Computational and Structural Biotechnology Journal 25 (2024) 3–8

Contents lists available at ScienceDirect

Computational and Structural Biotechnology Journal

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

Research Article

Predicting zeta potential of liposomes from their structure: A nano-QSPR model for DOPE, DC-Chol, DOTAP, and EPC formulations

Kamila Jarzynska [a,1], Agnieszka Gajewicz-Skretna [a,2], Krzesimir Ciura [a,b,*,3], Tomasz Puzyn [a,**,4]

[a] Laboratory of Environmental Chemoinformatics, Faculty of Chemistry, University of Gdansk, Wita Stwosza 63, 80-308 Gdansk, Poland
[b] Department of Physical Chemistry, Faculty of Pharmacy, Medical University of Gdansk, Gen. Hallera 107, 80-416 Gdańsk, Poland

ARTICLE INFO

Keywords:
Liposomes
Zeta potential
Nano-QSPR
KwLPR
PCA biplot
Nanodescriptors
Nanocarriers

ABSTRACT

Liposomes, nanoscale spherical structures composed of amphiphilic lipids, hold great promise for various pharmaceutical applications, especially as nanocarriers in targeted drug delivery, due to their biocompatibility, biodegradability, and low immunogenicity. Understanding the factors influencing their physicochemical properties is crucial for designing and optimizing liposomes. In this study, we have presented the kernel-weighted local polynomial regression (KwLPR) nano-quantitative structure-property relationships (nano-QSPR) model to predict the zeta potential (ZP) based on the structure of 12 liposome formulations, including 1,2-dioleoyl-sn-glycero-3-phosphoethanolamine (DOPE), 3ß-[N-(N′,N′-dimethylaminoethane)-carbamoyl]cholesterol (DC-Chol), 1,2-dioleoyl-3-trimethylammonium-propane (DOTAP), and L-α-phosphatidylcholine (EPC). The developed model is well-fitted ($R^2 = 0.96$, $RMSE_C = 5.76$), flexible ($Q^2_{CVloo} = 0.83$, $RMSE_{CVloo} = 10.77$), and reliable ($Q^2_{Ext} = 0.89$ $RMSE_{Ext} = 5.17$). Furthermore, we have established the formula for computing molecular nanodescriptors for liposomes, based on constituent lipids' molar fractions. Through the correlation matrix and principal component analysis (PCA), we have identified two key structural features affecting liposomes' zeta potential: hydrophilic-lipophilic balance (HLB) and enthalpy of formation. Lower HLB values, indicating a more lipophilic nature, are associated with a higher zeta potential, and thus stability. Higher enthalpy of formation reflects reduced zeta potential and decreased stability of liposomes. We have demonstrated that the nano-QSPR approach allows for a better understanding of how the composition and molecular structure of liposomes affect their zeta potential, filling a gap in ZP nano-QSPR modeling methodologies for nanomaterials (NMs). The proposed proof-of-concept study is the first step in developing a comprehensive and computationally based system for predicting the physicochemical properties of liposomes as one of the most important drug nano-vehicles.

1. Introduction

Liposomes, discovered in the 1960 s by the British hematologist Dr. Alec D. Bangham [1] are colloidal spherical structures with a diameter between 15 and 100 nm – so classified as nanomaterials (NMs) – composed of amphiphilic lipids molecules, such as phospholipids, that self-assemble in solution [2,3]. The liposomal membrane consists of a lipid bilayer organized around an inner aqueous core, with polar groups (heads) facing outwards and hydrophobic groups (tails) oriented inwards to the membrane [2]. Due to their biocompatibility, biodegradability and low immunogenicity, liposomes are increasingly being considered for various pharmaceutical applications, such as targeted

Abbreviations: DOPE, 1,2-dioleoyl-sn-glycero-3-phosphoethanolamine; DC-Chol, 3ß-[N-(N′,N′-dimethylaminoethane)-carbamoyl]cholesterol; DOTAP, 1,2-dioleoyl-3-trimethylammonium-propane; EPC, L-α-phosphatidylcholine; KwLPR, kernel-weighted local polynomial regression; QSAR/QSPR, quantitative structure-activity/property relationships; HLB, hydrophilic-lipophilic balance; NMs, nanomaterials; ZP, zeta potential; AOP, adverse outcome pathways; DoE, design of experiments.

* Corresponding author at: Laboratory of Environmental Chemoinformatics, Faculty of Chemistry, University of Gdansk, Wita Stwosza 63, 80-308 Gdansk, Poland
** Corresponding author.
E-mail addresses: krzesimir.ciura@gumed.edu.pl (K. Ciura), tomasz.puzyn@ug.edu.pl (T. Puzyn).
[1] orcid.org/0000-0002-2896-0463
[2] orcid.org/0000-0001-7702-210X
[3] orcid.org/0000-0001-6187-6039
[4] orcid.org/0000-0003-0449-8339

https://doi.org/10.1016/j.csbj.2024.01.012
Received 27 November 2023; Received in revised form 17 January 2024; Accepted 19 January 2024
Available online 24 January 2024

drug nanocarriers, medical imaging, vaccine carriers, and cosmetics. Their primary role is to improve the bioavailability of poorly soluble medical substances, but they can also increase the cellular uptake of drugs with a low ability to penetrate biological membranes [4,5]. The first milestone in liposome-based nanomedicine was the introduction of Doxil® to the U.S. market in 1995 (and still extensively used clinically) for treating ovarian cancer and AIDS-related Kaposi's sarcoma [6]. Moreover, liposomes can be adjusted to the applications – their safety, stability, and efficiency are determined by: (i) the selection of lipids, head groups, chain lengths, and the ratio of liposome components, (ii) the size, surface charge, lipids organization, and surface modification [2].

