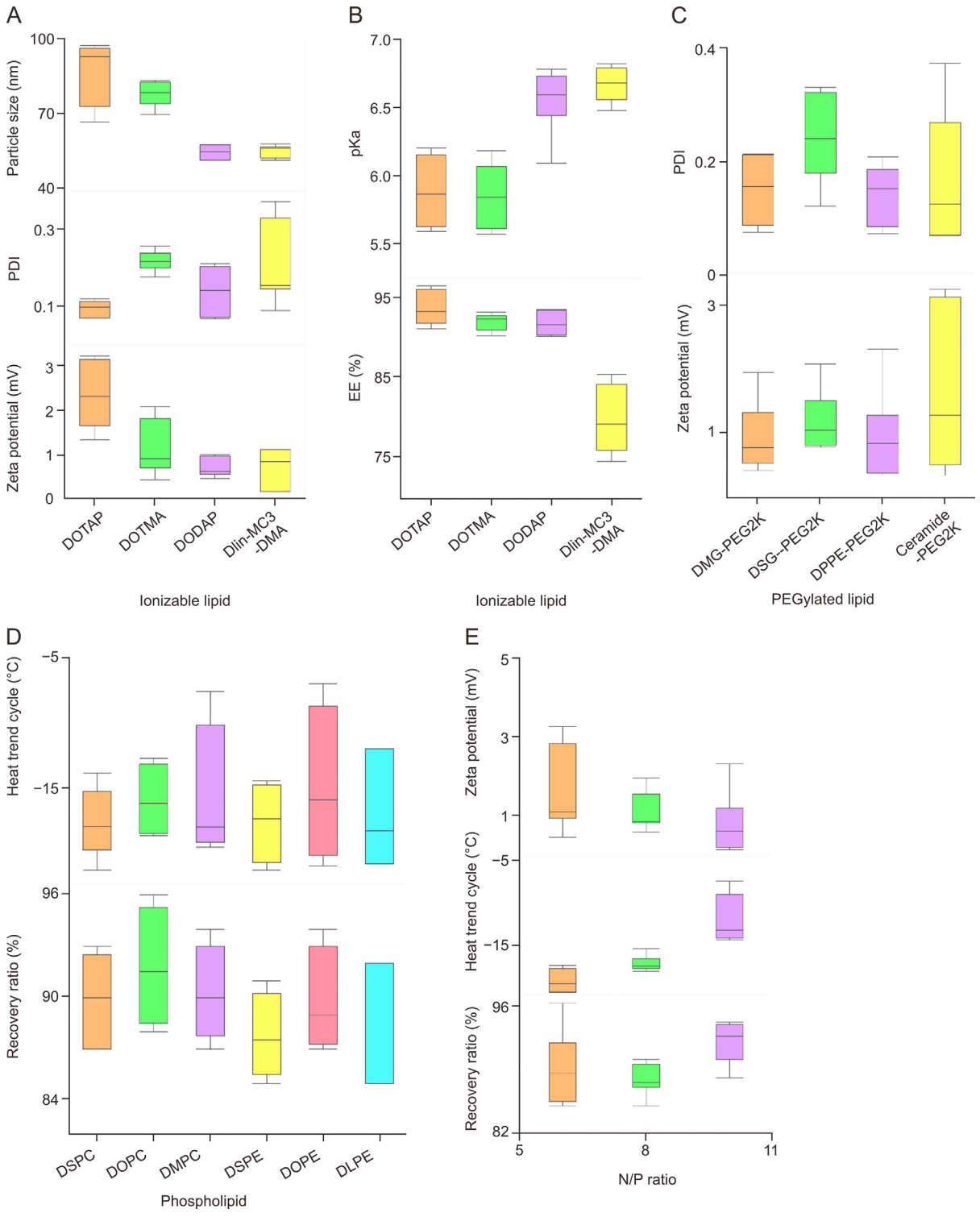

Fig. 2. Effect of factors on the responses of the messenger RNA-lipid nanoparticles (mRNA-LNP) vaccine formulations. (A) Effect of ionizable lipids on particle size, polydispersity index (PDI), and Zeta potential. (B) Effect of ionizable lipids on pKa and encapsulation efficiency (EE). (C) Effect of PEGylated lipid on PDI and Zeta potential. (D) Effect of phospholipid on heat trend cycle and recovery ratio. (E) Effect of nitrogen from lipid and phosphate from nucleic acid (N/P) ratio on Zeta potential, heat trend cycle, and recovery ratio.

3.2. Principal component analysis

The PCA tool is specifically implemented to identify the major responses and reduce the dimension [47]. In our study, the datasets are evaluated using PCA tools, which suggests that [PS & ZP] (0.80) and [EE and encapsulated mRNA] (0.91) have a strong (close to 1) and positive correlation, whereas [PS and pKa] and [ZP and pKa] have a negative correlation, as shown in scatterplot diagram (Fig. 3). The library for the PCA tool is built from different material attributes, processing conditions (Table 1), and experimental outcomes (Table 2) [48]. This approach is particularly beneficial to identify the particles in the protein formulations [49].

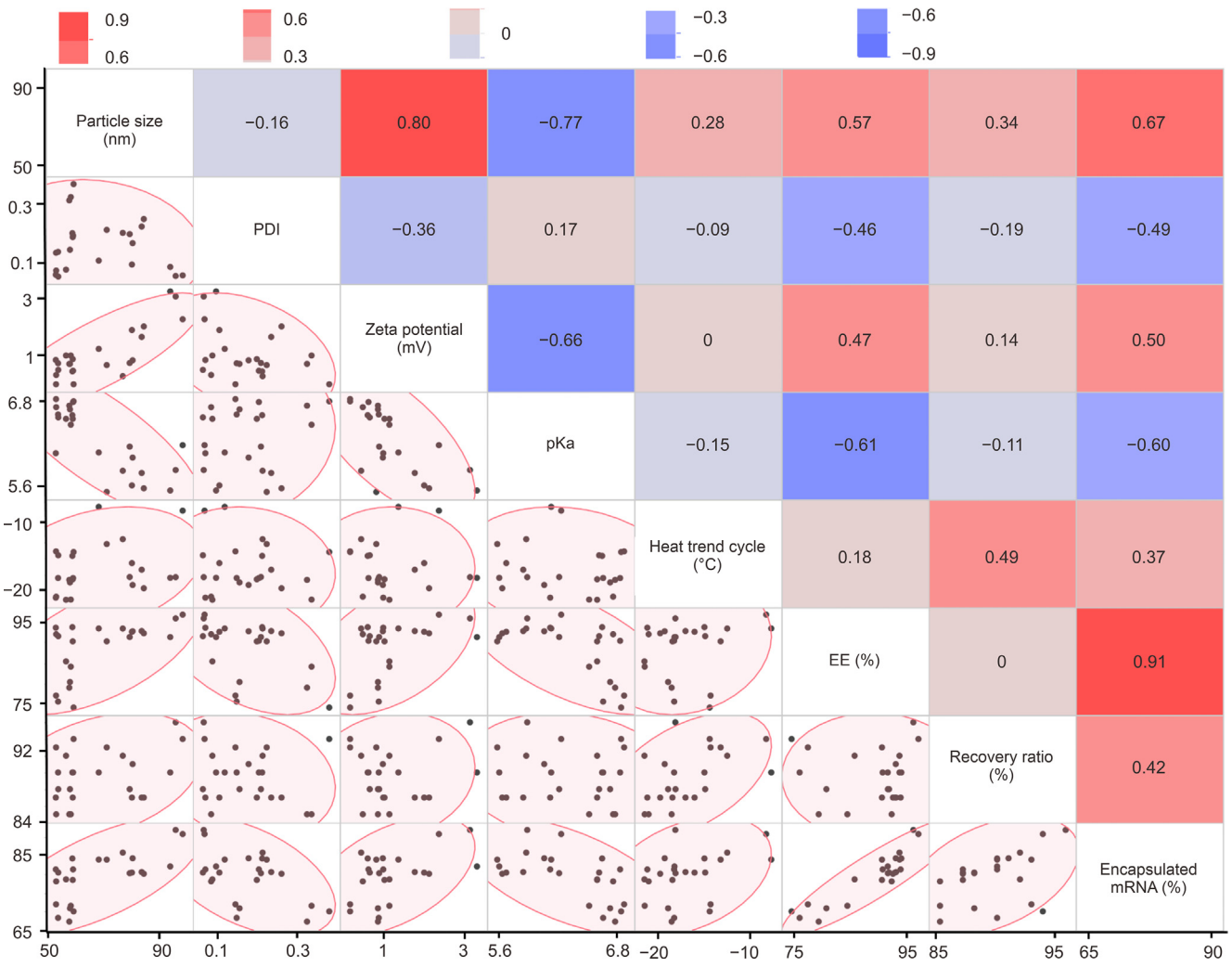

Fig. 3. Scatterplot matrix plotted between different outcome responses. Particle size, polydispersity (PDI), Zeta potential, pKa, heat trend cycle, encapsulation efficiency (EE), recovery ratio, and encapsulated mRNA illustrate the correlation between the responses. The principal component analysis tool from JMP Pro® was used to determine the correlations. Higher intensity of red indicates stronger positive correlation, whereas higher intensity of blue indicates stronger negative correlation.

