Exploring Cellular Interactions of Liposomes Using Protein Corona Fingerprints and Physicochemical Properties

Arafeh Bigdeli,[†] Sara Palchetti,[‡] Daniela Pozzi,[‡] Mohammad Reza Hormozi-Nezhad,[†] Francesca Baldelli Bombelli,[§] Giulio Caracciolo,*[,‡] and Morteza Mahmoudi*[,#,∥]

[†]Department of Chemistry, Sharif University of Technology, Tehran 1113658639, Iran

[‡]Department of Molecular Medicine, "Sapienza" University of Rome, Rome 00185, Italy

[§]Department of Chemistry, Materials and Chemical Engineering Giulio Natta, Politecnico di Milano, Milan 20133, Italy

[#]Nanotechnology Research Center, Faculty of Pharmacy, Tehran University of Medical Sciences, Tehran 1316943551, Iran

[∥]Stanford Cardiovascular Institute and Division of Cardiovascular Medicine, Stanford University School of Medicine, Stanford, California 94305, United States

ABSTRACT: To control liposomes fate and transport upon contact with biofluids, it is essential to consider several parameters affecting the synthetic and biological identity of liposomes, as well as liposome−protein corona (PC) aspects. As a powerful tool in this data mining adventure, quantitative structure−activity relationship (QSAR) ap-proach is used to correlate physicochemical properties of liposomes and their PC fingerprints to multiple quantified biological responses. In the present study, the relationship between cellular interactions of a set of structurally diverse liposomal formulations and their physicochemical and PC properties has been investigated *via* linear and nonlinear QSAR models. Significant parameters affecting cellular uptake and cell viability of liposomes in two important cancer cell lines (PC3 and HeLa) have been identified. The developed QSARs have the capacity to be implemented in advanced targeted delivery of liposomal drugs.

KEYWORDS: *liposomes, cellular uptake, cell viability, QSAR, protein corona, cancer*

Liposomes are organic nanoscale vesicles that are clinically established as outstanding drug delivery systems, as well as diagnostic and treatment tools for several diseases, particularly cancer and fungal infections.[1,2] Liposomes consist of single or multiple lipid bilayers encapsulating an aqueous compartment. Their unique self-assembled structure enables them to entrap desired hydrophilic/hydrophobic agents into their internal aqueous core/phospholipid bilayer, making them versatile tools for drug delivery systems. Given their biocompatibility, biodegradability, ease of size, and surface tuning properties, liposomes have attracted much interest in nanomedicine in the last few decades and have found the potential to be further applied in theranostics.[3]

As with any other nanostructure upon entering the biological media, liposomes are also immediately surrounded by high levels of proteins (or other biomolecules), forming a rich protein shell, referred to as "protein corona" (PC). These evolving collection of proteins associated within the lipid surface control the surface properties of liposomes in the biological environment for a certain amount of time, determining what is "seen" by the living organism.[4−6] Thus, the initial synthetic identity of the liposomes (*i.e.*, size, shape, and surface chemistry) will turn into a new biological identity (*i.e.*, size and aggregation state of the liposome, together with the structure and composition of the PC). The resulting biological identity may be far different from the synthetic identity of liposomes, effecting their fate and transport leading to unpredictable biological responses (*e.g.*, biodistribution, uptake, opsonization, toxicity, kinetics).[7]

Several factors such as size, shape, charge, surface chemistry, topography, and curvature of the liposome can regulate the formation of PC. There has been several recent studies investigating the formed PC around liposomes with different

Table 1. Physicochemical Properties of Liposomal Formulations, Including Size of Bare Liposomes and Liposome−HP complexes, Zeta Potential of Bare Liposomes and Liposome−HP complexes, Micrograms of Proteins Bound to Liposomes after 1 h Incubation with HP

	size (nm)		PDI		zeta potential (mV)		protein assay ($\mu g/\mu L$)
	−HP[a]	+HP[b]	−HP	−HP	−HP	+HP	+HP
L1	97 ± 4	235 ± 37	0.12 ± 0.01	0.21 ± 0.04	52 ± 2.1	−20.7 ± 0.6	7.29 ± 1.12
L2	112 ± 11	228 ± 4	0.17 ± 0.01	0.23 ± 0.02	48.5 ± 0.9	−17.9 ± 0.7	9.73 ± 1.10
L3	135 ± 8	259 ± 21	0.14 ± 0.04	0.12 ± 0.04	30.1 ± 3.4	−19.3 ± 1.8	5.06 ± 0.45
L4	128 ± 17	150 ± 6	0.11 ± 0.02	0.15 ± 0.01	47.4 ± 0.9	−23.0 ± 1.1	10.55 ± 0.89
L5	131 ± 7	182 ± 13	0.10 ± 0.01	0.16 ± 0.05	26.1 ± 2.9	−25.5 ± 1.4	5.51 ± 0.54
L6	157 ± 10	211 ± 14	0.17 ± 0.01	0.23 ± 0.01	58.2 ± 4.3	−28.5 ± 2.3	6.70 ± 0.62
L7	108 ± 4	216 ± 2	0.14 ± 0.02	0.22 ± 0.04	34.5 ± 5.5	−17.9 ± 2.3	4.02 ± 0.43
L8	125 ± 10	207 ± 12	0.18 ± 0.02	0.21 ± 0.02	46.2 ± 2.2	−20.4 ± 4.5	5.47 ± 0.56
L9	255 ± 23	329 ± 12	0.12 ± 0.02	0.15 ± 0.01	33.1 ± 2.4	−15.4 ± 1.2	4.36 ± 0.28
L10	193 ± 15	236 ± 18	0.13 ± 0.02	0.19 ± 0.03	42.1 ± 1.2	−19.3 ± 1.4	3.93 ± 0.27
L11	191 ± 23	301 ± 18	0.11 ± 0.02	0.15 ± 0.05	−8.4 ± 1.9	−22.1 ± 2.3	2.94 ± 0.32
L12	136 ± 1	152 ± 12	0.10 ± 0.03	0.15 ± 0.02	−26.6 ± 0.2	−18.2 ± 0.8	8.50 ± 1.09
L13	126 ± 23	171 ± 11	0.11 ± 0.03	0.13 ± 0.02	−14.2 ± 1.1	−28.6 ± 1.4	4.06 ± 0.51
L14	144 ± 4	178 ± 6	0.09 ± 0.01	0.14 ± 0.01	30.2 ± 2.9	−29.3 ± 3.5	4.91 ± 0.64
L15	114 ± 1	137 ± 6	0.08 ± 0.01	0.16 ± 0.03	18 ± 3.3	−13.9 ± 0.6	2.70 ± 0.21
L16	123 ± 6	169 ± 6	0.12 ± 0.03	0.17 ± 0.03	17.9 ± 1.2	−24.5 ± 1.2	1.50 ± 0.03
L17	103 ± 10	115 ± 1	0.09 ± 0.03	0.13 ± 0.04	9.9 ± 2.2	−13.2 ± 0.9	1.10 ± 0.12

[a]Bare liposomes. [b]Liposome−HP complexes.