In the case of colloidal suspensions, like liposomes, the general particle charge indicator that provides the initial information on whether the liposome is stable or not is the zeta potential (ZP). The zeta potential is defined as the electric potential at the shear plane, which is an imaginary surface separating the thin liquid layer (constituted of counter-ions) bound to the solid surface in motion. It is currently assumed that zeta potential over |30| mV (optimum > |60|) is required for total electrostatic stabilization, from |5| to |15| mV occurs with limited flocculation, and between |5| and |3| mV – maximum flocculation. The greater the ZP the more likely the suspension is to be stable because the charged particles repel one another and thus overcome the natural tendency to aggregate [7].

Developing effective nano-vehicles requires a broad understanding of their interactions with the biological environment. Likewise, nanotoxicology studies the impact of nanomaterials on living organisms to determine and recognize their toxicity. Bondarenko et al. [8] emphasize that these nanotechnology sub-disciplines, nanomedicine and nanotoxicology, share overlapping interests and challenges. Nanomedicine can apply existing knowledge and methodology to design safer, more effective, and stable drug formulations on the nanoscale.

One of the essential nanotoxicological approaches, utilizing machine learning and artificial intelligence, is nano-quantitative structure-activity/property relationships (nano-QSAR/QSPR). Nano-QSAR/QSPR aims to use experimental data to obtain a model based on nanomaterial descriptors (nanodescriptors) numerically expressing the variability of nanoforms' structure to predict the physicochemical properties or biological activity of NMs. An established nano-QSAR/QSPR model allows for identifying the most significant structural features affecting the properties of nanomaterials and determining their potential toxicity mechanisms [9,10]. Puzyn et al. [11] proposed the first nano-QSAR model to predict the cytotoxicity of metal oxide nanomaterials toward *Escherichia coli.* Since then, this methodology has been widely used to evaluate the critical physicochemical properties of nanoforms (including nano-mixtures) [12,13], their impact on the environment [14,15], or assess the risk at different levels – from in vitro tests to the prediction of doses activating key events in adverse outcome pathways (AOP) [16–19]. Recently, similarity-based techniques such as quantitative read-across structure-activity relationships (q-RASAR) have gained popularity. This method applies chemical similarity concepts of read-across in an unsupervised step to compute nanodescriptors and then develops a supervised learning model, especially useful for limited data of NMs [20–23].

Considering our previous experience in modeling the zeta potential of metal oxide nanoparticles [12,24,25] and polymeric nanomaterials [26], in the proof-of-the-concept presented here, we have implemented the nano-QSPR method to predict the ZP of liposomes consisting of 1, 2-dioleoyl-sn-glycero-3-phosphoethanolamine (DOPE), 3ß-[N-(N′, N′-dimethylaminoethane)-carbamoyl]cholesterol (DC-Chol), 1,2-dioleoyl-3-trimethylammonium-propane (DOTAP), and L-α-phosphatidylcholine (EPC). We have employed the kernel-weighted local polynomial regression (KwLPR), as this method solves several problems of modeling small datasets typically encountered in nanotechnology-related research [27].

To the best of our knowledge, there have been no attempts to model the zeta potential of liposomes using nano-QSPR. Although the design of experiments (DoE) approach has been used to optimize the composition of liposomes to achieve adequate zeta potential value [28], there is no information on how lipids' molecular structure affects this endpoint. Therefore, we have adopted molecular nanodescriptors, originally dedicated to small-molecule compounds, to quantitatively describe liposomes based on constituent lipids' molar fractions. Since the slow development of interpretable nanomaterial descriptors (especially for more complex structures) has been identified as an important feature limiting the achievement of nanoinformatics milestones [29,30], our nanodescriptors are intuitive and easily interpretable. Thus, the established nano-QSPR model makes it possible to understand interactions between the composition of liposomes and their properties at the next, more advanced molecular level.

In this study, we have proposed a workflow dedicated to liposomes as one of the most important nanocarriers, opening new opportunities for developing computational tools for nanomedical applications using the well-known nanotoxicological method, nano-QSAR/QSPR.

2. Results and discussion

The composition of 12 liposome formulations and their ZP values are shown in Table 1. As this is a proof-of-concept study, the data came from a single publication, performed in the same laboratory by the same group, following a single protocol [28]. Conducting experiments under differing conditions could introduce a substantial source of uncertainty in the results [9].

We have adopted molecular descriptors originally dedicated to small-molecule compounds to compute liposomes' nanodescriptors based on constituent's lipids molar fractions (details of the established formula are given in the Materials and Methods section). Thus, each liposome was described by physicochemical (PCH), geometrical (G), hydrogen bonding (H), and quantum-mechanical descriptors (QM). Additionally, the effect of peptide concentration (the delivery substance) was evaluated as a potential factor affecting zeta potential. All nanodescriptors and their notations are summarized in Table 2.

Along with the standardized correlation matrix analysis (Supplementary Information, Fig. S1), the factor loadings gained from the principal component analysis (PCA) were applied to the feature selection.