3.3. Partial least squares (PLS) analysis

The PLS analysis of the response datasets (particle size, PDI, ZP, pKa, heat trend cycle, EE, recovery ratio, and encapsulated mRNA) are provided in Fig. 4, which can interpret the significant factors and influence of the two-factor correlation. Both PS and PDI are positively affected by the ionizable-to-chol ratio and negatively influenced by the PEGylated lipid and phospholipid-to-PEGylated lipid ratios (Fig. 4A). Similarly, ZP is positively affected by the ionizable-to-chol ratio and negatively influenced by the PEGylated lipid, FRR, and N/P ratio (Fig. 4A). Other output responses (pKa, heat trend cycle, EE, recovery ratio, and encapsulated mRNA) are positively influenced by the PEGylated lipid and phospholipid-to-PEGylated lipid ratio (Fig. 4B). Although the ionizable-to-chol ratio appeared to be a significant factor (Fig. 4C), an ionizable lipid has no significant contribution toward EE (Fig. 4B), which is contrary to prior findings [20]. One possible reason for this discrepancy is that the evaluation of the entire dataset library made it difficult to identify significant ionizable lipids from the ionizable-to-chol ratio; thus, the formulations may require optimization as a next step, which is discussed in the subsequent section. In case of the heat trend cycle and recovery ratio, the outcome is also positively affected by the N/P ratio (Fig. 4B). When the variable importance parameter is set at 0.8, the major influencing factors are the ionizable-to-chol ratio (0.9285), PEGylated lipid (0.9942), phospholipid-to-PEGylated lipid (1.0970), and N/P ratio (1.5669, Figs. 4C and D).

3.4. Optimization of the mRNA-LNP formulations

The output responses are crucial for the stability, efficacy, and quality of the mRNA-LNP formulations and thus are termed as critical quality attributes (CQAs), as listed in Table 5. The constraints are set to the CQAs as PS in a range of 40–120 nm, PDI < 0.3, ZP in between 1 and 4 (irrespective of positive or negative charge), pKa in between 5 and 7, thermal stability when temperature rose by more than −25 °C, EE > 75%, recovery ratio > 85%, and encapsulated mRNA > 70% (Table 5). Similar study has been conducted with alphavirus-based self-amplifying RNA (saRNA) LNP as an alternative to mRNA vaccine where DOTAP/DOTMA lipid produce LNP with particle size < 200 nm (PDI < 0.3, EE > 90%). It is slightly large in size which requires optimization. Conversely, low toxicity, dose-sparing property, versatility, and simple manufacturing process facilitate a rapid response against future pandemics [50]. Another study found that antibody titer increased with increasing LNP size until 100 nm, and then started to decrease and confirmed that LNP size in a range of

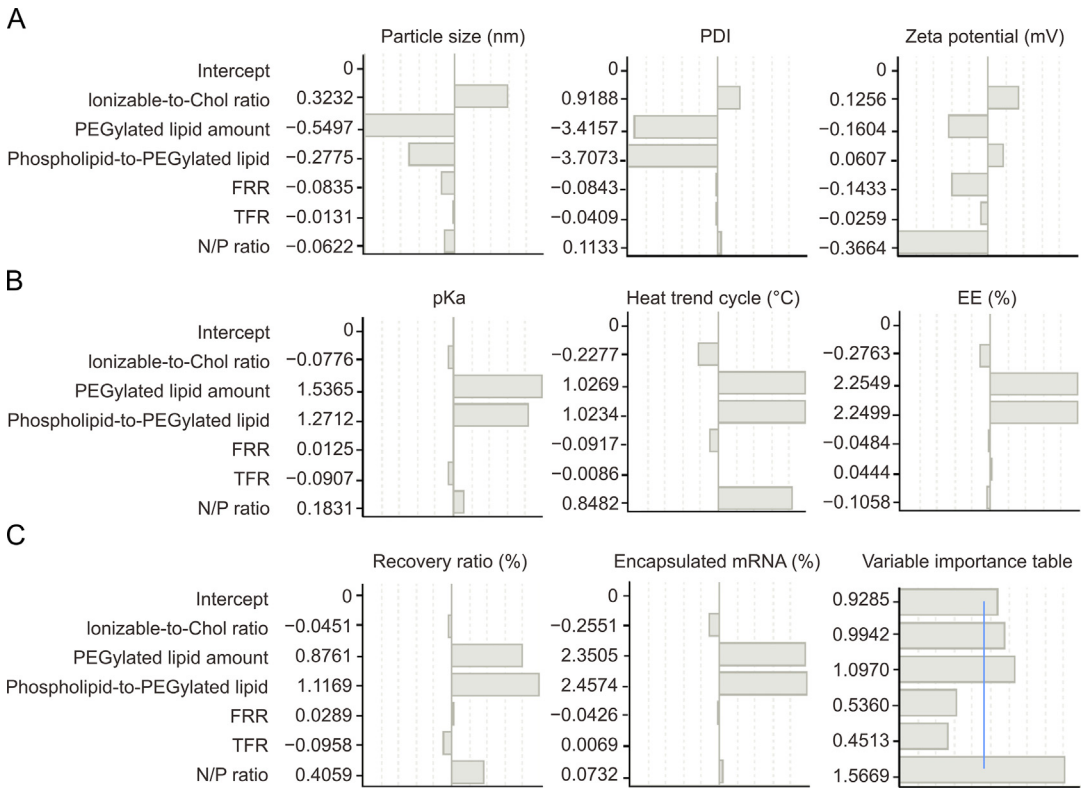

Fig. 4. Partial least squares analysis of the outcome responses. (A) Particle size, polydispersity (PDI), and Zeta potential (ZP). (B) pKa, heat trend cycle, and encapsulation efficiency (EE). (C) Recovery ratio, encapsulated mRNA, and variable importance table.

60–150 nm showed a robust immune response without impacting on immunogenicity in non-human primates [51]. Therefore, particle size in DOE set to 40–120 nm. Based on the constraints, the expected outcomes of PS and pKa are set to "in range," the likely outcome of PDI is set to "minimize," and the desired outcomes of the remaining responses (Fig. 5A): ZP, heat trend cycle, EE, recovery ratio, and encapsulated mRNA, are set to "maximize." Based on these optimization criteria, the graph is plotted against the phospholipid-to-PEGylated lipid vs. the ionizable-to-chol ratio, where the predicted responses are obtained as a PS of 62.91 nm (Fig. 5B), PDI of 0.19 (Fig. 5C), ZP of 4.52 (Fig. 5D), pKa of 6.32 (Fig. 5E), heat trend cycle of −14.63 °C (Fig. 5F), EE of 93.53% (Fig. 5G), recovery ratio of 99.80% (Fig. 5H), and encapsulated mRNA of 96.69%, when the factors are set as a PEGylated lipid amount of 1.58, a FRR of 3.50, a TFR of 19.88, and a N/P ratio of 9.91, with ionizable lipid DODAP, phospholipid DOPE, and PEGylated lipid DMG-PEG2K.

The probable zone of interest that meets the pre-defined criteria is allocated as the design space that falls under the defined constraints (shaded region in yellow). Similarly, as before, one of the mRNA-LNP formulations is chosen within the design space, which gives the predicted responses as a PS of 73.62 nm, PDI of 0.35, ZP of 0.16, pKa of 6.20, heat trend cycle of −21.44 °C, EE of 96.44%, recovery ratio of 96.63%, and encapsulated mRNA of 91.74%, when the factors are set to a PEGylated lipid amount of 2, a FRR of 4, a TFR of 16, and an N/P ratio of 8, with ionizable lipid DOTAP, phospholipid DOPE, and PEGylated lipid DSG-PEG2K (Fig. 5). This work is supported by previous research where the quality by design (QbD) approach has been useful to optimize the mRNA manufacturing process, reduce out-of-specification, and contribute to a continuous manufacturing platform [20].

Table 5
Constraints established on the critical quality attributes (CQAs) of messenger RNA-lipid nanoparticles (mRNA-LNP) formulations along with the desired outcomes of the responses.

