Calculating Apparent pK_a Values of Ionizable Lipids in Lipid Nanoparticles

Nicholas B. Hamilton[1], Steve Arns[2], Mee Shelley[1], Irene Bechis[3], and John C. Shelley[1*]

[1]*Schrödinger, Inc., 101 SW Main St., Suite 1300, Portland OR 97204, USA*

[2]*Acuitas Therapeutics, 6190 Agronomy Road, Suite 405, Vancouver, BC, Canada, V6T 1Z3*

[3]*Schrödinger, GmbH, Glücksteinallee 25, 68163 Mannheim, Germany*

Abstract:

Creating new ionizable lipids for use in lipid nanoparticles (LNPs) is an active field of research. One of the critical properties for selecting suitable ionizable lipids is the apparent pK_a value of the lipid as formulated in an LNP. We have developed a structure-based, computational methodology for the prediction of the apparent pK_a value of ionizable lipids within LNPs. This methodology has been validated for the three most successful ionizable lipids to date which are present in the mRNA LNP COVID-19 vaccines COMIRNATY® (Pfizer/BioNTech) and Spikevax® (Moderna), and the siRNA LNP therapeutic Onpattro® (Alnylam). The calculation was also applied to Lipid A, a variant of the ionizable lipid used in COMIRNATY®. We believe

that this new technology permits systematic computational prescreening of ionizable lipids to select the most promising candidates for synthesis and experimental testing, accelerating the formulation improvement process and reducing costs.

Introduction and Background

Due to their instrumental role in combating the SARS-CoV-2 pandemic[1,2], lipid nanoparticles (LNPs) and the mRNA vaccines they enable are now household names. For instance, as of Mar. 2024, ~4.6 billion doses of the COMIRNATY® (Pfizer/BioNTech) vaccine have been shipped world-wide[3]. These statistics represent a remarkable achievement for humanity during society's time of need and these vaccines would not have been possible without the earlier development of LNP delivery systems. As nucleic acid-based therapies, including mRNA, continue to be adapted for new indications and diseases[4,5], next generation LNP formulations providing more efficient and selective delivery systems are one way to further enable these therapies.

mRNA is an endogenous messenger molecule that bridges the gap between DNA and the ribosome, enabling the synthesis of protein from instructions in genetic code. The coding architecture of mRNA is readily sequenced[6] and synthetic methods are well understood[7]. As such, engineered mRNA sequences can harness the internal mechanisms of the cell to synthesize proteins that have broad therapeutic applications from vaccines to protein replacement to gene editing and more. Despite the elegance of this approach, susceptibility to endogenous nucleases combined with the size and negative charge inherent to long mRNA strands that limits cellular penetration make the efficient delivery of mRNA to the cytoplasm where the ribosome is located

a significant challenge[8]. mRNA encapsulated in an LNP is afforded both robust protection from degradation and a fine-tunable mechanism for optimizing payload delivery.

After uptake into target cells via endocytosis an LNP is exposed to an increasingly acidic environment as the endosome matures. As the pH drops, the amines of the ionizable lipids (ILs) are protonated and the LNP is understood to undergo structural changes as the ILs begin to associate with the anionic endosomal membrane. This process facilitates escape of the encapsulated mRNA from the endosome into the cytoplasm; a crucial step on the path to translation of a protein of interest. It is well known that the pK_a values of ILs in bulk aqueous solution (p${K_a}^S$) are typically much higher than the apparent pK_a (p${K_a}^A$) values for the same lipids as measured in LNP formulations[9]. p${K_a}^A$ values in the specific range of 6-7 is one requirement to observe functional activity with a typical LNP formulation. This p${K_a}^A$range allows for particle destabilization and endosomal escape of the mRNA as the endosome acidifies after the LNP is internalized. Additionally, a p${K_a}^A$ value lower than the pH of the blood (~7.4) avoids the LNP having significant net positive charge, which is a known driver of LNP toxicity. Finally, this p${K_a}^A$ range supports effective encapsulation of mRNA at the acidic pH of the formulation process.