The combination of PC1 and PC2 failed to capture any interesting trends (Supplementary Information, Fig. S2), while the PC1-PC3 biplot (explaining in total 58.8% of data variance) proved the relevance of certain nanodescriptors in these PCs, as well as in determining the zeta potential value.

As shown in Fig. 1, both principal components differentiated the liposomes in terms of zeta potential values. Liposomes with the lowest ZP values are in the upper left corner of the PCA biplot (high PC3 values),

Table 1
Characterization of 12 liposomes used for the development of the nano-QSPR model.

Formulation	Zeta potential [mV]	Lipids				Peptide concentration [μg/mL]
		DOPE	DC-Chol	DOTAP	EPC	
1	72	0	0	1	0	10
2	70	0	1	0	0	100
3	69	0	0	1	0	100
4	66	0.33	0.33	0.33	0	100
5	64	0	1	0	0	10
6	64	0	0.33	0.33	0.33	10
7	61	0.25	0.25	0.25	0.25	55
8	57	0.33	0	0.33	0.33	100
9	41	0.33	0.33	0	0.33	100
10	30.3	0.03	0.07	0.13	0.77	13.4
11	7	0	0	0	1	10
12	1	0	0	0	1	100

Table 2
Notation of 24 nanodescriptors considered in the nano-QSPR modeling.

Notation	Nanodescriptor	Notation	Nanodescriptor
PCH1	Molar refractivity [cm^3/mol]	**H1**	Hydrogen bond donor count
PCH2	logD	**H2**	Hydrogen bond acceptor count
PCH3	Hydrophilic-lipophilic balance (HLB)	**QM1**	HOMO [eV]
PCH4	logP	**QM2**	LUMO [eV]
G1	Van der Waals volume [A^3]	**QM3**	Ionization potential [eV]
G2	Van der Waals surface area [A^2]	**QM4**	Electron affinity [eV]
G3	Solvent accessible surface area [A^2]	**QM5**	Hardness [eV]
G4	Topological polar surface area [A^2]	**QM6**	Softness [eV]
G5	Minimum projection area [A^2]	**QM7**	Electronegativity [eV]
G6	Maximum projection area [A^2]	**QM8**	Polarizability [au]
G7	Minimum projection radius [A]	**QM9**	Enthalpy of formation [Hartree]
G8	Maximum projection radius [A]	**PEPTIDE**	Peptide concentration [μg/mL]

while those with the highest ZP values are in the bottom part with low values of the third principal component. The nanodescriptors with the highest contribution, defined by the value of normalized factor loading in PC3, were: (i) PCH3 – hydrophilic-lipophilic balance (HLB) and (ii) QM9 – enthalpy of formation, so both were selected as independent variables to quantify the relationship between liposome molecular structure and ZP.

Based on selected nanodescriptors, we have developed the KwLPR nano-QSPR model quantifying relationships between the liposome's structure and its zeta potential.

Close to unity values of the determination ($R^2 = 0.96$), leave-one-out cross-validation ($Q^2_{CVloo} = 0.83$), and external validation coefficients ($Q^2_{Ext} = 0.89$), and small differences between R^2 and Q^2_{CVloo} proved that the nano-QSPR model is well-fitted, flexible, stable and reliable. Similar values of root mean squared calibration error ($RMSE_C = 5.76$), root mean squared error of leave-one-out cross-validation ($RMSE_{CVloo} = 10.77$), and root mean squared error of external validation ($RMSE_{Ext} = 5.17$) indicated the model's ability to generalize information.

The plot of experimental versus predicted values (Fig. 2a) showed a high correlation between observed and predicted zeta potential values of the liposomes, in both training and validation sets.

The range of applicability was evaluated using a Williams plot (Fig. 2b). No residual value exceeded a value of 3 standard deviations from the residual average. In addition, none of the liposomes were significantly different from compounds in the training set – all were characterized by leverage values $h < h^* = 1$. The results demonstrated high prediction precision of the nano-QSPR model, so it can be used to predict the zeta potential of both liposomes from the dataset and new lipid compositions.

To show the independence of the results from the split, we also tested the 1:4 splitting algorithm (Supporting Information, Table S1, Fig. S3).

In preliminary studies, we have also considered simpler regression methods like linear regression (LR), partial least squares regression (PLS), and distance-weighted *k*-nearest neighbors' algorithm (k-NN), but their statistics did not meet the nano-QSPR modeling requirements (Supplementary Information, Table S2). Additionally, other methods such as the read-across and q-RASAR [21,22,31], which have recently been used to analyze limited datasets of NMs, were tested. However, KwLPR has outperformed all other approaches in terms of performance statistics.

As a non-parametric method, the KwLPR algorithm does not offer a single global model, but the interpretation of the PCA biplot confirmed the described relationships between structure and ZP. The results suggest that liposomes with lower HLB values, pointing to a more lipophilic nature [32], tend to have higher zeta potential, implying increased stability. In turn, enthalpy of formation refers to heat absorbed during liposome formation. Higher enthalpy makes the liposome a higher-energy compound and, thereby, more reactive. Consequently, liposomes characterized by high enthalpy are inherently less stable and have lower zeta potential. No significant correlation was found between