Responses	Symbols	Constraints	Desired outcomes
PS	(Y_1)	40 nm $\leq Y_1 \leq$ 120 nm	In range
PDI	(Y_2)	$Y_2 \leq 0.3$	Minimize
ZP	(Y_3)	(±) 1 $\leq Y_3 \leq$ (±) 4	In range
pKa	(Y_4)	$5 \leq Y_1 \leq 7$	In range
Heat trend cycle	(Y_5)	−25 °C $\leq Y_2$	Maximize
EE	(Y_6)	75% $\leq Y_4$	Maximize
Recovery ratio	(Y_7)	85% $\leq Y_4$	Maximize
Encapsulated mRNA	(Y_8)	70% $\leq Y_4$	Maximize

PS: particle size; PDI: polydispersity index; ZP: Zeta potential; EE: encapsulation efficiency.

3.5. Monte Carlo simulation

Based on the design space and optimization, Monte Carlo simulation is performed (n = 10,000) and an R^2 and P value of the predicted dataset are statistically evaluated. The PS (0.9981, 0.0197), PDI (0.9832, 0.0163), ZP (0.9954, 0.0477), pKa (0.9800, 0.0239), heat trend cycle (0.9988, 0.0122), EE (0.9980, 0.0215), recovery ratio (0.9200, 0.0365), and encapsulated mRNA (0.9500, 0.0463) are found to be statistically significant when the P value is considered at 0.05, which suggests that the statistical conclusions are robust. Therefore, predictive modeling is further supported by machine learning tools in subsequent studies. From the above observations, it can be assured that the DOE results show a superb regression with an $R^2 \geq 0.92$, and it can be considered that the regression is sufficiently accurate for the target outcomes.

3.6. Measurement system analysis (MSA) and process capability analysis

MSA is performed using X-bar and R chart for 10,000 simulations, as shown in Fig. 6. The acceptable range is determined by

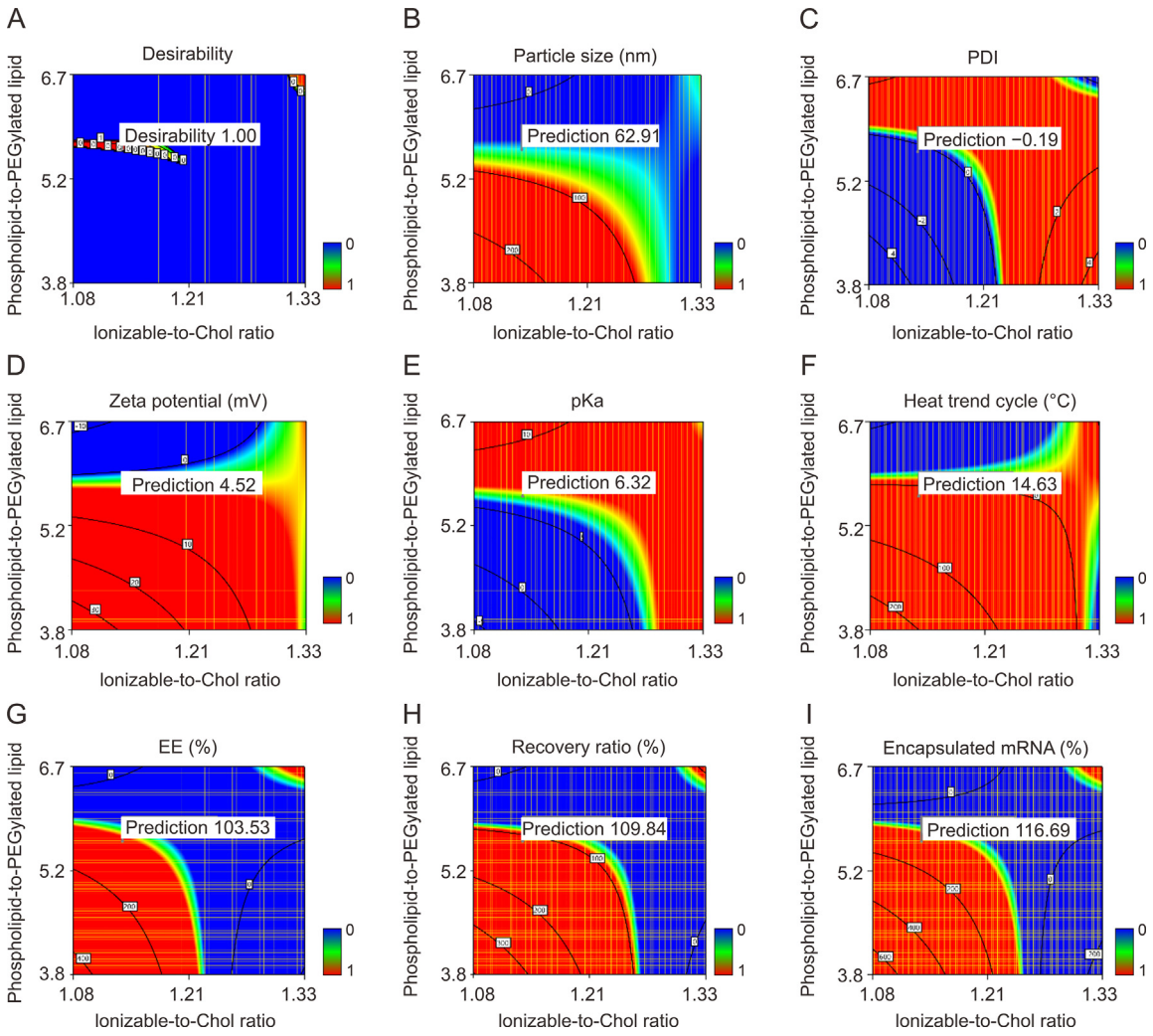

Fig. 5. Optimization of outcome responses. (A) Desirability, (B) particle size, (C) polydispersity index (PDI), (D) Zeta potential, (E) pKa, (F) heat trend cycle, (G) encapsulation efficiency (EE), (H) recovery ratio, and (I) encapsulated mRNA.

tolerance interval analysis using a portion of the simulated population (n = 172). The measurement values of the outcome responses, PS, PDI, ZP, pKa, heat trend cycle, EE, recovery ratio, and encapsulated mRNA were 51.59 nm, 82.09 nm, 0.08 to 0.36, 0.35−2.01 mV, 5.84 to 6.72, −20.50 °C to −12.68 °C, 82.19%%−95.13%, 86.43%%−92.65%, and 72.87%%−85.89%, respectively. For the above measurements, the lower and higher control limits of the responses are calculated using Eq. (S1) and Eq. (S2) and the X-bar chart limits are obtained as PS (Fig. 6A, 65.02 nm−69.06 nm), PDI (Fig. 6B, 0.16 to 0.20), ZP (Fig. 6C, 1.05 mV−1.30 mV), pKa (Fig. 6D, 6.15 to 6.35), heat trend cycle (Fig. 6E, −17.5 °C to −16.0 °C), EE (Fig. 6F, 87.8%−88.9%), recovery ratio (Fig. 6G, 88.7%−90.9%), and encapsulated mRNA (Fig. 6H, 77.9%−80.9%). A process "in control" refers to a stable and predictable process that avoids unnecessary adjustments.

The mean and range of average values are found to be within the control limits (Fig. 6), which are further tested with process capability analysis using Eqs. (S5), (S6), and (S7), and illustrated in Fig. 7. The blue solid curve and black dotted curve represent the samples within the group and overall groups, respectively. The *Cpk* values of the outcome responses, PS, PDI, ZP, pKa, heat trend cycle, EE, recovery ratio, and encapsulated mRNA are 2.046 (Fig. 7A), 1.584 (Fig. 7B), 1.260 (Fig. 7C), 1.475 (Fig. 7D), 1.313 (Fig. 7E), 1.294 (Fig. 7F), 1.315 (Fig. 7G), and 1.843 (Fig. 7H), respectively. A *Cpk* value > 1.300 indicates the proximity of the observed value to a target and consistency around the average performance of the manufacturing process with a 4 σ capacity. These values are important for designing 6σ and ensuring product quality during continuous manufacturing of mRNA-LNP formulation [50].