Considering these factors, a key design feature in the search for new ILs is having an appropriate p${K_a}^A$ value when formulated in an LNP. As such, there is considerable interest in developing computational tools to predict p${K_a}^A$ for LNPs[10] to support the systematic design of ILs and to increase the success rate of synthesized lipids having ideal LNP p${K_a}^A$ values. Herein, we present a computational method which can be performed in roughly one week and provides reliable p${K_a}^A$ values for ionizable LNPs. This methodology utilizes umbrella sampling[11-13] to quantify the pK_a shift (ΔpK_a) of a lipid upon transfer from bulk aqueous solution to an environment that is locally similar to that they would experience in an LNP when not directly associating with a RNA

molecule. This shift can in turn be combined with either experimental or computationally derived $pK_a{}^S$ values to yield $pK_a{}^A$ values for specific LNP formulations.

Methodology

The Structured Liquid Builder, which is a part of Schrödinger Materials Science Suite[14] and a front-end to Packmol[15], generated the initial lipid bilayers at a surface area per lipid of 61 Å^2. Figure 1 gives the compositions for each of the four systems studied (more detailed information is available in Table S1). Two of these systems represent the mRNA LNP formulations used in SARS-CoV-2 vaccines COMIRNATY® (Pfizer/BioNTech) and Spikevax® (Moderna) while the third represents Onpattro® (Alnylam), a siRNA LNP product for treating polyneuropathy of hATTR. Lipid A, a variant of the IL from the COMIRNATY® vaccine, ALC-0315, was also studied since it has an unusually low $pK_a{}^A$ value. We omitted the polyethylene glycol (PEG) lipids which comprise approximately 1-2 mol % of the lipids present in these formulations, because it is largely shed prior to endocytosis[16] and fine-tuning of the apparent $pK_a{}^A$ value seems to be most sensitive for endosomal escape. Leaving out these large flexible molecules also simplifies the calculations by reducing the required system size and simulation run times. The compositions of the ALC-0315 and Lipid A systems differ slightly because one additional cholesterol molecule was included in each half of the bilayer in the latter due to change in rounding off molecule counts late in the project. The mol % values of the other lipids were proportionately increased to compensate. The ILs for each of these formulations are depicted in Figure 1. In each system equal amounts of protonated and neutral forms of the IL are included,

implying that the pH is effectively equal to the $pK_a{}^A$ for that IL. The built structures included 0.15M NaCl, as well as additional Cl^- counterions to ensure that the system has a net overall charge of zero.

Our umbrella sampling approach involves accurately determining the distribution of distances along a straight line, in this case, the distance from the center of membrane to the amine atom in a lipid molecule (z), within a series of distance windows. Reliable results depend on effective sampling of conformational space for this special lipid molecule which presents a challenge for molecules as large and flexible as the ILs depicted in Figure 1. In the following we will refer to calculations for each type of system by the label used for the IL in Figure 1, i.e., ALC-0315 for COMIRNATY®*, MC3 (from DLin-MC3-DMA) for Onpattro®*, and SM-102 for Spikevax®* where the * indicates that the PEG lipids and RNA were left out as well as rounding to an integer number of molecules. Lipid A refers to the variant of the COMIRNATY®* formulation with that lipid instead of ALC-0315.

Initial equilibration of each bilayer utilized the standard relaxation protocol used in Schrödinger Suite's implementation of Desmond[17,18] within Maestro[19]. This protocol consisted of a Brownian dynamics step, a canonical ensemble step, and three subsequent isobaric-isothermal ensemble simulations. A 1μs NPγT molecular dynamics (MD) simulation at 310.15K, 1.013 bar and 0 surface tension was then performed to produce a well-relaxed bilayer.

A lipid buried within the membrane is selected for umbrella sampling. Umbrella sampling windows were then created for each system starting from the two windows centered above and below the initial lipid position (as defined by the headgroup N-atom), relaxing in each window for 10 ns before initiating the relaxation for the next window further out with a window separation of 1.0 Å, and an applied harmonic potential of 2.0 kcal/mol/Å^2. A harmonic potential with a 5.0

kcal/mol/Å^2 force constant centered on $z = 0$ was applied to the center of mass for the membrane. Each relaxed window was sampled for 100 ns. Both the relaxations and the sampling runs utilized NPγT MD under the same conditions used for the 1 μs simulation. The range of z values sampled was -2 to 50 Å for all systems except MC3 where the range was -2 to 55 Å, covering the range of z values from the central region of the membrane to bulk aqueous solution. The distribution of z values visited was then calculated for each window for times between 30 and 100 ns and, utilizing the Weighed Histogram Analysis Method (WHAM)[20], the unbiased Potential of Mean Force (PMF) was determined at a 0.2 Å resolution for each system in Figure 1. To obtain more accurate estimates and statistics, we have sampled each system in Figure 1 across six replicas for each of the positive and neutral forms of the ILs, by selecting different lipids within the equilibrated bilayer for the umbrella sampling. For each system the PMF was smoothed over 11 adjacent values and then shifted to 0 energy at large bilayer-lipid separations.