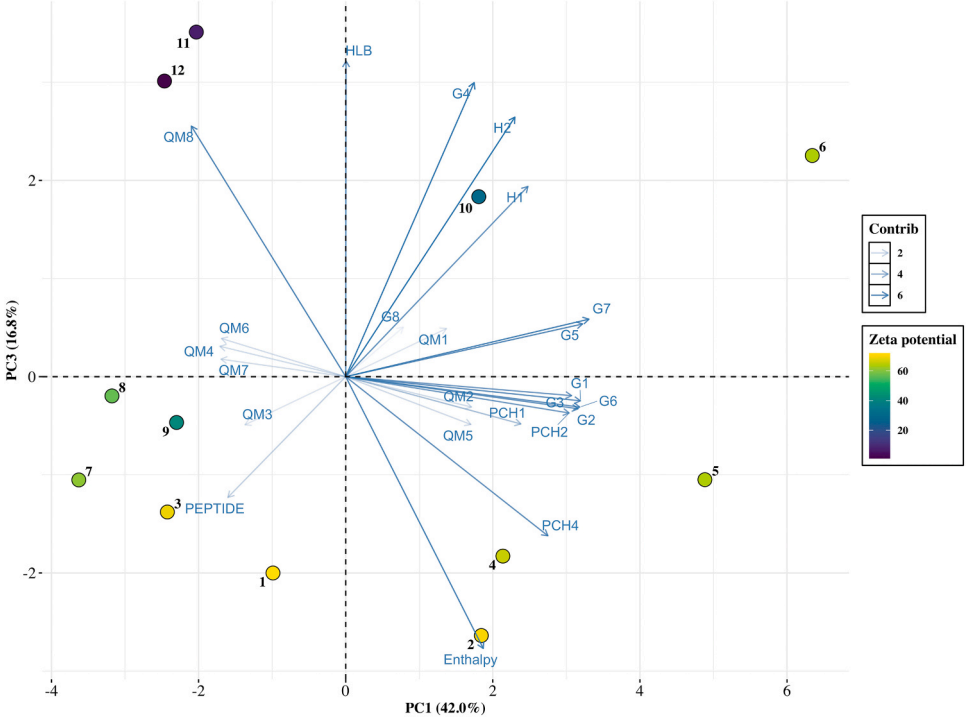

Fig. 1. PCA biplot for the first and third principal components.

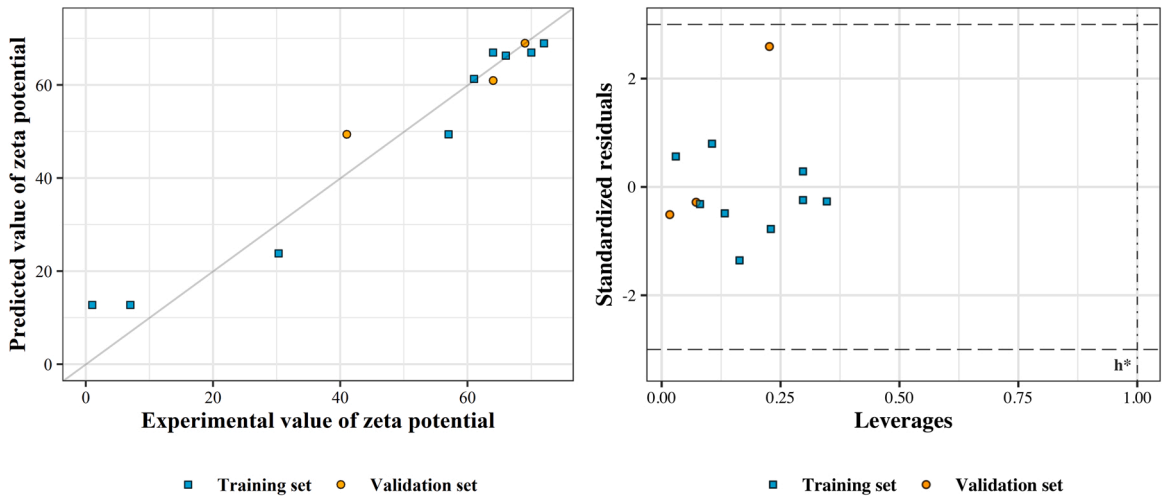

Fig. 2. (a) Scatter plot of experimentally determined versus predicted zeta potential values for training and validation compounds of the nano-QSPR model; (b) Williams plot illustrating the applicability domain of the nano-QSPR model.

selected nanodescriptors, which proved that they reflect different aspects of liposome behavior and do not show a direct relationship. As expected [28], the peptide is not an important nanodescriptor in determining the zeta potential of liposomes.

The study of Soema et al. [28] showed that cationic lipids (DOTAP and/or DC-Chol) in liposome composition contribute to an increase in zeta potential value, while their absence/addition of EPC (zwitterionic lipid) causes its decrease. Extending those findings with the knowledge obtained from this research leads to the conclusions that: (i) liposomes containing cationic lipids in their structures, showing significant lipophilicity and low enthalpies of formation, have high zeta potential, and thus stability, (ii) liposomes lacking cationic lipids/consisted of zwitterionic ones, exhibiting high hydrophilicity and enthalpies of formation, have low zeta potential values (so stability).

To date, nano-QSPR modeling of zeta potential has been mainly concentrated on metal oxide NMs generally considered more straightforward to characterize than soft NMs, which possess complex structure, composition, and dynamic behavior. Metal oxide NMs' modeling often relies on nanodescriptors derived from their simple properties like enthalpy of formation [33,34], electronegativity [34] ionic radius, a charge of metal ion, the electronegativity of metal, etc. [15]. Additionally, DFT calculations using molecular nanoclusters are commonly used to characterize metal-based NMs and widely accepted for building nano-QSAR/QSPR models [12,33,34]. However, this approach cannot be directly applied to soft NMs due to inherent differences in their nature.