synthetic identities.[8−11] For instance, liposomes with charged surfaces tend to adsorb more proteins than those with neutral surfaces. Accordingly, liposomes with anionic or cationic phospholipid head groups activate a complement more efficiently than those bearing neutral head groups. In addition to the physicochemical properties of the liposome itself, proteins source and concentration, incubation time and temperature are also included among environmental factors shaping the liposome-PC.[12−15] For instance, it has been shown that liposomes incubated in mouse plasma are more negatively charged than those in human plasma and bear PCs less enriched in opsonins and plentiful in apolipoproteins. This implies a dissimilar circulation time and pharmacokinetic profile of liposomes in the bloodstream of mice, compared to humans.[16]

Quantitative mapping of the relationships between liposomes' synthetic identity, biological identity and biological response enables researchers to thoroughly understand and control their nano-bio interactions in biofluids, which are essential for future developments in this research area. However, developing these correlations will require extensive studies, high-throughput techniques, and new strategies. As a promising approach in this regard, quantitative structure−activity relationship (QSAR) allows to predict the biological impact of PC formation on liposomes bioresponse and facilitates the identification of more meaningful relationships between the synthetic/biological identity of liposomes and their biological outcome. Recently, Chan *et al.*[17] established a quantitative database of PC characterization and used serum PC fingerprints along with nanoparticle properties to predict the cell association of a library of gold nanoparticles with diverse sizes and surface coatings. According to the presented framework, in another study, Cohen *et al.*[18] developed linear and nonlinear QSAR models for predicting nano-bio interactions of the previously introduced library of inorganic nanoparticles. However, Pozzi *et al.*[19] have demonstrated that efficient predictive modeling of nanoparticle behavior *in vivo* requires accurate knowledge of PC fingerprints in circulating biological media, rather than static incubation, which is used as a model for protein adsorption so far. Thus, in order to take into account the effect of PC formation in QSAR studies, the PC composition formed under suitable dynamic conditions is likely the most accurate descriptor to be considered in this regard.

Moving toward liposomes as representative organic nanoparticles, a QSAR approach has been proposed herein to show how PC fingerprints and physicochemical properties of liposomes can predict their cellular interactions. Cellular uptake and cell viability were chosen as the most common biological profiles, revealing the physiological behavior of liposomes in different application of interest. Unlike the predetermined reports, in this study we probed several cellular responses to the corona coated liposomes, leading to deeper understanding of the biological behavior of these nanospecies. Moreover, the effect of protein corona formation has been carefully considered by not only introducing protein corona fingerprints within the descriptor set, but also measuring the cellular responses with and without precoating the liposomes with human plasma proteins.

RESULTS AND DISCUSSION

A set of 17 liposomal formulations was synthesized, followed by characterization steps before and after incubation in Human Plasma (HP). The prepared formulations differed in their lipid composition and were characterized by different size and surface charge (see Methods for details). The formulations included anionic, cationic, and neutral-charged liposomes with an averaged hydrodynamic size ranging from about 100 to 250 nm, creating a library of physicochemical properties. Hydro-dynamic sizes and zeta potential measurements of bare liposomes and liposome-HP complexes, together with protein assay results (micrograms of proteins bound to liposomes after 1 h incubation with HP), are provided in Table 1. Interaction with pure HP was generally found to promote a significant increase in the hydrodynamic size (ranging between 12 and 138 nm) and the polydispersity index (PDI). Following interaction

Table 2. Descriptors Used in QSAR Models Including LPP and PCF

Liposome Physicochemical Properties (LPP)

−HP	+HP
Size (nm)	Size (nm)
Molecular Weight (g/mol)	Zeta potential (mV)
Surface area per liposome (cm^2)	Protein Density ($\mu g/\mu L$)
Mol (per liposome)	
Zeta Potential (mV)	

Protein Corona Fingerprints (PCF)

abbrev.	full name	abbrev.	full name
VTN	Vitronectin	HPX	Hemopexin
FGB	Fibrinogen beta chain	YWHAZ	14-3-3 protein zeta/delta
F2	Prothrombin	SERPINC1	Antithrombin-III
ALB	Serum albumin	PLTP	Phospholipid transfer protein
APOA2	Apolipoprotein A-II	CFH	Complement factor H
FGG	Fibrinogen gamma chain	ILK	Integrin-linked protein kinase
IGKC	Ig kappa chain C region	ANGPTL6	Angiopoietin-related protein 6
C3	Complement C3	CP	Ceruloplasmin
FGA	Fibrinogen alpha chain	COL18A1	Collagen alpha-1(XVIII) chain
APOC2	Apolipoprotein C-II	MMRN1	Multimerin-1
Clusterin	Clusterin	C1R	Complement C1r subcomponent
HABP2	Hyaluronan-binding protein 2	IGLC2	Ig lambda-2 chain C regions
APOC3	Apolipoprotein C−III	IGHG1	Ig gamma-1 chain C region
PROS1	Vitamin K-dependent protein S	KRT1	Keratin, type II cytoskeletal 1
C4BPA	C4b-binding protein alpha chain	APOC1	Apolipoprotein C-I
TTR	Transthyretin	IGHG2	Ig gamma-2 chain C region
HBB	Hemoglobin subunit beta	APOB	Apolipoprotein B-100
AMBP	Protein AMBP	TF	Serotransferrin
HBA1	Hemoglobin subunit alpha	KRT2	Keratin, type II cytoskeletal 2 epidermal
ITIH2	Interalpha-trypsin inhibitor heavy chain H2	ACTB	Actin, cytoplasmic 1
ITIH1	Interalpha-trypsin inhibitor heavy chain H1	KRT10	Keratin, type I cytoskeletal 10
ITIH3	Interalpha-trypsin inhibitor heavy chain H3	C4B	Complement C4-B
PROC	Vitamin K-dependent protein C	IGHK8	Ig kappa chain V−III region HAH
PROZ	Vitamin K-dependent protein Z	A2M	Alpha-2-macroglobulin
LUM	Lumican	HPR	Haptoglobin-related protein
APOA1	Apolipoprotein A-I	IGHM	Ig mu chain C region
KNG1	Kininogen-1	ORM1	Alpha-1-acid glycoprotein 1
SERPIND1	Heparin cofactor 2	HP	Haptoglobin
APOE	Apolipoprotein E	KRT5	Keratin, type II cytoskeletal 5
C1S	Complement C 1s subcomponent	KRT14	Keratin, type I cytoskeletal 14
SERPINA1	Alpha-1-antitrypsin	PON1	Serum paraoxonase/arylesterase 1
SERPINA10	Protein Z-dependent protease inhibitor	IGHA1	Ig alpha-1 chain C region
C4BPB	C4b-binding protein beta chain	RBP4	Retinol-binding protein 4
GSN	Gelsolin	C1QB	Complement C1q subcomponent subunit B
MYL6	Myosin light polypeptide 6	APOA5	Apolipoprotein A-V
RAP1B	Ras-related protein Rap-1b	SHBG	Sex hormone-binding globulin
LBP	Lipopolysaccharide-binding protein	ORM2	Alpha-1-acid glycoprotein 2
C9	Complement component C9	KRT16	Keratin, type I cytoskeletal 16
SPP2	Secreted phosphoprotein 24	APOC4	Apolipoprotein C-IV
TLN1	Talin-1	APOF	Apolipoprotein F
CFB	Complement factor B	F9	Coagulation factor IX
ITGA2B	Integrin alpha-IIb	IGHC8	Ig heavy chain V−III region VH26
STOM	Erythrocyte band 7 integral membrane protein	C1QC	Complement C1q subcomponent subunit C
MYH9	Myosin-9	A1BG	Alpha-1B-glycoprotein
APOM	Apolipoprotein M	F10	Coagulation factor X
MASP1	Mannan-binding lectin serine protease 1	KRT6A	Keratin, type II cytoskeletal 6A
IgKC5	Ig kappa chain V−IV region Len	DCD	Dermcidin
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase	AGLL5	Immunoglobulin lambda-like polypeptide 5
HRG	Histidine-rich glycoprotein	TUBB1	Tubulin beta-1 chain
APOA4	Apolipoprotein A-IV	SAA4	Serum amyloid A-4 protein
CRP	C-reactive protein	TUBA4A	Tubulin alpha-4A chain
PFN1	Profilin-1	FN1	Fibronectin