Figure 1: Compositions and lipids used to create LNP bilayers. The four lipid compositions used in this study are based on published compositions of contemporary LNP systems[3], omitting the PEG-based lipids, increasing the amounts of the other three lipids proportionally to their mol% values in the full formulation and rounding off to integers for the actual numbers of lipids. CHOL and DSPC stand for cholesterol and distearoylphosphaticdylcholine. The *'s reflect the absence of mRNA and the PEG-based lipids in our calculations as well as the effect of using an integer number of each lipid in the calculation. The structures for the full ILs for each formulation are also depicted.

COMIRNATY® / ALC-0315*

Lipid	Mol %
ALC-0315	48
CHOL	42
DSPC	10

Lipid A

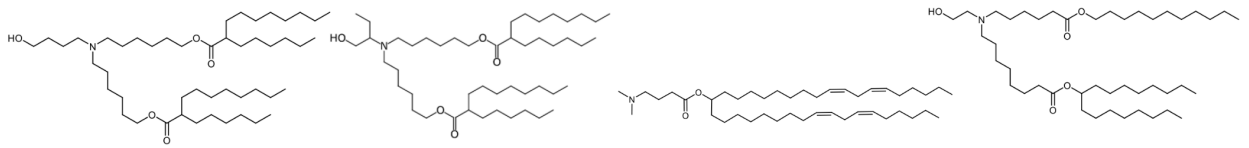
Lipid	Mol %
Lipid A	47.1
CHOL	43.1
DSPC	9.8

Onpattro® / MC3*

Lipid	Mol %
Din-MC3-DMA	49
CHOL	40.8
DSPC	10.2

SpikeVax® / SM-102*

Lipid	Mol %
SM-102	49
CHOL	40.8
DSPC	10.2



ALC-0315 Lipid A DLin-MC3-DMA SM-102

$\Delta \mathrm{p}K_a$ is calculated using:

$$\Delta \mathrm{p}K_a = \frac{k_B T}{2.303}\left\{\ln\left(< e^{-\frac{PMF(z)_{positive}}{k_B T}} >_z\right) - \ln\left(< e^{-\frac{PMF(z)_{neutral}}{k_B T}} >_z\right)\right\} \qquad (1)$$

Where the averages run over z ranges from 0 (the center of membrane) to where the PMF no-longer deviates from the bulk value (ALC-0315: 45.5 Å, for Lipid A: 49 Å, for MC3: 52 Å, for SM-102: 40 Å), k_B is the Boltzmann constant and T is the temperature in Kelvin. $\mathrm{p}{K_a}^A$, is given by:

$$\mathrm{p}{K_a}^A = \mathrm{p}{K_a}^S + \Delta\mathrm{p}K_a$$

Despite the simplicity of this methodology, the overall accuracy also relies on having a highly accurate $\mathrm{p}{K_a}^S$ value for each of the lipids. Unfortunately, due to both common lipid solubility challenges and the limited accuracy of $\mathrm{p}{K_a}^S$ prediction methods, care needs to be taken in selecting $\mathrm{p}{K_a}^S$ values. We used the recently created ML-based version of Epik[21] which has been extensively parametrized for a wide range of organic molecules in water to calculate $\mathrm{p}{K_a}^S$.