Within the realm of soft NMs, the development of nano-QSPR methodology has been limited to just one model specifically designed to predict the zeta potential of polymeric NMs [26], highlighting the shortcomings of this field. Moreover, that model has its limitations, i.e., it reflects core and coating to be made of one component, while liposomes are usually multi-component NMs. In this context, our study is pioneering as it addresses the application of the nano-QSPR approach to model the zeta potential of liposomes.

While previous investigations have mainly focused on optimizing the composition of liposome formulations and their effect on ZP (by DoE approach) [28], we have taken a step further by recognizing the influence of the liposome's molecular features on the modeled property through their quantitative description by nano-QSPR modeling. The obtained nano-QSPR model shows that *in silico* methods can be successfully applied to assess the stability (expressed by ZP) of liposomes. However, it should be acknowledged that a larger dataset would be required to broaden its range of applications and improve the reliability and confidence of predictions even if the proposed model met statistical criteria. Nevertheless, our findings fill a critical gap in the existing nano-QSPR methodology for modeling the zeta potential of NMs and open the door to a deeper understanding of liposomes' physicochemical properties, contributing to advances in the design, characterization, and application of the most used targeted drug nano-vehicles.

3. Conclusions

In this proof-of-concept study, we successfully quantified the relationship between liposome molecular structure and its zeta potential using the KwLPR nano-QSPR approach, overcoming the challenges of limited datasets in nanotechnology research. The performance characteristics of the developed nano-QSPR model demonstrate that it is well-fitted (R^2 = 0.96, $RMSE_C$ = 5.76), stable (Q^2_{CVloo} = 0.83, $RMSE_{CVloo}$ = 10.77), and able to generalize information (Q^2_{Ext} = 0.89 $RMSE_{Ext}$ = 5.17). We have established a simple, intuitive method to compute molecular nanodescriptors for liposomes based on constituent lipids' molar fractions. The most important of these, HLB and enthalpy of formation, were identified based on the correlation matrix and PCA biplot interpretation. We indicated that liposomes with lower HLB values (more lipophilic) are correlated with higher ZP, and thus stability. In turn, an increase in enthalpy causes a decrease in ZP and stability. Based on initial findings [28], the desired stable structure of liposome (liposome nano-vehicle) should consist of cationic lipids, be strongly lipophilic, and have a low enthalpy of formation value. The presented methodology states a computationally-based framework for modeling liposomes' physiochemical properties, such as zeta potential, filling a gap in ZP nano-QSPR modeling methodologies for NMs. We have demonstrated that our approach enhances understanding of factors affecting properties and enables more efficient optimization of liposomes. Our findings significantly contribute to the development of computational tools for the design of liposome nano-vehicles, which is particularly important in nanomedicine.

4. Materials and methods

4.1. Dataset

We used empirically measured zeta potential values for 12 liposomes from the literature [28]. The pH of the experiment was 7.4. Each liposome was composed of one to four different lipids: 1,2-dioleoyl-sn-glycero-3-phosphoethanolamine (DOPE), 3ß-[N-(N′,N′-dimethylaminoethane)-carbamoyl] cholesterol (DC-Chol), 1,2-dioleoyl-3-trimethylammonium-propane (DOTAP), and L-α-phosphatidylcholine (EPC), in various molar fractions,

with the addition of HLA-A2.1-restricted influenza peptide GILGFVFTL (M158–66). Formulations containing only DOPE do not form liposomes, so they were not included in the dataset [28].

The dataset was ranked to decrease the value of zeta potential and then divided into training and validation sets according to the 1:3 (T:V) splitting algorithm.

4.2. Nanodescriptors computation

To describe a whole liposome, we have developed the established formula (Eq. 1) that involves multiplying the value of a lipid's descriptor by its molar fraction and then summing the values thus obtained for all lipids which build the liposome, resulting in the one nanodescriptor representing the structure of the liposome. We treated the peptide concentration as a separate, additional nanodescriptor. There were, in total, 24 nanodescriptors representing the properties of liposomes.

$$\boldsymbol{D_L} = \sum_{i}^{n} d_i \frac{n_i}{\sum n_i} \tag{1}$$

where: D_L denotes the nanodescriptor of the whole liposome, n is the number of lipids forming the liposome, and d_i represents the descriptor for the i-th lipid.

Each liposome was described using two sets of molecular nanodescriptors. The first set included 14 physicochemical, geometrical, and hydrogen bonding nanodescriptors, achieved using CHEMICALIZE (Csizmadia, P. & F., Hungary) [35] software. The second set, which consisted of 9 quantum-mechanical nanodescriptors, was computed using the GAUSSIAN 16 (Frisch, M., U.S.) [36] software based on the pre-optimized geometry of molecular structures via the Density Functional Theory (DFT) method with hybrid density functional B3LYP with the 6–31 G basis set (i.e., B3LYP/6–31 G++) [37]. The molecular structures of the lipids were generated using MOLDEN 6.1 (Schaftenaar, G., Holland) [38] molecular editor.

Data autoscaling (standardizing) was performed to guarantee that all nanodescriptors are equal and gain the same resources of variability in revealing high-dimensional associations between these variables and zeta potential. That means that they were transformed so that the average value of each was equal to 0 and the standard deviation was equal to one. In this way, we ensure an equal scale and range of all variables [39]. A full list of computed molecular nanodescriptors' values can be found in Supplementary Information, Table S3.