Table 2. continued

Protein Corona Fingerprints (PCF)			
abbrev.	full name	abbrev.	full name
APOH	Beta-2-glycoprotein 1	COMP	Cartilage oligomeric matrix protein
THBS1	Thrombospondin-1	SERPINA4	Kallistatin
ITGB3	Integrin beta-3	GP1BA	Platelet glycoprotein Ib alpha chain
KRT9	Keratin, type I cytoskeletal 9	IGHC8	Ig heavy chain V−III
AHSG	Alpha-2-HS-glycoprotein	TUBA1B	Tubulin alpha-1B chain
LPA	Apolipoprotein(a)	FBLN1	Fibulin-1
FLNA	Filamin-A	THBS4	Thrombospondin-4
IGJ	Immunoglobulin J chain	AFM	Afamin
C5	Complement C5	IGHL8	Ig lambda chain V−III region LO
AGT	Angiotensinogen	F5	Coagulation factor V
GC	Vitamin D-binding protein	VWF	von Willebrand factor
SEPP1	Selenoprotein P	SLC4A1	Band 3 anion transport protein
COLEC10	Collectin-10	SERPINF1	Pigment epithelium-derived factor
APOD	Apolipoprotein D	CALR	Calreticulin
SERPINF2	Alpha-2-antiplasmin	CFHR5	Complement factor H-related protein 5
TPM4	Tropomyosin alpha-4 chain	LCAT	Phosphatidylcholine-sterol acyltransferase
C8G	Complement component C8 gamma chain	CF1	Complement factor I
APOL1	Apolipoprotein L1	CPN2	Carboxypeptidase N subunit 2
VCL	Vinculin	HSP90B1	Endoplasmin
PGLYRP2	N-acetylmuramoyl-L-alanine amidase	CNDP1	Beta-Ala-His dipeptidase
C7	Complement component C7	HRNR	Hornerin
C8B	Complement component C8 beta chain	PRG4	Proteoglycan 4
C6	Complement component C6	FLG2	Filaggrin-2
SERPING1	Plasma protease C1 inhibitor	TUBB	Tubulin beta chain
C8A	Complement component C8 alpha chain	COLEC11	Collectin-11
C1QA	Complement C1q subcomponent subunit A	IGHG3	Ig gamma-3 chain C region
GP9	Platelet glycoprotein IX	IGHG8	Ig heavy chain V−III region TRO
FERMT3	Fermitin family homologue 3	CD14	Monocyte differentiation antigen CD14
PLG	Plasminogen	LGALS3BP	Galectin-3-binding protein
CETP	Cholesteryl ester transfer protein	LYZ	Lysozyme C
SERPINA3	Alpha-1-antichymotrypsin	IGBUT	Ig heavy chain V−III region BUT
ITIH4	Interalpha-trypsin inhibitor heavy chain H4	SAA1	Serum amyloid A-1 protein
VTN	Vitronectin	HPX	Hemopexin
FGB	Fibrinogen beta chain	YWHAZ	14-3-3 protein zeta/delta
F2	Prothrombin	SERPINC1	Antithrombin-III
ALB	Serum albumin	PLTP	Phospholipid transfer protein
APOA2	Apolipoprotein A-II	CFH	Complement factor H
FGG	Fibrinogen gamma chain	ILK	Integrin-linked protein kinase
IGKC	Ig kappa chain C region	ANGPTL6	Angiopoietin-related protein 6
C3	Complement C3	CP	Ceruloplasmin
FGA	Fibrinogen alpha chain	COL18A1	Collagen alpha-1(XVIII) chain
APOC2	Apolipoprotein C−II	MMRN1	Multimerin-1
Clusterin	Clusterin	C1R	Complement C1r subcomponent
HABP2	Hyaluronan-binding protein 2	IGLC2	Ig lambda-2 chain C regions
APOC3	Apolipoprotein C−III	IGHG1	Ig gamma-1 chain C region
PROS1	Vitamin K-dependent protein S	KRT1	Keratin, type II cytoskeletal 1
C4BPA	C4b-binding protein alpha chain	APOC1	Apolipoprotein C−I
TTR	Transthyretin	IGHG2	Ig gamma-2 chain C region
HBB	Hemoglobin subunit beta	APOB	Apolipoprotein B-100
AMBP	Protein AMBP	TF	Serotransferrin
HBA1	Hemoglobin subunit alpha	KRT2	Keratin, type II cytoskeletal 2 epidermal
ITIH2	Interalpha-trypsin inhibitor heavy chain H2	ACTB	Actin, cytoplasmic 1
ITIH1	Interalpha-trypsin inhibitor heavy chain H1	KRT10	Keratin, type I cytoskeletal 10
ITIH3	Interalpha-trypsin inhibitor heavy chain H3	C4B	Complement C4−B
PROC	Vitamin K-dependent protein C	IGHK8	Ig kappa chain V−III region HAH
PROZ	Vitamin K-dependent protein Z	A2M	Alpha-2-macroglobulin
LUM	Lumican	HPR	Haptoglobin-related protein
APOA1	Apolipoprotein A-I	IGHM	Ig mu chain C region
KNG1	Kininogen-1	ORM1	Alpha-1-acid glycoprotein 1