3.7. Processing conditions optimization using XGBoost and Bayesian optimization

Precise control of CQAs in mRNA-LNP manufacturing is critical for the desired output responses. XGBoost is used to understand the importance of the parameters (Fig. 8). The experimental values (y-axis) and predicted values (x-axis) of PS, PDI, ZP, pKa, heat trend cycle, EE, recovery ratio, and encapsulated mRNA are shown in Figs. 8A−H, respectively. Based on the predicted values of the outcome responses, the predictive accuracy of the XGBoost model is calculated for PS, PDI, ZP, pKa, heat trend cycle, EE, recovery ratio, and encapsulated mRNA in terms of R^2, *RMSE*, and *MAE*, where the corresponding values are obtained as (0.9971, 0.4932, 0.3663), (0.9756, 0.0078, 0.0045), (0.9897, 0.0411, 0.0230), (0.9886, 0.0277, 0.0200), (0.9981, 0.1136, 0.0732), (0.9516, 0.7034, 0.4109), (0.9496, 0.4041, 0.2791), and (0.9414, 0.9054, 0.6322), respectively. All of the parameters have $R^2 > 0.95$, *RMSE* < 0.91, and *MAE* < 0.63 (Table 6), which suggests that the XGBoost model has good predictive ability.

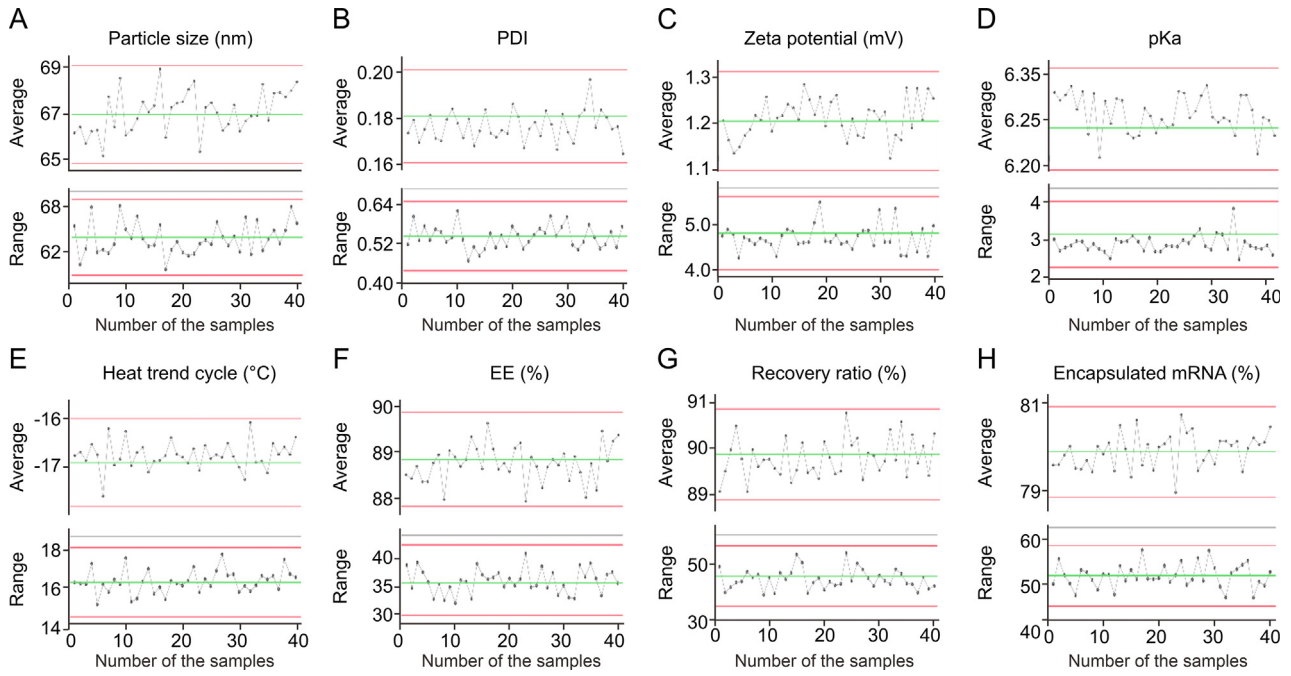

Fig. 6. *X*−bar (top) and R chart (bottom) plots for simulated dataset, where the average and range of each response values are within the respective upper and lower control limits. (A) Particle size, (B) polydispersity index (PDI), (C) Zeta potential), (D) pKa, (E) heat trend cycle, (F) encapsulation efficiency (EE), (G) recovery ratio, and (H) encapsulated mRNA.

Furthermore, Bayesian optimization is implemented on the above predicted datasets using Gaussian process regression, where the input factors are fed to an acquisition function to selectively obtain a desired outcome. In the current scenario, the PS is targeted between 62.70 nm and 72.64 nm (Fig. 9A) and a PDI between 0.07 and 0.17 (Fig. 9B), which is achieved when the input factors, i.e., ionizable-to-chol ratio, PEGylated lipid amount, phospholipid-to-PEGylated lipid, FRR, TFR, and N/P ratio, are set to 1.17, 1.50, 3.76, 4.22 mL/min, 17.85 mL/min, and 9.76, respectively. Similarly, ZP in a range of +0.55 mV to +0.90 mV (Fig. 9C) and preferred pKa in a

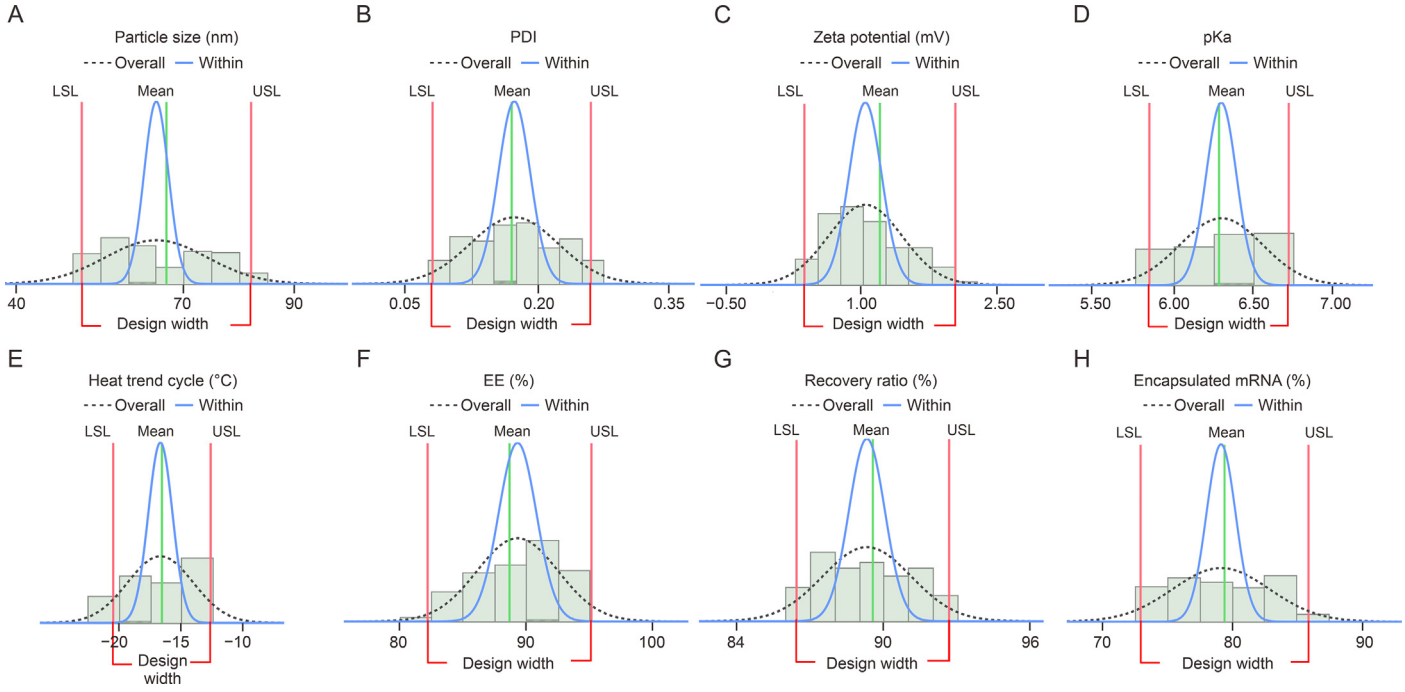

Fig. 7. Process capability analysis illustrated by the blue curve (distribution within the group) and dotted curve (distribution of overall groups). (A) Particle size, (B) polydispersity index (PDI), (C) Zeta potential, (D) pKa, (E) heat trend cycle, (F) encapsulation efficiency (EE), (G) recovery ratio, and (H) encapsulated mRNA. Left and right (red) lines represent the lower and upper specification limits (LSL and USL) while middle green line represent its mean value. The sharp intensity in y-axis direction shows that the conditions easily meet the specification.