4.3. Data analysis and nano-QSPR model development

We performed the principal component analysis to verify the homogeneity of the dataset and find trends and patterns within the data. The results of the PCA were visualized as a biplot, which yields mechanistic insight into the relationships between descriptors and links them with modeled quantity. It allows not only to obtain information on whether the correlation between original variables and the principal component (PC) is positive or negative but also introduces the strength of this correlation (expressed by vector length) and the degree of correlation among the variables (expressed by angles between vectors: an adjacent angle – high positive, a straight angle – high negative, and a right angle – no correlation between two variables) [27].

To overcome data shortage and deliver the best possible, in terms of accuracy and predictive ability of the nano-QSPR model for predicting the zeta potential of liposomes, the kernel-weighted local polynomial regression (KwLPR) approach, was applied. The KwLPR approach was purposefully developed to deal with the limited training data set. As discussed elsewhere [27,40], an integral part of the KwLPR algorithm is to use only the most similar compounds to directly estimate the property of interest (here, the zeta potential of liposomes). Since the KwLPR is a single-model equation-free approach, mechanistic interpretation should be based on the PCA biplot analysis [27]. The nano-QSPR model was validated against the Organization for Economic Cooperation and Development (OECD) standard principles [41]. Multiple classical rigorous metrics were used to check the model's goodness-of-fit (determination coefficient R^2, root mean square calibration error $RMSE_C$), robustness (leave-one-out cross-validation coefficient Q^2_{CVloo}, root mean square error of leave-one-out cross-validation $RMSE_{CVloo}$) and predictive capability (external validation coefficient Q^2_{Ext}, root mean square error of prediction $RMSE_{Ext}$) [39]. All statistical values were calculated according to the formulas summarized in Supporting Information, Fig. S4.

In essence, a good nano-QSAR/QSPR model should be characterized by values of R^2, Q^2_{CVloo}, and Q^2_{Ext} as close to one as possible, and comparable and as small as possible values of mean squares errors. Among these metrics, error-based are more important, because the occurrence of significant differences between their values indicates that nanoforms belonging to the training set are overfitted, and thus model's low ability to generalize information [39].

The nano-QSAR/QSPR model's applicability domain (AD) describes the boundaries of the theoretical area in the feature space of the nanomaterials in which the predictions are reliable. This space is determined by plotting the leverage coefficients h ($h = \boldsymbol{X}^T(\boldsymbol{X}^T\boldsymbol{X})$) on the abscissa axis and the values of standardized residuals $y_{pred} - y_{obs}$ on the ordinate axis (Williams plot). Leverage coefficients reveal the structural similarity of NMs to the training set. The critical value of leverage is expressed by the formula: $h^* = 3pn^{-1}$, where p is the number of descriptors in the model and n is the number of compounds in the training set. Nanoforms with values greater than h^* are considered as lying outside the AD. The ordinate axis represents the model's prediction precision with standardized residuals expected to fall within 3 standard deviations. Defining the AD allows for mapping the error's surface based on descriptors used and evaluating model quality [39].

4.4. Software

All analyses were performed using R Statistical Software (v4.2.2, R Core Team, 2022) [42]. The correlation matrix was visualized via corrplor R package [43]. The KwLPR nano-QSPR modeling was performed using KwLPR.Rmd script [27]. Scatter plots of experimentally determined versus predicted zeta potential values and Williams plots were created using ggplot R package [44]. PCA analysis was performed and visualized using factoextra [45], factoMineR [46], and ggrepel [47] R packages.

Funding

This work was funded by Poland National Science Centre within the project NanoCARRIERS (ID: 2021/40/Q/ST5/00117).

CRediT authorship contribution statement

Kamila Jarzynska – conceptualization, methodology, formal analysis, visualization, writing-original draft **Agnieszka Gajewicz-Skretna** – methodology, edition of the manuscript. **Krzesimir Ciura** – conceptualization, supervision, edition of the manuscript. **Tomasz Puzyn** – conceptualization, supervision, edition of the manuscript.

Declaration of Competing Interest

The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Acknowledgment

Computations in GAUSSIAN 09 were carried out using the Centre of Informatics Tricity Academic Supercomputer & Network computers.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.csbj.2024.01.012.