Results and Discussion

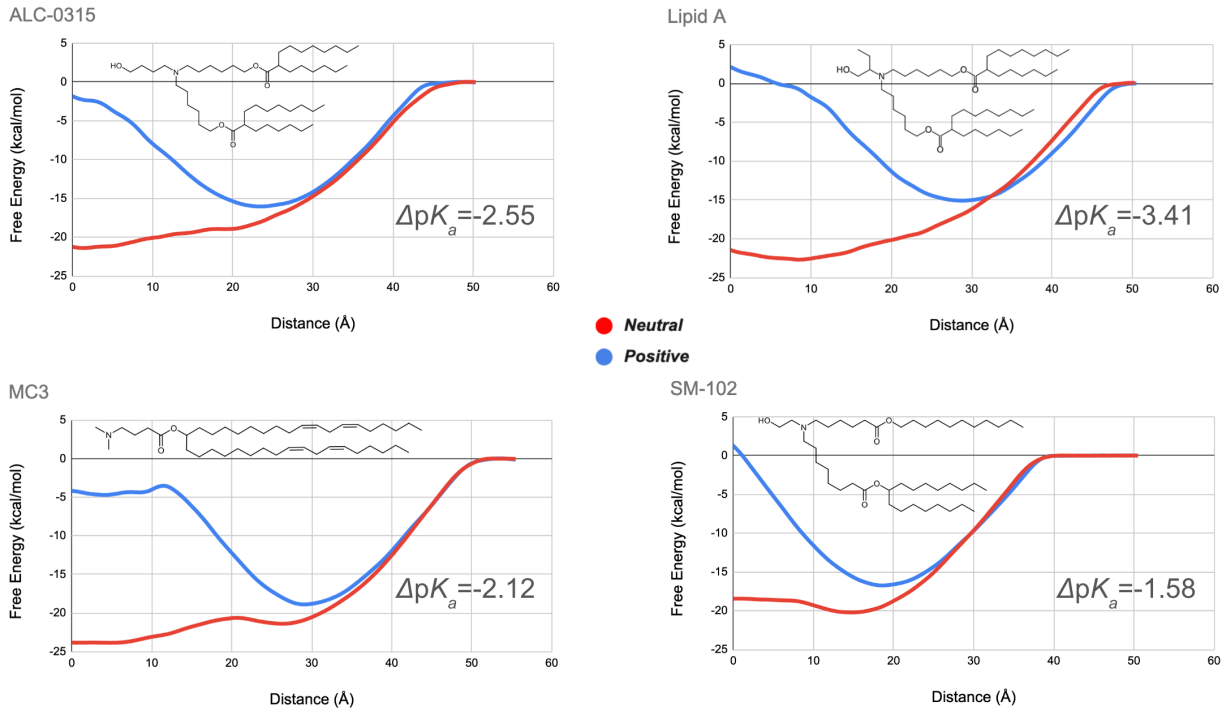


Figure 2*:* The PMF for charged (blue) and neutral (red) lipids, averaged over all 6 replicates relative to the center of the bilayer (Distance = 0) out into bulk aqueous solution, where these values become flat. The ΔpK_a calculated for each of the lipids is included for each formulation. The PMF plots for the replicates are provided in the Supporting Information.

As shown in Figure 2, the PMF for the neutral and charged lipids are quite different. The MC3 system required greater distances from the center of the bilayer in order to obtain flat free energy curves than the other 3 systems. All in their neutral form have broad, deep minima in the bilayer with MC3 having the strongest favorable free energy for integration of the lipid molecules. The positive forms have deep minima within the head-group regions in the bilayer for each formulation at distances ranging from 18 to 28 Å. At smaller distances the free energy rises approaching or exceeding that for bulk water (0 kcal/mol) in the center of the bilayer, reflecting the free energy cost for burying the charged headgroup inside a low-dielectric region. The

minimum for the positive form is deepest for MC3. These minima are weaker and narrower than the corresponding wide and flat low free energy regions for the neutral lipids in all cases. The stronger overall binding of the neutral forms leads to the effective drop in the $pK_a{}^A$ relative to the $pK_a{}^S$ value (i.e., a negative ΔpK_a value). For MC3, the effect of the strong minima largely cancels out yielding one of the smaller $|\Delta pK_a|$. Published MD simulations of bilayers containing ILs indicate that the neutral head groups exist and perhaps favor burial inside membranes while the positive head groups remain on the surface of lipid structures and retain their exposure to water[22-24]. The contrast between the positioning of the neutral and charged ILs seems most extreme for MC3 in our studies.