Table 2. continued

Protein Corona Fingerprints (PCF)			
abbrev.	full name	abbrev.	full name
SERPIND1	Heparin cofactor 2	HP	Haptoglobin
APOE	Apolipoprotein E	KRT5	Keratin, type II cytoskeletal 5
C1S	Complement C 1s subcomponent	KRT14	Keratin, type I cytoskeletal 14
SERPINA1	Alpha-1-antitrypsin	PON1	Serum paraoxonase/arylesterase 1
SERPINA10	Protein Z-dependent protease inhibitor	IGHA1	Ig alpha-1 chain C region
C4BPB	C4b-binding protein beta chain	RBP4	Retinol-binding protein 4
GSN	Gelsolin	C1QB	Complement C1q subcomponent subunit B
MYL6	Myosin light polypeptide 6	APOA5	Apolipoprotein A-V
RAP1B	Ras-related protein Rap-1b	SHBG	Sex hormone-binding globulin
LBP	Lipopolysaccharide-binding protein	ORM2	Alpha-1-acid glycoprotein 2
C9	Complement component C9	KRT16	Keratin, type I cytoskeletal 16
SPP2	Secreted phosphoprotein 24	APOC4	Apolipoprotein C−IV
TLN1	Talin-1	APOF	Apolipoprotein F
CFB	Complement factor B	F9	Coagulation factor IX
ITGA2B	Integrin alpha-IIb	IGHC8	Ig heavy chain V−III region VH26
STOM	Erythrocyte band 7 integral membrane protein	C1QC	Complement C1q subcomponent subunit C
MYH9	Myosin-9	A1BG	Alpha-1B-glycoprotein
APOM	Apolipoprotein M	F10	Coagulation factor X
MASP1	Mannan-binding lectin serine protease 1	KRT6A	Keratin, type II cytoskeletal 6A
IgKC5	Ig kappa chain V−IV region Len	DCD	Dermcidin
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase	AGLL5	Immunoglobulin lambda-like polypeptide 5
HRG	Histidine-rich glycoprotein	TUBB1	Tubulin beta-1 chain
APOA4	Apolipoprotein A-IV	SAA4	Serum amyloid A-4 protein
CRP	C-reactive protein	TUBA4A	Tubulin alpha-4A chain
PFN1	Profilin-1	FN1	Fibronectin
APOH	Beta-2-glycoprotein 1	COMP	Cartilage oligomeric matrix protein
THBS1	Thrombospondin-1	SERPINA4	Kallistatin
ITGB3	Integrin beta-3	GP1BA	Platelet glycoprotein Ib alpha chain
KRT9	Keratin, type I cytoskeletal 9	IGHC8	Ig heavy chain V−III
AHSG	Alpha-2-HS-glycoprotein	TUBA1B	Tubulin alpha-1B chain
LPA	Apolipoprotein(a)	FBLN1	Fibulin-1
FLNA	Filamin-A	THBS4	Thrombospondin-4
IGJ	Immunoglobulin J chain	AFM	Afamin
C5	Complement C5	IGHL8	Ig lambda chain V−III region LO
AGT	Angiotensinogen	F5	Coagulation factor V
GC	Vitamin D-binding protein	VWF	von Willebrand factor
SEPP1	Selenoprotein P	SLC4A1	Band 3 anion transport protein
COLEC10	Collectin-10	SERPINF1	Pigment epithelium-derived factor
APOD	Apolipoprotein D	CALR	Calreticulin
SERPINF2	Alpha-2-antiplasmin	CFHR5	Complement factor H-related protein 5
TPM4	Tropomyosin alpha-4 chain	LCAT	Phosphatidylcholine-sterol acyltransferase
C8G	Complement component C8 gamma chain	CF1	Complement factor I
APOL1	Apolipoprotein L1	CPN2	Carboxypeptidase N subunit 2
VCL	Vinculin	HSP90B1	Endoplasmin
PGLYRP2	N-acetylmuramoyl-L-alanine amidase	CNDP1	Beta-Ala-His dipeptidase
C7	Complement component C7	HRNR	Hornerin
C8B	Complement component C8 beta chain	PRG4	Proteoglycan 4
C6	Complement component C6	FLG2	Filaggrin-2
SERPING1	Plasma protease C1 inhibitor	TUBB	Tubulin beta chain
C8A	Complement component C8 alpha chain	COLEC11	Collectin-11
C1QA	Complement C1q subcomponent subunit A	IGHG3	Ig gamma-3 chain C region
GP9	Platelet glycoprotein IX	IGHG8	Ig heavy chain V−III region TRO
FERMT3	Fermitin family homologue 3	CD14	Monocyte differentiation antigen CD14
PLG	Plasminogen	LGALS3BP	Galectin-3-binding protein
CETP	Cholesteryl ester transfer protein	LYZ	Lysozyme C
SERPINA3	Alpha-1-antichymotrypsin	IGBUT	Ig heavy chain V−III region BUT
ITIH4	Interalpha-trypsin inhibitor heavy chain H4	SAA1	Serum amyloid A-1 protein
VTN	Vitronectin	HPX	Hemopexin
FGB	Fibrinogen beta chain	YWHAZ	14−3−3 protein zeta/delta

Table 2. continued

Protein Corona Fingerprints (PCF)			
abbrev.	full name	abbrev.	full name
F2	Prothrombin	SERPINC1	Antithrombin-III
ALB	Serum albumin	PLTP	Phospholipid transfer protein
APOA2	Apolipoprotein A-II	CFH	Complement factor H
FGG	Fibrinogen gamma chain	ILK	Integrin-linked protein kinase
IGKC	Ig kappa chain C region	ANGPTL6	Angiopoietin-related protein 6
C3	Complement C3	CP	Ceruloplasmin
FGA	Fibrinogen alpha chain	COL18A1	Collagen alpha-1(XVIII) chain
APOC2	Apolipoprotein C−II	MMRN1	Multimerin-1
Clusterin	Clusterin	C1R	Complement C1r subcomponent
HABP2	Hyaluronan-binding protein 2	IGLC2	Ig lambda-2 chain C regions
APOC3	Apolipoprotein C−III	IGHG1	Ig gamma-1 chain C region
PROS1	Vitamin K-dependent protein S	KRT1	Keratin, type II cytoskeletal 1
C4BPA	C4b-binding protein alpha chain	APOC1	Apolipoprotein C−I
TTR	Transthyretin	IGHG2	Ig gamma-2 chain C region