References

[1] Bangham AD, Horne RW. Negative staining of phospholipids and their structural modification by surface-active agents as observed in the electron microscope. J Mol Biol 1964;8(5):660–8. https://doi.org/10.1016/S0022-2836(64)80115-7.
[2] Guimarães D, Cavaco-Paulo A, Nogueira E. Design of liposomes as drug delivery system for therapeutic applications. Elsevier B.V. May 15 Int J Pharm 2021. https://doi.org/10.1016/j.ijpharm.2021.120571.
[3] Chibowski E, Szcześ A. Zeta potential and surface charge of DPPC and DOPC liposomes in the presence of PLC enzyme. Adsorption 2016;22(4–6):755–65. https://doi.org/10.1007/s10450-016-9767-z.
[4] Li M, Du C, Guo N, Teng Y, Meng X, Sun H, et al. Composition design and medical application of liposomes. Elsevier Masson s.r.l. February 15 Eur J Med Chem 2019: 640–53. https://doi.org/10.1016/j.ejmech.2019.01.007.
[5] Nsairat H, Khater D, Sayed U, Odeh F, Al Bawab A, Alshaer W. Liposomes: structure, composition, types, and clinical applications. Elsevier Ltd May 1 Heliyon 2022. https://doi.org/10.1016/j.heliyon.2022.e09394.
[6] Bulbake U, Doppalapudi S, Kommineni N, Khan W. Liposomal formulations in clinical use: an updated review. MDPI AG June 1 Pharmaceutics 2017. https://doi.org/10.3390/pharmaceutics9020012.
[7] Heurtault B, Saulnier P, Pech B, Proust JE, Benoit JP. Physico-chemical stability of colloidal lipid particles (Elsevier BV) Biomaterials 2003:4283–300. https://doi.org/10.1016/S0142-9612(03)00331-4.
[8] Bondarenko O, Mortimer M, Kahru A, Feliu N, Javed I, Kakinen A, et al. Nanotoxicology and nanomedicine: the Yin and Yang of nano-bio interactions for the new decade. Elsevier B.V. August 1 Nano Today 2021. https://doi.org/10.1016/j.nantod.2021.101184.
[9] Wyrzykowska E, Mikolajczyk A, Lynch I, Jeliazkova N, Kochev N, Sarimveis H, et al. Representing and describing nanomaterials in predictive nanoinformatics. Nature Research September 1 Nat Nanotechnol 2022:924–32. https://doi.org/10.1038/s41565-022-01173-6.
[10] Yan X, Yue T, Winkler DA, Yin Y, Zhu H, Jiang G, et al. Converting nanotoxicity data to information using artificial intelligence and simulation. American Chemical Society July 12 Chem Rev 2023:8575–637. https://doi.org/10.1021/acs.chemrev.3c00070.
[11] Puzyn T, Rasulev B, Gajewicz A, Hu X, Dasari TP, Michalkova A, et al. Using nano-QSAR to predict the cytotoxicity of metal oxide nanoparticles. Nat Nanotechnol 2011;6(3):175–8. https://doi.org/10.1038/nnano.2011.10.
[12] Mikolajczyk A, Gajewicz A, Rasulev B, Schaeublin N, Maurer-Gardner E, Hussain S, et al. Zeta potential for metal oxide nanoparticles: a predictive model developed by a nano-quantitative structure-property relationship approach. Chem Mater 2015; 27(7):2400–7. https://doi.org/10.1021/cm504406a.
[13] Toropov AA, Achary PGR, Toropova AP. Quasi-SMILES and Nano-QFPR: the predictive model for zeta Potentials of metal oxide nanoparticles. Chem Phys Lett 2016;660:107–10. https://doi.org/10.1016/j.cplett.2016.08.018.
[14] Mikolajczyk A, Gajewicz A, Mulkiewicz E, Rasulev B, Marchelek M, Diak M, et al. Nano-QSAR modeling for ecosafe design of heterogeneous TiO2-based nano-photocatalysts. Environ Sci Nano 2018;5(5):1150–60. https://doi.org/10.1039/c8en00085a.
[15] Mikolajczyk A, Sizochenko N, Mulkiewicz E, Malankowska A, Rasulev B, Puzyn T. A chemoinformatics approach for the characterization of hybrid nanomaterials: safer and efficient design perspective. Nanoscale 2019;11(24):11808–18. https://doi.org/10.1039/c9nr01162e.
[16] Buglak AA, Zherdev AV, Dzantiev BB. Nano-(Q)SAR for cytotoxicity prediction of engineered nanomaterials. MDPI AG December 11 Molecules 2019. https://doi.org/10.3390/molecules24244537.
[17] Jagiello K, Halappanavar S, Rybińska-Fryca A, Willliams A, Vogel U, Puzyn T. Transcriptomics-based and AOP-informed structure–activity relationships to predict pulmonary pathology induced by multiwalled carbon nanotubes. Small 2021;17(15). https://doi.org/10.1002/smll.202003465.
[18] Jagiello K, Ciura K. In vitro to in vivo extrapolation to support the development of the next generation risk assessment (NGRA) strategy for nanomaterials (Royal Society of Chemistry) Nanoscale 2022. https://doi.org/10.1039/d2nr00664b.
[19] Shirokii N, Din Y, Petrov I, Seregin Y, Sirotenko S, Razlivina J, et al. Quantitative prediction of inorganic nanomaterial cellular toxicity via machine learning. Small 2023. https://doi.org/10.1002/smll.202207106.
[20] Banerjee, A.; Kar, S.; Gajewicz-Skretna, A.; Roy, K. Q-RASAR Modeling of Cytotoxicity of TiO2-Based Multi-Component Nanomaterials. 2022. https://doi.org/10.20944/preprints202210.0402.v1.
[21] Banerjee A, Roy K. First report of Q-RASAR modeling toward an approach of easy interpretability and efficient transferability. Mol Divers 2022;26(5):2847–62. https://doi.org/10.1007/S11030-022-10478-6/METRICS.