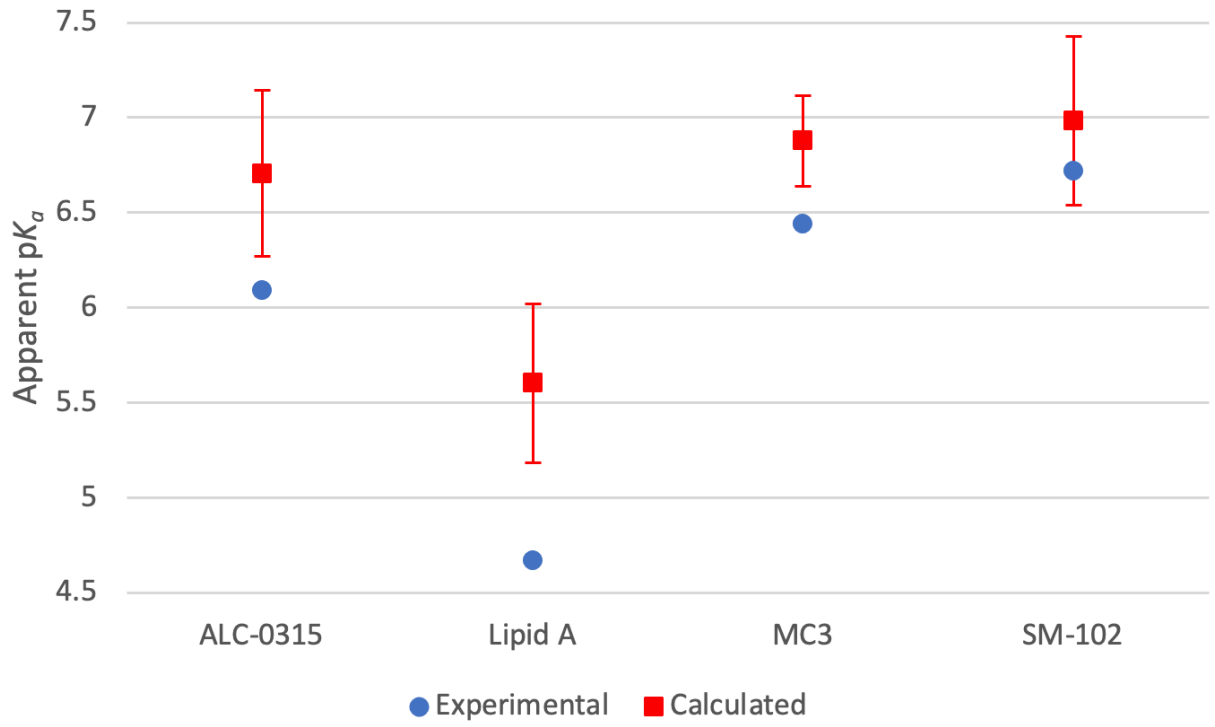


***Figure 3**: Calculated and experimental apparent pK_a value trends. The experimental values are shown in blue, while the calculated values are red. The error bars for the calculated values represent one standard deviation of the average.*

Table 1 contains the $pK_a{}^S$, ΔpK_a and $pK_a{}^A$ values for all 4 lipid systems. Interestingly for the three lipids used in therapeutics, the trend in the $pK_a{}^A$ values for both our calculations and the

experimental values orders ALC-0315 < MC3 < SM-102 which is the opposite of the trend for the pK_a^S values, i.e., the shift in the pK_a values is anticorrelated with the pK_a^S values for these formulations. Lipid A does not fit this trend having a mid-range pK_a^S value and the most negative ΔpK_a value yielding the lowest pK_a^A value.

Table 1. pK_a^S, ΔpK_a and pK_a^A values. Experiment refers to experimental pK_a^A values measured for lipid nanoparticles from the references provided.

Lipid / formulation	pK_a^S	ΔpK_a	pK_a^A	
			Calculated	Experiment
ALC-0315	9.26	-2.55	6.71	6.09[25]
Lipid A	9.01	-3.41	5.60	4.67[26]
MC3	9.00	-2.12	6.88	6.44[25]
SM-102	8.50	-1.58	6.98	6.75[25]

Figure 3 is a plot of the experimental and calculated pK_a^A values. Error bars were the standard deviation of the average value calculated from the spread in calculated ΔpK_a values for the six replicates. The size of the error bars could be reduced by adding more replicas or running longer umbrella sampling simulations. Overall, the trend in calculated as compared to experimental pK_a^A values is well reproduced with a R^2 value of 0.99 (for either the 3 commercialized formulations or all 4 formulations). The calculated values are somewhat higher and span a smaller range.