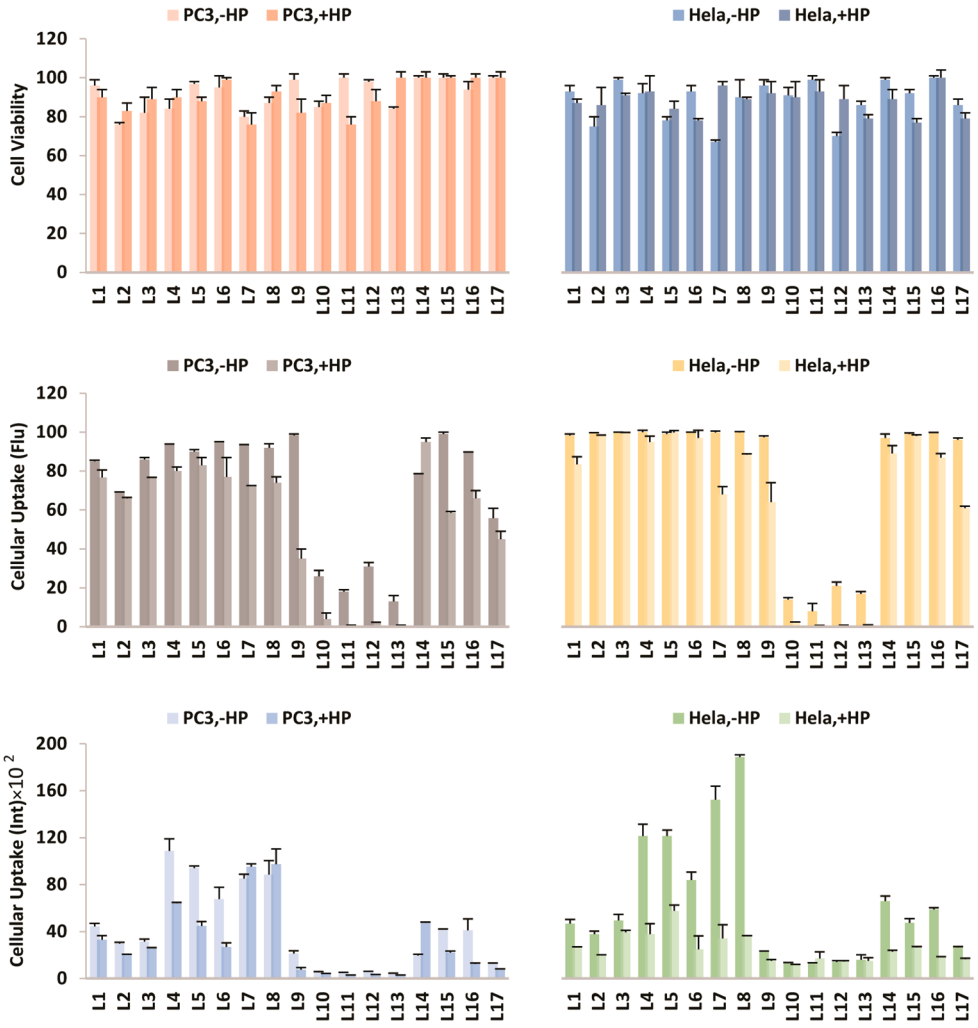

Figure 1. Cell viability, percentage of fluorescent cells, and mean intensity in PC3 and HeLa cell lines of bare liposomes (−HP) and liposome−HP complexes (+HP).

with HP, zeta potential of cationic liposomes reversed to negative values, while it remained negative for anionic ones (L11−L13). In particular, incubation of the liposomes in HP led to a "normalization" of the zeta potential to an average value of about −21.04 ± 4.98 mV independently of the pristine surface charge. This behavior was already observed for metal NPs incubated in HP indicating that likely the main driving force in the protein−NP interaction is not electrostatic.[20−23] After incubation with HP for 1 h at 37 °C, HP−liposome complexes were isolated from excess of plasma and an average of 5.20 ± 0.64 μg/mL adsorbed proteins was quantified per each liposomal formulation. In particular, it can be seen that

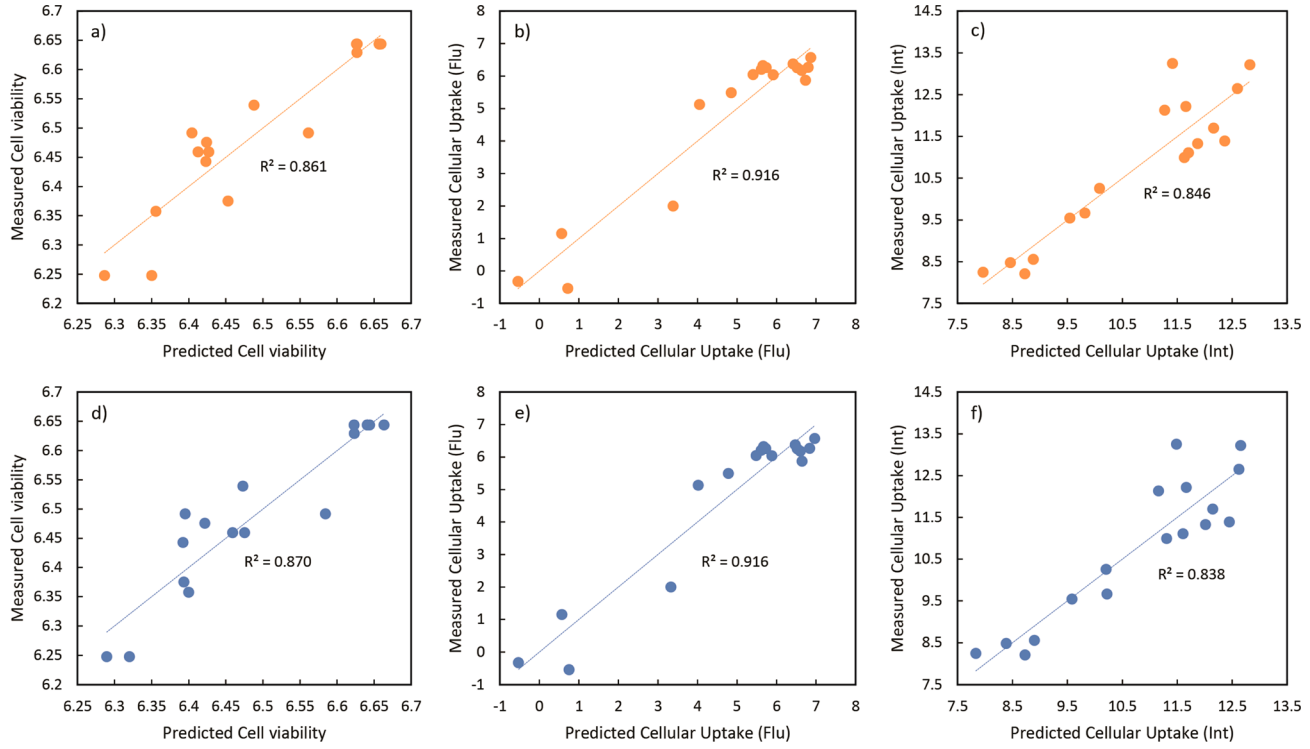

Figure 2. Measured bioresponse values (log2-transformed) of PC3 cells after administration of liposome−HP complexes *versus* their predicted values by linear MLR (top) and nonlinear ANN (bottom) QSARs. (a and d) Cell viability; (b and e) cellular uptake_Fluorescence; (c and f) cellular uptake_Intensity. The data for HeLa cell line and −HP/+HP is depicted in Figures S1−S3.

higher amount of adsorbed protein was associated with highly positive lipid surface, while a lower protein enrichment was found for slightly positive lipid surfaces that were partially coated by polyethyleneglygol (PEG) chains. The characterization information on the liposomes, before and after incubation in HP, referred to as "liposome physicochemical properties (LPP)", was used as the first descriptor set, including 23 LPPs (briefly listed in Table 2).