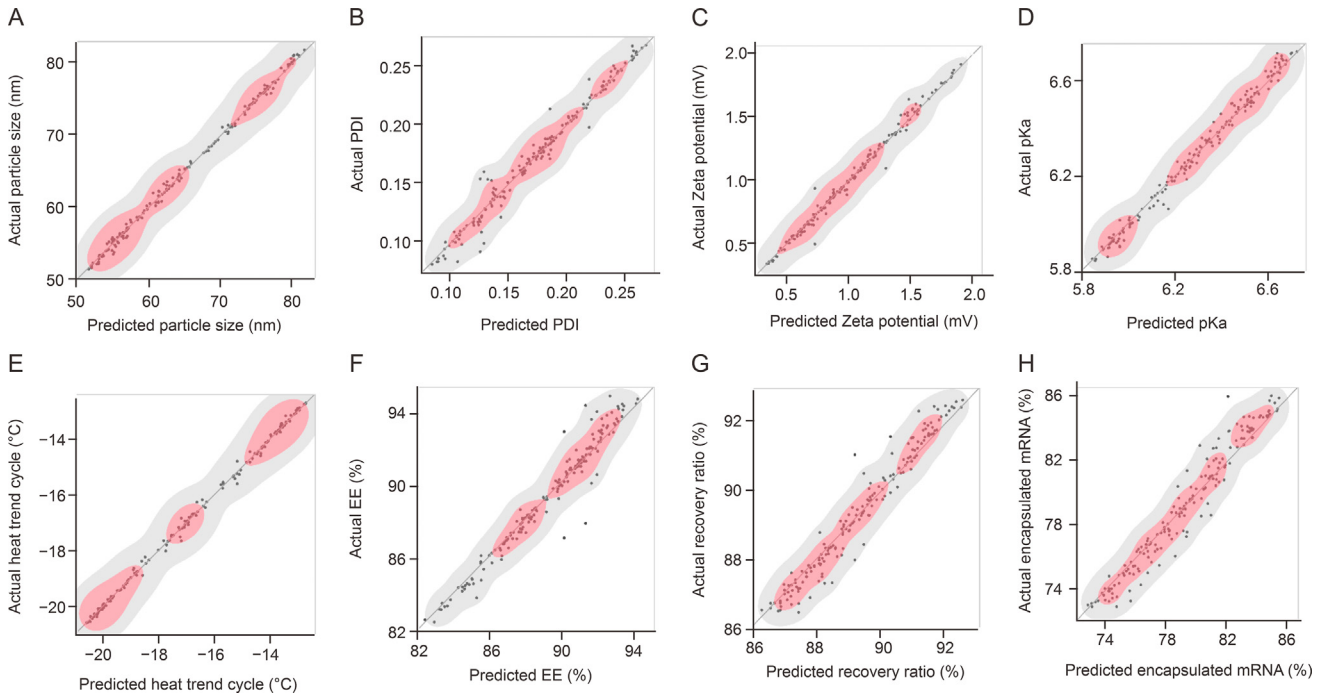

Fig. 8. Plot of actual vs. predicted values of outcome responses using XGBoost tool. (A) Particle size (PS), (B) polydispersity index (PDI), (C) Zeta potential (ZP), (D) pKa, (E) heat trend cycle, (F) encapsulation efficiency (EE), (G) recovery ratio, and (H) encapsulated mRNA.

range of 6.36–6.54 (Fig. 9D) can be obtained when the input factors, i.e., ionizable-to-chol ratio, phospholipid-to-PEGylated lipid, FRR, TFR, and N/P ratio, are set to 1.33, 3.76, 3 mL/min, 12 mL/min, and 10, respectively. The PEGylated lipid amount slightly differs in those two output conditions. For ZP, the predicted PEGylated lipid amount is 2.5%, whereas in the case of pKa, the predicted PEGylated lipid amount is 1.5% (Figs. 9C and D).

A stable heat-resistant mRNA-LNP formulation is targeted at a temperature of −11.60 °C (between −12.82 °C and −10.39 °C) (Fig. 9E), which can be attained if the input factors, i.e., ionizable-to-chol ratio, PEGylated lipid amount, phospholipid-to-PEGylated lipid, FRR, TFR, and N/P ratio, are fixed at 1.15, 2.22%, 6.10, 3.89 mL/min, 17.49 mL/min, and 9.21, respectively. Similarly, an mRNA-LNP formulation with a high EE (>88%) can be predicted when Bayesian optimization for the input factors (i.e., ionizable-to-chol ratio, PEGylated lipid amount, phospholipid-to-PEGylated lipid, FRR, TFR, and N/P ratio) is set as 1.15, 2.02%, 5.15, 4.00 mL/min, 16.78 mL/min, and 8.22, respectively (Fig. 9F). Alternatively, a recovery ratio > 89% can be predicted when the input factors, i.e., ionizable-to-chol ratio, PEGylated lipid amount, phospholipid-to-PEGylated lipid, FRR, TFR, and N/P ratio, are set to 1.14, 1.61%, 4.84, 4.13 mL/min, 15.14 mL/min, and 7.27, respectively (Fig. 9G). In another scenario, encapsulated mRNA levels >78% can be estimated if those six factors are changed to 1.19, 1.56%, 5.34, 3.94 mL/min, 15.83 mL/min, and 7.96, respectively (Fig. 9H).

Table 6
Based on the experimental and predicted outcomes, the prediction accuracy of the XGBoost model is calculated and expressed in terms of R^2, *RMSE*, and *MSE*.

Responses	R^2	*RMSE*	*MSE*
PS	0.9971	0.4932	0.3663
PDI	0.9756	0.0078	0.0045
ZP	0.9897	0.0411	0.0230
pKa	0.9886	0.0277	0.0200
Heat trend cycle	0.9981	0.1136	0.0732
EE	0.9516	0.7034	0.4109
Recovery ratio	0.9496	0.4041	0.2791
Encapsulated mRNA	0.9414	0.9054	0.6322

PS: particle size; PDI: polydispersity index; ZP: Zeta potential; EE: encapsulation efficiency; RMSE: root mean square error; MSE: mean squared error.

3.8. Lipid mix ratio prediction for desired mRNA-LNPs using the SVEM model

SVEM model is implemented to predict the probable lipid mix ratio, which will give the targeted output responses as mentioned in above Bayesian optimization. The feasible lipid mix ratio (ionizable, phospholipid, and others (chol + PEGylated lipid)) is shaded in red, while the least feasible lipid ratio is shaded in blue (Fig. 10). The low amount of ionizable lipid (0.10–0.20), high amount of phospholipid (0.85–0.90), and medium quantity of others (chol + PEGylated lipid, 0.30 to 0.65) favor the targeted responses (Fig. 10A). Specifically, the amounts of chol and PEGylated lipid in the other components are high (0.95–1.00) and low (0–0.05), respectively (Fig. 10B). A prior study demonstrated that such in silico nanoparticle composition screening had potential applications in predicting de novo biological experiments [51]. In the present case, Fig. 10 shows the possible region of the lipid mix composition (high intensity in red color) that meets the optimized (target) conditions discussed in above section.

The SVEM model is fed to forward regression, Lasso, and neural network, which give a predictive accuracy in terms of R^2, *RASE*, and *AAE* (Table 7). All three SVEM models have delivered the desired outcomes. When the models are thoroughly compared, the predictive accuracy of the models is in the order of SVEM–Fwd_regression < SVEM–Lasso < SVEM–Neural model, listed in this order based on high R^2 and low *RASE* and *AAE* values (Table 7). For instance, the predictive accuracy of PS in terms of R^2 is in the order of SVEM–Fwd_regression: 0.9910 < SVEM–Lasso: 0.9977 < SVEM–Neural: 0.9980. Similarly, the predictive accuracy of PS in terms of *RASE* is in the order of SVEM–Fwd_regression: 1.3921 > SVEM–Lasso: 0.7012 > SVEM–Neural: 0.6556 and in terms of *AAE*, is in the order of SVEM–Fwd_regression:

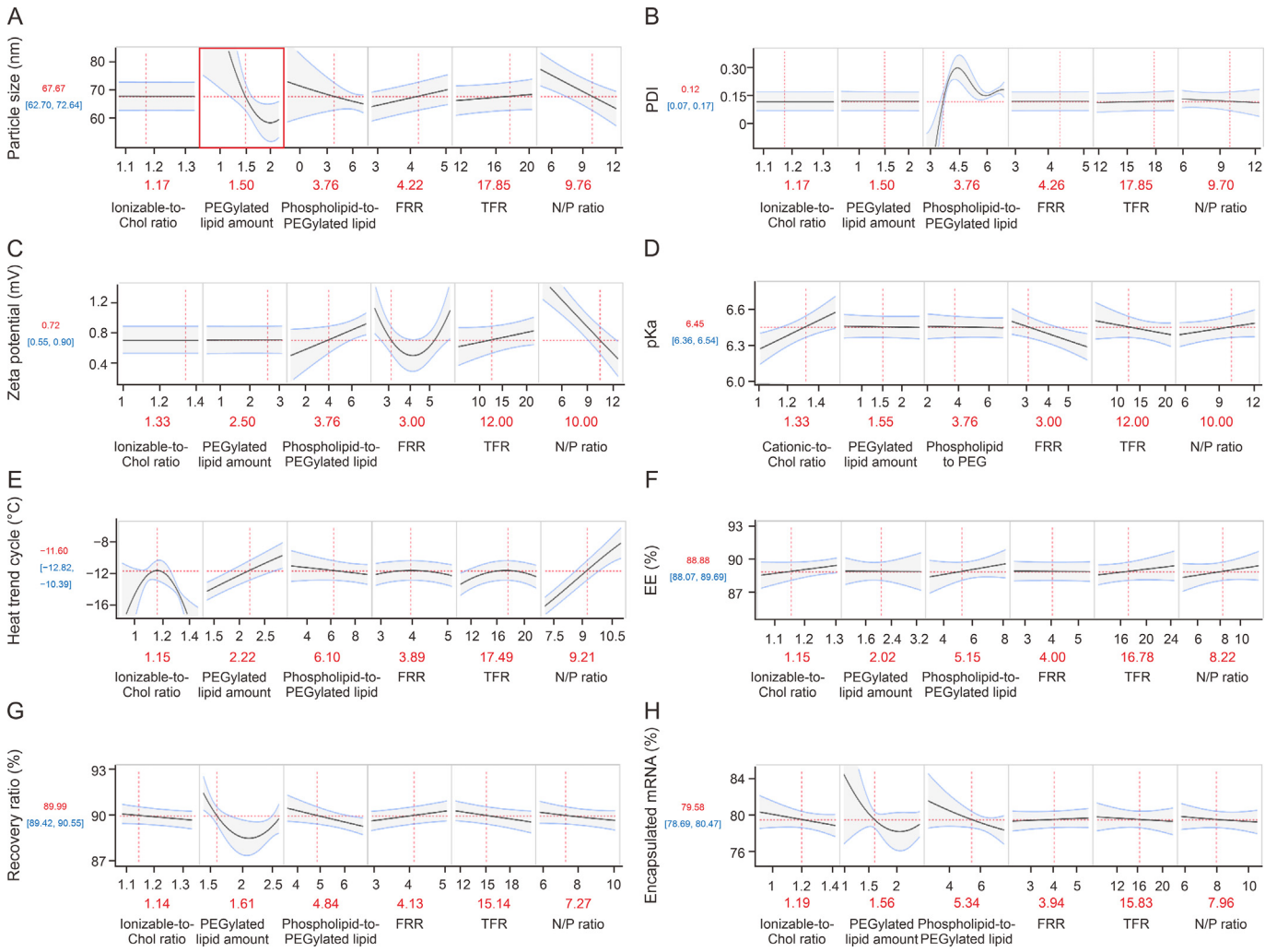

Fig. 9. Bayesian optimization of outcome responses. (A) Particle size, (B) polydispersity index (PDI), (C) Zeta potential), (D) pKa, (E) heat trend cycle, (F) encapsulation efficiency (EE), (G) recovery ratio, and (H) encapsulated mRNA. N/P: nitrogen from lipid and phosphate from nucleic acid; FRR: flow rate ratio; TFR: total flow rate.

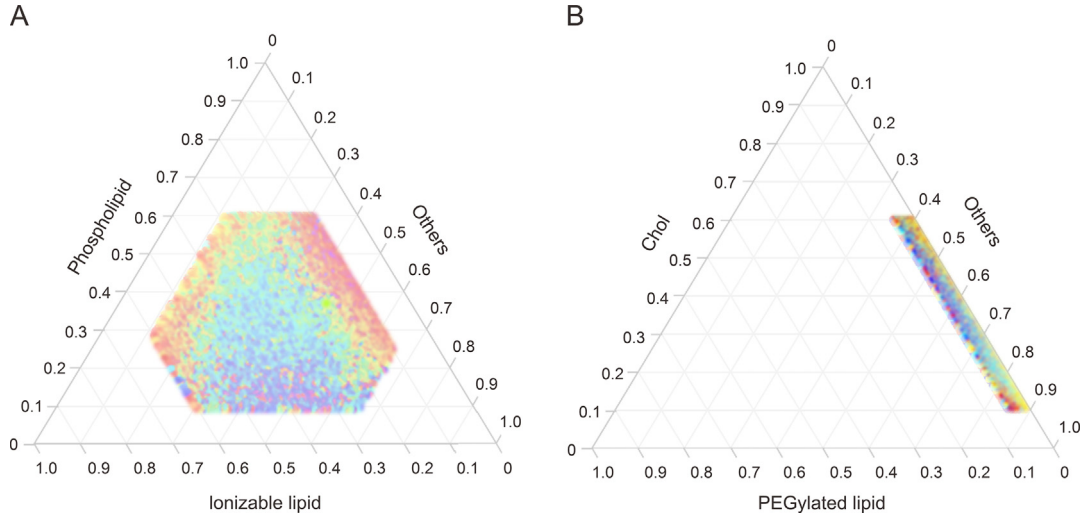

Fig. 10. Determination of the sweet spot from the lipid mix ratio that can give the target outcomes using a self-validated ensemble model (SVEM).(A) Prediction of the response (red: most desirable, blue: least desirable) using a lipid mix ratio of ionizable, phospholipid, and the others (Chol + PEGylated lipid). (B) Prediction of the response using Chol, PEGylated lipid, and the others (ionizable lipid + phospholipid).

Table 7
Prediction accuracy of three subclasses of the self-validated ensemble (SVEM) model: SVEM-Fwd_regression, SVEM-Lasso, and SVEM-Neural are evaluated and expressed in terms of R^2, *RASE*, and *AAE*.

Responses	SVEM−Fwd_regression			SVEM−Lasso			SVEM−Neural		
	R^2	*RASE*	*AAE*	R^2	*RASE*	*AAE*	R^2	*RASE*	*AAE*
PS	0.9910	1.3921	0.8270	0.9977	0.7012	0.4935	0.9980	0.6556	0.3337
PDI	0.9851	0.0101	0.0070	0.9893	0.0086	0.0048	0.9908	0.0080	0.0027
ZP	0.9689	0.1402	0.0761	0.9693	0.1392	0.0814	0.9742	0.1276	0.0374
pKa	0.9919	0.0382	0.0221	0.9990	0.0137	0.0095	0.9994	0.0103	0.0048
Heat trend cycle	0.9710	0.6386	0.4194	0.9696	0.6545	0.4541	0.9816	0.5094	0.1542
EE	0.9941	0.4671	0.3532	0.9964	0.3663	0.2459	0.9976	0.3004	0.0900
Recovery ratio	0.9929	0.2464	0.1561	0.9940	0.2275	0.1236	0.9945	0.2177	0.0868
Encapsulated mRNA	0.9875	0.6964	0.4069	0.9884	0.6703	0.3686	0.9899	0.6270	0.1911

PS: particle size; PDI: polydispersity index; ZP: Zeta potential; EE: encapsulation efficiency; RMSE: root mean square error; AAE: average absolute error.

Table 8
Independent experimental runs of messenger RNA-lipid nanoparticles (mRNA-LNP) vaccine formulations along with the actual outcomes and the predicted outcomes.