Conclusion

In summary, we have employed a methodology for predicting the pK_a shift for a lipid between bulk aqueous solution and a lipid bilayer as a stand-in for the environment within an LNP. This shift can be combined with the bulk aqueous solution pK_a value to yield usefully accurate apparent pK_a values for the IL for LNP formulations. To our knowledge, this work represents the first reported non-experimental methodology for calculating the apparent p$K_a$values of ILs in LNPs.

We plan to continue to expand and verify this new method with more subtle variations in lipid structure and also examine some of the underlying, collective structural features that influence the pK_a values in a future work. This methodology can be applied directly to new formulations and is expected to expedite the development of new nanoparticle systems for therapeutic delivery.

Supporting Information

The supporting information contains the PMF plots and calculated values for each of the individual simulations for all lipids. As well, the composition of the system for each lipid is provided.

References

(1) Polack, F. P.; Thomas, S. J.; Kitchin, N.; Absalon, J.; Gurtman, A.; Lockhart, S.; Perez, J. L.; Pérez Marc, G.; Moreira, E. D.; Zerbini, C.; et al. Safety and Efficacy of the BNT162b2 mRNA

Covid-19 Vaccine. *New England Journal of Medicine* **2020**, *383*, 2603-2615. DOI: 10.1056/NEJMoa2034577

(2) Wilson, B.; Geetha, K. M. Lipid nanoparticles in the development of mRNA vaccines for COVID-19. *J Drug Deliv Sci Technol* **2022**, *74*, 103553. DOI: 10.1016/j.jddst.2022.103553.

(3) Pfizer, Inc. Working To Make COVID-19 Vaccines Accessible to Everyone, Everywhere. https://www.pfizer.com/science/coronavirus/vaccine/working-to-reach-everyone-everywhere (accessed 2024 Mar. 21th)

(4) Swetha, K.; Kotla, N. G.; Tunki, L.; Jayaraj, A.; Bhargava, S. K.; Hu, H.; Bonam, S. R.; Kurapati, R. Recent Advances in the Lipid Nanoparticle-Mediated Delivery of mRNA Vaccines. *Vaccines (Basel)* **2023**, *11*, 658. DOI: 10.3390/vaccines11030658.

(5) Seo, Y.; Lim, H.; Park, H.; Yu, J.; An, J.; Yoo, H. Y.; Lee, T. Recent Progress of Lipid Nanoparticles-Based Lipophilic Drug Delivery: Focus on Surface Modifications. *Pharmaceutics* **2023**, *15*, 772. DOI: 10.3390/pharmaceutics15030772

(6) Deshpande, D.; Chhugani, K.; Chang, Y.; Karlsberg, A.; Loeffler, C.; Zhang, J.; Muszyńska, A.; Munteanu, V.; Yang, H.; Rotman, J.; Tao, L.; Balliu, B.; Tseng, E.; Eskin, E.; Zhao, F.; Mohammadi, P.; Łabaj, P. P.; , Mangul, S. RNA-seq data science: From raw data to effective interpretation. *Front. Genet.* **2023**, *14*, 997383. doi: 10.3389/fgene.2023.997383

(7) Shi, Y.; Zhen, X.; Zhang, Y.; Li, Y.; Koo, S.; Saiding, Q.; Kong, N.; Liu, G.; Chen, W.; Tao, W. Chemically Modified Platforms for Better RNA Therapeutics. *Chem. Rev.* **2024**, *124*, 929-1033. https://doi.org/10.1021/acs.chemrev.3c00611

(8) Nitika; Wei, J.; Hui, A. M. The Delivery of mRNA Vaccines for Therapeutics. *Life (Basel)* **2022**, *12*, 1254. DOI: 10.3390/life12081254

(9) Patel, P.; Ibrahim, N. M.; Cheng, K. The Importance of Apparent pK_a in the Development of Nanoparticles Encapsulating siRNA and mRNA. *Trends Pharmacol Sci* **2021**, *42*, 448-460. DOI: 10.1016/j.tips.2021.03.002

(10) Carrasco, M. J.; Alishetty, S.; Alameh, M.-G.; Said, H.; Wright, L.; Paige, M.; Soliman, O.; Weissman, D.; Cleveland, T. E.; Grishaev, A.; et al. Ionization and structural properties of mRNA lipid nanoparticles influence expression in intramuscular and intravascular administration. *Communications Biology* **2021**, *4*, 956. DOI: 10.1038/s42003-021-02441-2.