The PC composition of each liposomal formulation was quantitatively characterized by liquid chromatography tandem mass spectrometry (LC−MS/MS). The results of the obtained PC compositions are presented in Table S1 (*i.e.*, descriptors 1−633) of the Supporting Information. For formulation, a quantitative profile of 633 HP proteins was provided. The relative abundance of the adsorbed proteins was named "PC fingerprints (PCF)" and it represents the second descriptor set (Table 2). The PCF obtained for the liposome formulations confirmed that the physicochemical properties of liposomes can dictate the composition of the PC.[6,8] It also showed that the PC composition is not necessarily enriched with the most abundant proteins in plasma.[8,24]

Twelve different biological end points were measured for each liposomal formulation, including their toxicity and cellular uptake, either with or without incubation with HP and in two different cancer cell lines. The corresponding data do not show any global trend, emphasizing the necessity of utilizing a comprehensive modeling approach for a thorough investiga-tion. For instance, it can be seen in Figure 1 that the overall percentage of fluorescent cells is significantly lower for L10−L13 in both cell lines and this reduction is enhanced after incubation in HP. We can notice that those formulations are all formed with negative charged liposomes, but the charge does not seem to be the only parameter influencing the uptake as also L10 (positive charged) shows a modest uptake. Thus, the phenomenon has to be due to the interplay between PC, size and charge variations. In fact, this hypothesis could not justify the low uptake value for L10, bearing a positive surface charge, implying that probably more variables other than solely charge might be responsible in this particular bioresponse. In addition, PC formation (together with its effects on surface charge of the particles) has shown to significantly affect the biological response of certain liposome formulation (*e.g.*, cellular uptake intensity values for HeLa cells has dramatically decreased after HP incubation for L7−L8), while in some other cases, none or very little changes upon corona formation were observed (*e.g.*, cell viability in PC3 cells in for L14−L17). These examples confirmed that for an in-depth and systematic investigation on the impact of PC on liposomes biological behavior, it is necessary to make some step further and implement it with QSAR analysis, rather than only considering the effect of one variable at a time. Moreover, it is necessary to take into account both descriptor sets: LPPs (revealing synthetic and biological identities of liposomes) and PCFs (covering the PC information) for a comprehensive analysis.

To develop linear and nonlinear correlations between the determined descriptors (x) and the measured biological end points (y), multiple linear regressions (MLR) and artificial neural networks (ANN) were accomplished to model y as a function of x, $y = f(x)$. Figure 2 compares the performance of the constructed linear and nonlinear regression models for predicting the three different biological response profiles of the liposome data set (L1−L17). The results for the remaining nine response profiles can be found in the Supporting Information in Figure S1−S3. The R^2 values displayed in Figure 2 were obtained including all liposome formulations in the model that was created using the best-selected descriptors chosen during

Table 3. Summary of Linear and Nonlinear QSAR Models for All Bioresponses

		cell viability				cellular uptake (flu)				cellular uptake (int)			
		PC3 −HP	PC3 +HP	HeLa −HP	HeLa +HP	PC3 −HP	PC3 +HP	HeLa −HP	HeLa +HP	PC3 −HP	PC3 +HP	HeLa −HP	HeLa +HP
Linear MLR model	R^2_{Fit}	0.79	0.88	0.82	0.71	0.93	0.94	0.91	0.93	0.88	0.89	0.84	0.86
	R^2_{cv}	0.59	0.65	0.55	0.55	0.64	0.66	0.69	0.65	0.60	0.60	0.57	0.56
	$RMSE_{Fit}$	0.06	0.04	0.07	0.05	0.23	0.57	0.34	0.71	0.52	0.53	0.48	0.23
	$RMSE_{cv}$	0.20	0.16	0.26	0.18	1.07	3.42	1.95	3.91	2.02	1.89	1.55	0.79
	$R^2_{y\text{-rand}}$	0.12	0.14	0.19	0.14	0.20	0.14	0.23	0.21	0.17	0.13	0.16	0.14
Nonlinear ANN model	R^2_{Fit}	0.64	0.84	0.77	0.72	0.97	0.99	0.98	0.99	0.83	0.77	0.76	0.82
	R^2_{cv}	0.53	0.58	0.49	0.65	0.69	0.74	0.67	0.73	0.57	0.54	0.52	0.58
	$RMSE_{Fit}$	0.06	0.04	0.06	0.08	0.09	0.19	0.10	0.15	0.48	0.62	0.52	0.23
	$RMSE_{cv}$	0.17	0.13	0.23	0.13	0.76	1.94	1.09	2.00	1.58	1.63	1.35	0.64
	$R^2_{y\text{-rand}}$	0.09	0.16	0.14	0.13	0.11	0.12	0.15	0.16	0.11	0.15	0.12	0.10

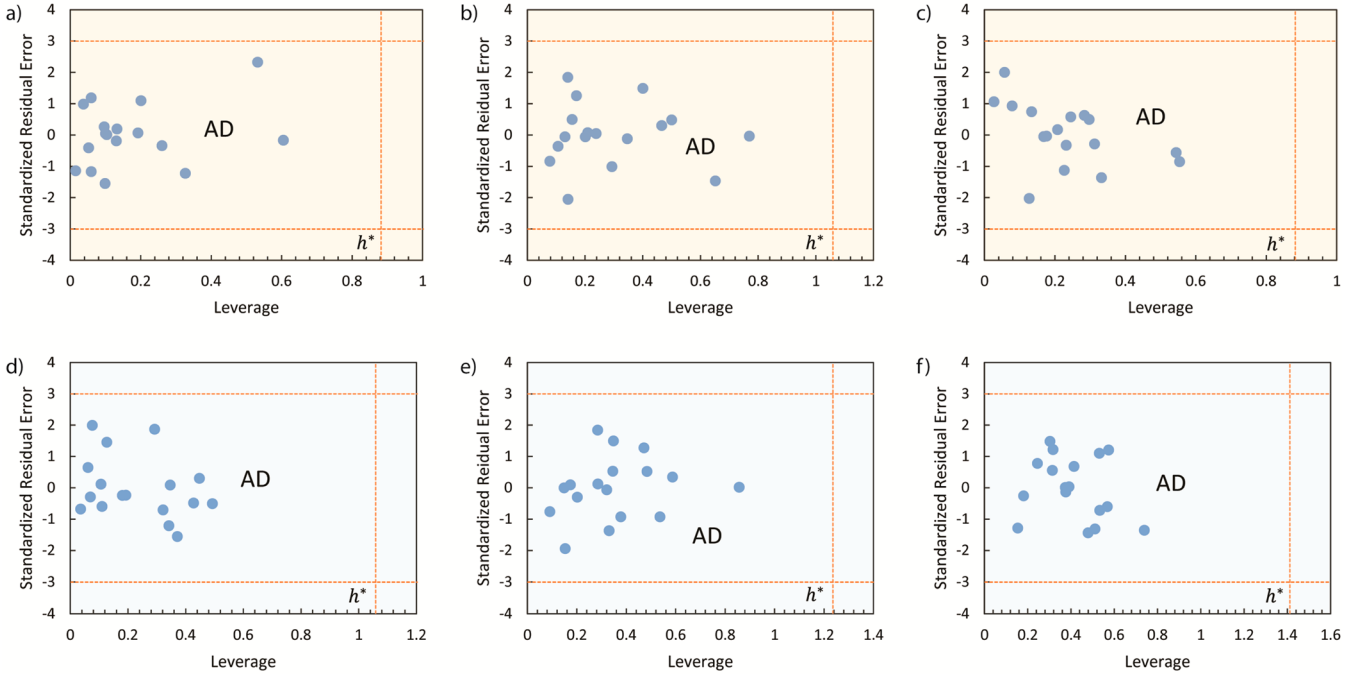

Figure 3. Applicability domain of the linear and nonlinear QSAR models identified by William's plot. (a) MLR, cell viability; (b) MLR, cellular uptake_Flu; (c) MLR, cellular uptake_Int; (d) ANN, cell viability; (e) ANN, cellular uptake_Flu; (f) ANN, cellular uptake_Int. The data shown here is for PC3,+HP. The AD for HeLa cell line and −HP/+HP can be found in Figures S4−S6.