	Factors (independent variables)								
S. No.	Ionizable lipid	Phospholipid	PEGylated lipid	PEGylated lipid	Ionizable-to-chol ratio	Phospholipid-to- PEGylated lipid ratio	N/P ratio	FRR (mL/min)	TFR (mL/min)
1	DOTMA	DOPE	Ceramide	1.6	1.3137	3.83	10	4	13
2	MC3	DMPC	DSG	1.9	1.0888	5.96	6	4	13

	Responses (Experimental)							
S. No.	PS (nm)	PDI	ZP (mV)	pKa	Heat trend cycle (℃)	EE (%)	Recovery ratio (%)	Encapsulated mRNA (%)
1	94.42	0.069	3.09	5.87	−17.93	95.558	96.282	91.736
2	96.95	0.071	2.33	6.21	−7.62	96.439	94.389	90.653

	Responses (Predicted)							
S. No.	PS (nm)	PDI	ZP (mV)	pKa	Heat trend cycle (℃)	EE (%)	Recovery ratio (%)	Encapsulated mRNA (%)
1	95.16	0.061	3.29	5.91	−17.32	95.880	96.444	92.268
2	97.34	0.069	2.26	6.23	−7.78	96.768	94.486	90.810

DOTMA: 1,2-di-*O*-octadecenyl-3-triMAP(-methylammonium propane); MC3: (6Z,9Z,28Z,31Z)-heptatriacont-6,9,28,31-tetraene-19-yl-4-(dimethylamino) butanoate; DMPC: 1,2-dimyristoyl-*sn*-glycero-3-PC; DOPE: 1,2-dioleoyl-*sn*-glycero-3-PE; DSG: 1,2-distearoyl-rac-glycerol-(PEG2K-polyethylene glycol-2K); Ceramide: N-palmitoyl-sphingosine-1-succinyl(methoxyPEG2K); PS: particle size; PDI: polydispersity index; ZP: Zeta potential; EE: encapsulation efficiency; N/P: nitrogen from lipid and phosphate from nucleic acid; FRR: flow rate ratio; TFR: total flow rate.

0.8270 > SVEM−Lasso: 0.4935 > SVEM−Neural: 0.3337 (Table 7). Consistently, equivalent results have been obtained for other output responses (PDI, ZP, pKa, heat trend cycle, EE, recovery ratio, encapsulated mRNA). Therefore, it can be implied that high prediction performance is achieved with a small sample size [22]. This is possible because of the fractional random weight bootstrap that assigns anti-correlated training and validation weights to individual observations. The eventual SVEM model is the outcome of the average of bootstrapped models. SVEM model performs better than single model selection approaches. The mRNA-LNP vaccine formulations under the zone marked in red (Fig. 10) can be experimentally obtained and optionally conducted comprehensive second-phase investigation [19]. In order to achieve the desired outcomes from the practical experiments, both ML tools focus on lipid composition and process factors. Specifically, SVEM is simpler compared to XGBoost/Bayesian optimization as it does not need separate training and validation datasets, avoids complications that arise in traditional DOE with mixture-process experiments, and enables QbD approach into the RNA vaccine production process [11,19]. A prior study implemented lightGBM algorithm to predict the mRNA-LNP formulations with better IgG titer, validated with *in vivo* test on ACE2 expression, and attempted to recognize the ionizable lipid, which serves a major role in nucleic acid therapeutics (with > 87% accuracy) [25]. Alternatively, we have attempted to predict ionizable lipid type, ratio of different lipids, and process conditions with other two ML tools and found that SVEM could be a simpler and reliable option to optimize and predict the conditions (accuracy >97%). Two independent mRNA-LNP vaccine formulations are prepared based on the optimization and compared to the SVEM−Fwd_regression prediction (Table 8), which suggests the SVEM prediction is close to the actual experimental outcomes. For instance, PS (experimental), 94−96 nm is approximately close to PS (predicted), 95−97 nm, which is consistent with a prior study [52]. mRNA vaccine with a particle size close to 110 nm showed improved immunoprophylactic and therapeutic tumor suppression in mice [53]. When delving into the model, incorporation of the neural network further improves the existing model [39−41]. Incorporation of ML tools in prediction and determination of optimal conditions lead to improve accuracy, reproducibility, and speed while it reduces cost, saves time, eliminates chemical consumption, reduces equipment use, and decreases human error, which can be significant in the development of efficient mRNA-LNP vaccine, scale-up, and even continuous manufacturing [54]. Another benefit can be the simultaneous prediction of all responses at the same time.

4. Conclusions

It is challenging to optimize both bioprocessing conditions and the lipid mix ratio of the mRNA-LNP formulations simultaneously while considering eight possible outcomes. In this case, dimensionality reduction is performed to select the significant factors, followed by XGBoost and Bayesian optimization for bioprocess optimization and the SVEM model for lipid mix ratio prediction. These ML tools i.e., XGBoost and Bayesian optimization are particularly appealing to formulation scientists because of the

possibility to simultaneously analyze multiple outcomes (eight responses) by conducting a small-size experimental design (24 experiments). The experimental outcomes were tested with predicted responses > 94% using the XGBoost model and tuned with Bayesian optimization. Based on the prediction, the lipid mix ratio can also be designed to give targeted outcomes with predictive accuracy > 97% using the SVEM model.

CRediT authorship contribution statement

Ravi Maharjan: Conceptualization, Investigation, Methodology, Visualization, Writing – original draft, Writing – review & editing. **Ki Hyun Kim:** Investigation, Validation. **Kyeong Lee:** Methodology, Validation. **Hyo-Kyung Han:** Validation. **Seong Hoon Jeong:** Funding acquisition, Project administration, Resources, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that there are no conflicts of interest.

Acknowledgments

This work was partially supported by the Advance Production of Vaccine Raw Materials (Grant Nos.: 20022404 and 20018168) funded by the Ministry of Trade, Industry & Energy (MOTIE, Korea). This work was also partially supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (Grant No.: NRF-2018R1A5A2023127) and Dongguk University Research Fund of 2023 (Grant No.: S-2023-G0001-00099).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jpha.2024.100996.