(11) Armacost, K. A.; Riniker, S.; Cournia, Z. Novel Directions in Free Energy Methods and Applications. *Journal of Chemical Information and Modeling* **2020**, *60*, 1-5. DOI: 10.1021/acs.jcim.9b01174.

(12) Patey, G. N.; Valleau, J. P. A Monte Carlo method for obtaining the interionic potential of mean force in ionic solution. *Journal of Chemical Physics* **2008**, *63*, 2334-2339. DOI: 10.1063/1.431685.

(13) Roux, B. The calculation of the potential of mean force using computer simulations. *Computer Physics Communications* **1995**, *91*, 275-282. DOI: 10.1016/0010-4655(95)00053-I

(14) Schrödinger Release 2023-2: Materials Science Suite; 2023.

(15) Martínez, L.; Andrade, R.; Birgin, E. G.; Martínez, J. M. PACKMOL: a package for building initial configurations for molecular dynamics simulations. *J Comput Chem* **2009**, *30*, 2157-2164. DOI: 10.1002/jcc.21224

(16) Suzuki, T.; Suzuki, Y.; Hihara, T.; Kubara, K.; Kondo, K.; Hyodo, K.; Yamazaki, K.; Ishida, T.; Ishihara, H. PEG shedding-rate-dependent blood clearance of PEGylated lipid nanoparticles in mice: Faster PEG shedding attenuates anti-PEG IgM production. *International Journal of Pharmaceutics* **2020**, *588*, 119792. DOI: 10.1016/j.ijpharm.2020.119792.

(17) Schrödinger, L., New York, NY. Schrödinger Release 2023-2: Desmond. 2023.

(18) Proceedings of the 2006 ACM/IEEE conference on Supercomputing. Tampa, Florida, 2006; Association for Computing Machinery.

(19) Schrödinger, L., New York, NY. Schrödinger Release 2023-2: Maestro. 2023.

(20) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *Journal of Computational Chemistry* **1992**, *13*, 1011-1021. DOI:10.1002/jcc.540130812.

(21) Johnston, R. C.; Yao, K.; Kaplan, Z.; Chelliah, M.; Leswing, K.; Seekins, S.; Watts, S.; Calkins, D.; Chief Elk, J.; Jerome, S. V.; et al. Epik: pK_a and Protonation State Prediction through Machine Learning. *Journal of Chemical Theory and Computation* **2023**, *19*, 2380-2388. DOI: 10.1021/acs.jctc.3c00044.

(22) Ramezanpour, M.; Schmidt, M. L.; Bodnariuc, I.; Kulkarni, J. A.; Leung, S. S. W.; Cullis, P. R.; Thewalt, J. L.; Tieleman, D. P. Ionizable amino lipid interactions with POPC: implications for lipid nanoparticle function. *Nanoscale* **2019**, *11*, 14141-14146, 10.1039/C9NR02297J. DOI: 10.1039/C9NR02297J.

(23) Trollmann, M. F. W.; Böckmann, R. A. mRNA lipid nanoparticle phase transition. *Biophys J* **2022**, *121*, 3927-3939. DOI: 10.1016/j.bpj.2022.08.037.

(24) Paloncýová, M.; Šrejber, M.; Čechová, P.; Kührová, P.; Zaoral, F.; Otyepka, M. Atomistic Insights into Organization of RNA-Loaded Lipid Nanoparticles. *J Phys Chem B* **2023**, *127*, 1158-1166. DOI: 10.1021/acs.jpcb.2c07671.

(25) Zhang, C.; Ma, Y.; Zhang, J.; Chun-Tien Kuo, J.; Zhang, Z.; Xie, H.; Zhu, J.; Liu, T. Modification of Lipid-Based Nanoparticles: An Efficient Delivery System for Nucleic Acid-Based Immunotherapy. *Molecules* **2022**, *27*, 1943. DOI: 10.3390/molecules27061943

(26) The pK_a value for Lipid A was measured by Acuitas using a TNS fluorescence assay.

Graphical Abstract