the variable selection step. The accuracy of the developed QSARs was further assessed *via* Monte Carlo cross validation, in which the data set was repeatedly partitioned into random training and prediction sets. The former was used to build the model and also to choose the most suitable variables, while the latter was used to evaluate the prediction ability of the created model. The training and prediction accuracies were quantified by the average coefficient of determination for the training and prediction sets, referred to as fitting (R^2_{fit}) and cross-validation (R^2_{cv}), respectively (Table 3). Both linear and nonlinear models showed high accuracy, deduced by their acceptable R^2_{fit} and R^2_{cv} values. The results revealed that ANN models provided a slightly better correlation for the fluorescent cellular uptake of liposomes compared to the MLR models. This disclosed the nonlinear nature of the corresponding biointeraction. In contrast, for the other two bioresponses, the MLR models seemed to perform better in correlating the liposomes properties to their physiological end points. Determining the linear or nonlinear behavior beyond a specific biological interaction can lead to superior design of the liposomes. Another index, the root-mean-square error (RMSE), was also measured for the training and prediction sets for evaluating the predictability of the developed models. Accordingly, the low $RMSE_{fit}$ and $RMSE_{cv}$ values also confirmed the satisfactory accuracy of the developed models. The linear and nonlinear QSARs demonstrated to have a wide range of applicability as they predict the right behavior of all liposome formulations except for L14 in its linear correlation to cellular uptake in HeLa cells (Figures 3 and S4−S6).

The selected variables in each model are summarized in Table 4. The first point that can be concluded is that the best variables selected in all QSAR models included both LPP and PCF descriptors, implying the importance of the information arising from both physicochemical properties and liposome− PC. This was further tested by building the QSAR models using only LPP descriptors (Table 5). Comparing the results between this table and Table 3 confirms the role of information-rich PCF descriptors that contribute to better QSAR models. The

Table 4. List of Selected Variables in Linear and Nonlinear QSAR Models for Each Bioresponse

	descriptor rank		1st	2nd	3rd	4th	5th	6th	7th
Cell Viability	PC3, −HP	MLR	DCD	SA	CP	RBP4			
		ANN	CP	C9	Size, change	SA	Tot Pr		
	PC3, +HP	MLR	SAA4	LPA	COL18A1	Size, change			
		ANN	SAA4	LPA	COL18A1	CP			
	HeLa, −HP	MLR	IGHL8	Size, change	CP	ZP, change			
		ANN	IGHL8	Size, change	CP	ZP, change	CETP		
	HeLa, +HP	MLR	KRT5	IGLC2	ZP, +HP				
		ANN	SAA4	IGLC2	IGHL8	ZP, +HP	KRT5		
Cellular Uptake (Flu)	PC3, −HP	MLR	IgKC5	Size, −HP	APOF	GSN	PRG4		
		ANN	KRT6A	ZP, +HP	APOF	HPX	GSN	C8B	
	PC3, +HP	MLR	Size,-HP	AHSG	ZP, −HP	APOF			
		ANN	PRG4	ZP, −HP	Size, change	ZP, +HP			
	HeLa, −HP	MLR	HPX	LPA	APOF				
		ANN	PRG4	Size, −HP	AHSG	HPX	SHBG	APOF	
	HeLa, +HP	MLR	PRG4	APOF	AHSG	ZP, −HP			
		ANN	ZP, −HP	AHSG	SHBG	APOF	PRG4		
Cellular Uptake (Int)	PC3, −HP	MLR	ZP, change	APOF	Size, −HP	ZP, −HP			
		ANN	AGT	C1R	ZP, −HP	GSN	F9	APOF	ZP, +HP
	PC3, +HP	MLR	APOF	PROC	Size, −HP	ZP, −HP			
		ANN	Size, −HP	ZP, change	PROC	ZP, −HP	APOF		
	HeLa, −HP	MLR	ZP, change	LYZ	ZP, −HP				
		ANN	COMP	APOF	Size, −HP	PROC			
	HeLa, +HP	MLR	TUBB1	HPR	SA	ZP, −HP			
		ANN	HPR	TUBB1	SERPING1	SA	ZP, +HP	ZP, −HP	

Table 5. Summary of Linear and Nonlinear QSAR Models When Considering Only LPPs as Descriptors

		cell viability				cellular uptake (flu)				cellular uptake (int)			
		PC3 −HP	PC3 +HP	HeLa −HP	HeLa +HP	PC3 −HP	PC3 +HP	HeLa −HP	HeLa +HP	PC3 −HP	PC3 +HP	HeLa −HP	,HeLa +HP
Linear MLR model	R^2_{Fit}	0.67	0.73	0.67	0.38	0.76	0.88	0.80	0.85	0.75	0.81	0.78	0.71
	R^2_{cv}	0.63	0.66	0.49	0.60	0.66	0.71	0.68	0.70	0.62	0.64	0.63	0.61
	$RMSE_{Fit}$	0.07	0.07	0.10	0.08	0.42	0.78	0.51	1.05	0.75	0.71	0.57	0.33
	$RMSE_{cv}$	0.20	0.15	0.28	0.19	1.07	2.09	1.20	2.79	1.75	1.82	1.22	0.67
Nonlinear ANN model	R^2_{Fit}	0.36	0.65	0.36	0.01	0.63	0.89	0.88	0.87	0.61	0.70	0.64	0.61
	R^2_{cv}	0.41	0.57	0.39	0.01	0.67	0.66	0.64	0.69	0.50	0.59	0.50	0.54
	$RMSE_{Fit}$	0.09	0.07	0.11	0.09	0.39	0.45	0.26	0.65	0.85	0.83	0.65	0.34
	$RMSE_{cv}$	0.16	0.12	0.21	0.11	0.91	2.07	1.21	2.99	1.51	1.50	1.08	0.55