References

[1] J. Pang, M.X. Wang, I.Y.H. Ang, et al., Potential rapid diagnostics, vaccine and therapeutics for 2019 novel coronavirus (2019-nCoV): A systematic review, J. Clin. Med. 9 (2020), 623.
[2] S.C. Semple, R. Leone, C.J. Barbosa, et al., Lipid nanoparticle delivery systems to enable mRNA-based therapeutics, Pharmaceutics 14 (2022), 398.
[3] L. Miao, Y. Zhang, L. Huang, mRNA vaccine for cancer immunotherapy, Mol. Cancer 20 (2021), 41.
[4] Q. Xiong, G.Y. Lee, J. Ding, et al., Biomedical applications of mRNA nanomedicine, Nano Res. 11 (2018) 5281–5309.
[5] G. Prakash, A. Shokr, N. Willemen, et al., Microfluidic fabrication of lipid nanoparticles for the delivery of nucleic acids, Adv. Drug Deliv. Rev. 184 (2022), 114197.
[6] K. Osouli-Bostanabad, S. Puliga, D.R. Serrano, et al., Microfluidic manufacture of lipid-based nanomedicines, Pharmaceutics 14 (2022), 1940.
[7] A. Algarni, E.H. Pilkington, E.J.A. Suys, et al., *In vivo* delivery of plasmid DNA by lipid nanoparticles: The influence of ionizable cationic lipids on organ-selective gene expression, Biomater. Sci. 10 (2022) 2940–2952.
[8] M. Kawaguchi, M. Noda, A. Ono, et al., Effect of cholesterol content of lipid composition in mRNA-LNPs on the protein expression in the injected site and liver after local administration in mice, J. Pharm. Sci. 112 (2023) 1401–1410.
[9] E. Oude Blenke, E. Örnskov, C. Schöneich, et al., The storage and In-use stability of mRNA vaccines and therapeutics: Not A cold case, J. Pharm. Sci. 112 (2023) 386–403.
[10] H. Helgers, A. Schmidt, L.J. Lohmann, et al., Towards autonomous operation by advanced process control—Process analytical technology for continuous biologics antibody manufacturing, Processes 9 (2021), 172.
[11] S. Daniel, Z. Kis, C. Kontoravdi, et al., Quality by Design for enabling RNA platform production processes, Trends Biotechnol. 40 (2022) 1213–1228.
[12] H.H. Ly, S. Daniel, S.K.V. Soriano, et al., Optimization of lipid nanoparticles for saRNA expression and cellular activation using a design-of-experiment approach, Mol. Pharm. 19 (2022) 1892–1905.
[13] K. Nag, M.E.H. Sarker, S. Kumar, et al., DoE-derived continuous and robust process for manufacturing of pharmaceutical-grade wide-range LNPs for RNA-vaccine/drug delivery, Sci. Rep. 12 (2022), 9394.
[14] Y. Suzuki, T. Miyazaki, H. Muto, et al., Design and lyophilization of lipid nanoparticles for mRNA vaccine and its robust immune response in mice and nonhuman Primates, Mol. Ther. Nucleic Acids 30 (2022) 226–240.
[15] S. Chen, X. Huang, Y. Xue, et al., Nanotechnology-based mRNA vaccines, Nat. Rev. Methods Primers 3 (2023), 63.
[16] Z. Ghaemmaghamian, R. Zarghami, G. Walker, et al., Stabilizing vaccines via drying: Quality by design considerations, Adv. Drug Deliv. Rev. 187 (2022), 114313.
[17] A. Ortiz-Perez, D. van Tilborg, R. van der Meel, et al., Machine learning-guided high throughput nanoparticle design, Digit. Discov. 3 (2024) 1280–1291.
[18] A.W. Freyn, C. Atyeo, P.L. Earl, et al., A monkeypox mRNA-lipid nanoparticle vaccine targeting virus binding, entry, and transmission drives protection against lethal orthopoxviral challenge, bioRxiv. (2022). https://doi.org/10.1101/2022.12.17.520886.
[19] A.T. Karl, S. Essex, J. Wisnowski, et al., A workflow for lipid nanoparticle (LNP) formulation optimization using designed mixture-process experiments and self-validated ensemble models (SVEM), J. Vis. Exp. 198 (2023).
[20] A. Schmidt, H. Helgers, F.L. Vetter, et al., Process automation and control strategy by quality-by-design in total continuous mRNA manufacturing platforms, Processes 10 (2022), 1783.
[21] K.J. Kauffman, J.R. Dorkin, J.H. Yang, et al., Optimization of lipid nanoparticle formulations for mRNA delivery *in vivo* with fractional factorial and definitive screening designs, Nano Lett. 15 (2015) 7300–7306.
[22] T. Lemkus, P. Ramsey, C. Gotwalt, et al., Self-validated ensemble models for design of experiments, arXiv. (2021). https://doi.org/10.48550/arXiv.2103.09303.
[23] M. Zhao, R. Wang, K. Yang, et al., Nucleic acid nanoassembly-enhanced RNA therapeutics and diagnosis, Acta Pharm. Sin. B 13 (2023) 916–941.
[24] E. Yeung, P. Ramsey, Optimization of a conventional glycosylation analytical method using machine learning and experimental design, BioProcessing Journal 20 (2021).
[25] W. Wang, S. Feng, Z. Ye, et al., Prediction of lipid nanoparticles for mRNA vaccines by the machine learning algorithm, Acta Pharm. Sin. B 12 (2022) 2950–2962.
[26] R. Maharjan, S. Hada, J.E. Lee, et al., Comparative study of lipid nanoparticle-based mRNA vaccine bioprocess with machine learning and combinatorial artificial neural network-design of experiment approach, Int. J. Pharm. 640 (2023), 123012.
[27] Y. Xu, S. Ma, H. Cui, et al., AGILE platform: A deep learning-powered approach to accelerate LNP development for mRNA delivery, bioRxiv, 2023. https://doi.org/10.1101/2023.06.01.543345.
[28] Y. Qin, A.A. Walters, N. Rouatbi, et al., Evaluation of a DoE based approach for comprehensive modelling of the effect of lipid nanoparticle composition on nucleic acid delivery, Biomaterials 299 (2023), 122158.
[29] A. Ouranidis, C. Davidopoulou, R.K. Tashi, et al., Pharma 4.0 continuous mRNA drug products manufacturing, Pharmaceutics 13 (2021), 1371.
[30] R. Maharjan, S.H. Jeong, Application of different models to evaluate the key factors of fluidized bed layering granulation and their influence on granule characteristics, Powder Technol. 408 (2022), 117737.
[31] L. Schoenmaker, D. Witzigmann, J.A. Kulkarni, et al., mRNA-lipid nanoparticle COVID-19 vaccines: Structure and stability, Int. J. Pharm. 601 (2021), 120586.
[32] M.M. Tosi, A.P. Ramos, B.S. Esposto, et al., Chapter Six Dynamic light scattering (DLS) of nanoencapsulated food ingredients, in: S.M. Jafari (Ed.), Characterization of Nanoencapsulated Food Ingredients, Vol. 4, Academic Press, New York, 2020, pp. 191–211.
[33] M.J. Carrasco, S. Alishetty, M.G. Alameh, et al., Ionization and structural properties of mRNA lipid nanoparticles influence expression in intramuscular and intravascular administration, Commun. Biol. 4 (2021), 956.
[34] M.I. Henderson, Y. Eygeris, A. Jozic, et al., Leveraging biological buffers for efficient messenger RNA delivery via lipid nanoparticles, Mol. Pharm. 19 (2022) 4275–4285.
[35] Y. Fan, C.W. Yen, H.C. Lin, et al., Automated high-throughput preparation and characterization of oligonucleotide-loaded lipid nanoparticles, Int. J. Pharm. 599 (2021), 120392.
[36] A. Hasanzadeh, M.R. Hamblin, J. Kiani, et al., Could artificial intelligence revolutionize the development of nanovectors for gene therapy and mRNA vaccines? Nano Today 47 (2022), 101665.
[37] D.Y. Ding, Y. Zhang, Y. Jia, et al., Machine learning-guided lipid nanoparticle design for mRNA delivery, arXiv. (2023). https://doi.org/10.48550/arXiv.2308.01402.
[38] S.A. Damiati, D. Rossi, H.N. Joensson, et al., Artificial intelligence application for rapid fabrication of size-tunable PLGA microparticles in microfluidics, Sci. Rep. 10 (2020), 19517.
[39] P. Hersh, SVEM in JMP 17. https://community.jmp.com/t5/Elevated-Thinking/SVEM-in-JMP-17/ba-p/538553. (Accessed 3 May 2024).
[40] P. Kay, P. Hersh, Re-Thinking the Design and Analysis of Experiments? (2021-EU-30MP-776). https://community.jmp.com/t5/Discovery-Summit-Europe-2021/Re-Thinking-the-Design-and-Analysis-of-Experiments-2021-EU-30MP/ta-p/349240. (Accessed 25 December 2023).
[41] P. Ramsey, W. Levin, T. Lemkus, et al., SVEM: A Paradigm Shift in Design and Analysis of Experiments (2021-EU-45MP-779). https://community.jmp.com/t5/Discovery-Summit-Europe-2021/SVEM-A-Paradigm-Shift-in-Design-and-Analysis-of-Experiments-2021/ta-p/349244. (Accessed 30 November 2023).

[42] K. Okuda, Y. Sato, K. Iwakawa, et al., On the size-regulation of RNA-loaded lipid nanoparticles synthesized by microfluidic device, J. Control. Release 348 (2022) 648–659.
[43] X. Hou, T. Zaks, R. Langer, et al., Lipid nanoparticles for mRNA delivery, Nat. Rev. Mater. 6 (2021) 1078–1094.
[44] A. Lamoot, J. Lammens, E. De Lombaerde, et al., Successful batch and continuous lyophilization of mRNA LNP formulations depend on cryoprotectants and ionizable lipids, Biomater. Sci. 11 (2023) 4327–4334.
[45] C. Malburet, L. Leclercq, J.F. Cotte, et al., Size and charge characterization of lipid nanoparticles for mRNA vaccines, Anal. Chem. 94 (2022) 4677–4685.
[46] A. Ji, M. Xu, Y. Pan, et al., Lipid microparticles show similar efficacy with lipid nanoparticles in delivering mRNA and preventing cancer, Pharm. Res. 40 (2023) 265–279.
[47] R.A. Meyer, S.Y. Neshat, J. Green, et al., Targeting strategies for mRNA delivery, Mater. Today Adv. 14 (2022), 100240.
[48] Y. Fan, Z. Shi, S. Ma, et al., Spectroscopy-based local modeling method for high-throughput quantification of nucleic acid loading in lipid nanoparticles, Anal. Chem. 94 (2022) 9081–9090.
[49] X.G. Chen, M. Graužinytė, A.W. van der Vaart, et al., Applying pattern recognition as a robust approach for silicone oil droplet identification in flow-microscopy images of protein formulations, J. Pharm. Sci. 110 (2021) 1643–1651.
[50] J. Yang, F. Meng, S. Huang, et al., Process capability analysis for manufacturing processes based on the truncated data from supplier products, Int. J. Prod. Res. 58 (2020) 6235–6251.
[51] D. Gong, E. Ben-Akiva, A. Singh, et al., Machine learning guided structure function predictions enable in silico nanoparticle screening for polymeric gene delivery, Acta Biomater. 154 (2022) 349–358.
[52] Y. Lin, Y. Wang, J. Ding, et al., Reactivation of the tumor suppressor PTEN by mRNA nanoparticles enhances antitumor immunity in preclinical models, Sci. Transl. Med. 13 (2021), eaba9772.
[53] M.A. Islam, J. Rice, E. Reesor, et al., Adjuvant-pulsed mRNA vaccine nanoparticle for immunoprophylactic and therapeutic tumor suppression in mice, Biomaterials 266 (2021), 120431.
[54] B. Yarahmadi, S.M. Hashemianzadeh, S.M. Milani Hosseini, Machine-learning-based predictions of imprinting quality using ensemble and non-linear regression algorithms, Sci. Rep. 13 (2023), 12111.