[22] Banerjee A, Roy K. Read-across and RASAR tools from the DTC laboratory. Chall Adv Comput Chem Phys 2023;35:239–68. https://doi.org/10.1007/978-3-031-33871-7_9.
[23] Banerjee A, Roy K. Machine-learning-based similarity meets traditional QSAR: "Q-RASAR" for the enhancement of the external predictivity and detection of prediction confidence outliers in an HERG toxicity dataset. Chemom Intell Lab Syst 2023;237. https://doi.org/10.1016/J.CHEMOLAB.2023.104829.
[24] Wyrzykowska E, Mikolajczyk A, Sikorska C, Puzyn T. Development of a novel in silico model of zeta potential for metal oxide nanoparticles: a nano-QSPR approach. Nanotechnology 2016;27(44). https://doi.org/10.1088/0957-4484/27/44/445702.
[25] Sizochenko N, Mikolajczyk A, Syzochenko M, Puzyn T, Leszczynski J. Zeta potentials (ζ) of metal oxide nanoparticles: a meta-analysis of experimental data and a predictive neural networks modeling. NanoImpact 2021;22. https://doi.org/10.1016/j.impact.2021.100317.
[26] Sengottiyan S, Mikolajczyk A, Jagiełło K, Swirog M, Puzyn T. Core, coating, or corona? The importance of considering protein coronas in nano-QSPR modeling of zeta potential. ACS Nano 2023;17(3):1989–97. https://doi.org/10.1021/acsnano.2c06977.
[27] Gajewicz-Skretna A, Kar S, Piotrowska M, Leszczynski J. The kernel-weighted local polynomial regression (KwLPR) approach: an efficient, novel tool for development of QSAR/QSAAR toxicity extrapolation models. J Chemin- 2021;13(1). https://doi.org/10.1186/s13321-021-00484-5.
[28] Soema PC, Willems GJ, Jiskoot W, Amorij JP, Kersten GF. Predicting the influence of liposomal lipid composition on liposome size, zeta potential and liposome-induced dendritic cell maturation using a design of experiments approach. Eur J Pharm Biopharm 2015;94:427–35. https://doi.org/10.1016/j.ejpb.2015.06.026.
[29] Oksel Karakus C, Winkler DA. Overcoming roadblocks in computational roadmaps to the future for safe nanotechnology. Nano Futures 2021;5(2):022002. https://doi.org/10.1088/2399-1984/ABE560.
[30] Haase, A.; Klaessig, F. EU US Roadmap Nanoinformatics 2030. 2018.
[31] Gajewicz A. What if the number of nanotoxicity data is too small for developing predictive nano-QSAR models? An alternative read-across based approach for filling data gaps. Nanoscale 2017;9(24):8435–48. https://doi.org/10.1039/C7NR02211E.
[32] Griffin WC. Classification of surface-active agents by "HLB". J Cosmet Sci 1949;1: 311–26.
[33] Puzyn T, Rasulev B, Gajewicz A, Hu X, Dasari TP, Michalkova A, et al. Using nano-QSAR to predict the cytotoxicity of metal oxide nanoparticles. Nat Nanotechnol 2011;6(3):175–8. https://doi.org/10.1038/nnano.2011.10.
[34] Gajewicz A, Schaeublin N, Rasulev B, Hussain S, Leszczynska D, Puzyn T, et al. Towards understanding mechanisms governing cytotoxicity of metal oxides nanoparticles: hints from nano-QSAR studies. Nanotoxicology 2015;9(3):313–25. https://doi.org/10.3109/17435390.2014.930195.
[35] *ChemAxon.* ⟨https://chemicalize.com/⟩*(02-2023).*
[36] Frisch, M.J., Trucks, G.W., Schlegel, H.B., Scuseria, G.E., Robb, M.A., Cheeseman, J.R., et al., Gaussian16 Revision C.01. 2016.
[37] Foresman, J.; Frisch, a. Exploring Chemistry with Electronic Structure Methods, 1996. *Gaussian Inc, Pittsburgh, PA* 1996.
[38] Schaftenaar G, Noordik JH, Molden A. Pre- and post-processing program for molecular and electronic structures. J Comput Aided Mol Des 2000;14(2):123–34. https://doi.org/10.1023/A:1008193805436.
[39] Gramatica P. Principles of QSAR modeling. Int J Quant Struct-Prop Relatsh 2020;5 (3):61–97. https://doi.org/10.4018/ijqspr.20200701.oa1.
[40] Gajewicz-Skretna A, Furuhama A, Yamamoto H, Suzuki N. Generating accurate in silico predictions of acute aquatic toxicity for a range of organic chemicals: towards similarity-based machine learning methods. Chemosphere 2021;280. https://doi.org/10.1016/j.chemosphere.2021.130681.
[41] Guidance Document on the Validation of (Quantitative) Structure-Activity Relationship [(Q)SAR] Models. 2014. https://doi.org/10.1787/9789264085442-EN.
[42] R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria 2022. ⟨https://www.R-project.org/⟩.
[43] Wei, T.; Simko, V. R Package "Corrplot": Visualization of a Correlation. 2021. ⟨https://github.com/taiyun/corrplot⟩.
[44] Wickham H. Ggplot2: Elegant Graphics for Data Analysis. New York: Springer-Verlag; 2016.
[45] Kassambara, A.; Mundt, F. Factoextra: Extract and Visualize the Results of Multivariate Data Analyses. 2020. ⟨https://CRAN.R-project.org/package=factoextra⟩.
[46] Lê S, Josse J, Husson F. FactoMineR: a package for multivariate analysis. J Stat Softw 2008;25(1):1–18. https://doi.org/10.18637/jss.v025.i01.
[47] Slowikowski, K. Ggrepel: Automatically Position Non-Overlapping Text Labels with "Ggplot2." 2023. ⟨https://CRAN.R-project.org/package=ggrepel⟩.