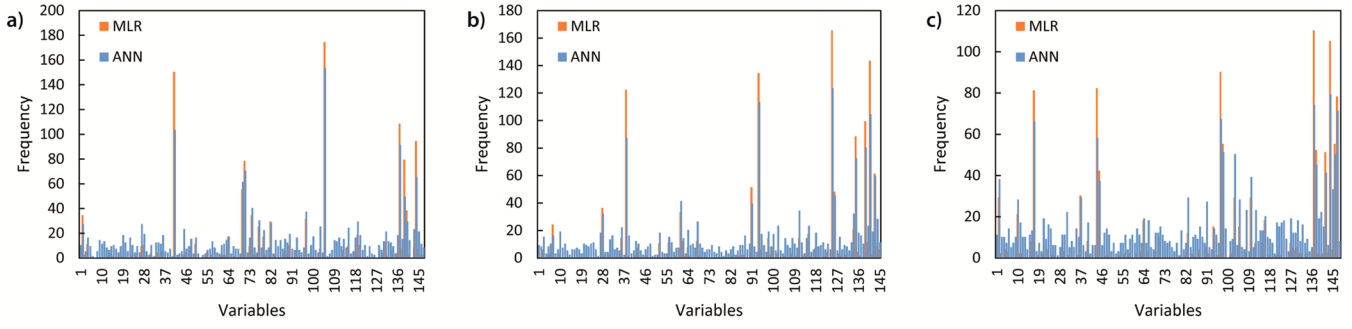

Figure 4. Frequency plot of variables in linear (orange) and nonlinear (blue) QSAR models. The higher frequency of a variable indicates it has been more selected during the variable selection iterations. (a) Cell viability, (b) cellular uptake_Flu, (c) cellular uptake_Int. The frequency plots in this figure are related to PC3 cell line and +HP. The results for HeLa cells and −HP/+HP can be found in Figures S7−S9.

descriptors listed for each model provide a comprehensive comparison framework for the important parameters in each biological response. For example, it seems that the IGHL8, IGLC2, CP, SAA4. and COL18A1 are the selected descriptors responsible for interpreting the cell viability values rather than the cellular uptake, which is more related to APOF. Moreover, regarding to cell viability, IGHL8, IGLC2. and CP governed the HeLa cell interaction, while SAA4 and COL18A1 seemed to correlate better to PC3 cell interactions. From another point of view, considering and comparing the number of times that ZP

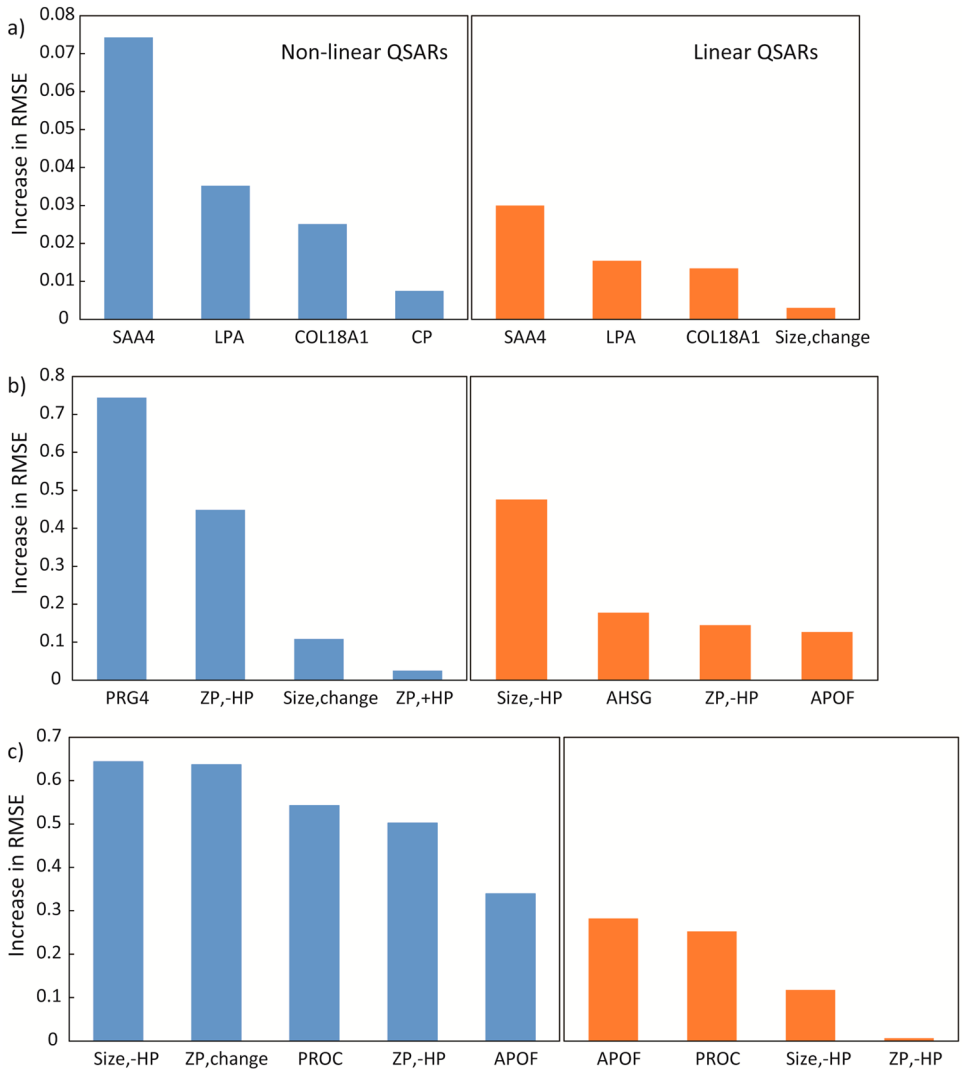

Figure 5. Variables' importance plot measured by sensitivity analysis. Increase in prediction error (RMSE) upon randomly permutation of each descriptor identifies the greater effect of that descriptor in developing the QSAR model. (a) Cell viability, (b) cellular uptake_Flu, (c) cellular uptake_Int. The VIPs in this figure are related to PC3 cell line and +HP. The results for HeLa cells and −HP/+HP can be found in Figures S10−S12.

emerged as the important descriptor in −HP in contrast to +HP subgroups, we can guess that surface charge is an important parameter associated with the synthetic identity of liposomes rather than their biological identity. These phenomena can be further explained by the "charge level-off" issue occurring for charged particles surrounded by serum/plasma proteins. Moreover, it seemed that zeta potential did not affect the interaction between liposomes and PC3 cells, regarding to the cell viability. In contrast, the appearance of ZP-related descriptors among the important variables for HeLa cells expressed the impact of charge in this cell line. Several other conclusions could be made by a deep screening of the data reported in Table 4. Generally, a consistency was found among the selected variables in linear and nonlinear models, which increased the confidence regarding the significance of the selected descriptors. This consistency is simply justified from the variables frequency plot (Figures 4 and S7−S9).

The most significant variables in each model were ranked by sensitivity analysis (Figures 5 and S10−S12), in which the most important variable caused a higher increase in the prediction error of the resulting model (RMSE) upon its permutation, indicating its significant role for building the model. To make sure that the developed models were significant (*i.e.*, not chance correlations), Y-scrambling test was also performed. The very low R^2 values after scrambling the **y** vector in each case indicates the robustness of the corresponding QSAR models (Table 3). Finally, a histogram of the correlation accuracies through the sampling iterations (see Methods) is illustrated in Figure 6 which shows acceptable distribution over the mean R^2 values reported for each model.

CONCLUSION

Protein corona fingerprints and physicochemical properties of 17 liposomal formulations were provided to predict multiple biological interactions. The cellular uptake and cell viability of PC3 and HeLa cells were linearly and nonlinearly correlated to the synthetic identity, biological identity, and protein corona of liposomes. Accurate QSAR models were established and the selected descriptors were further used for a thorough investigation on the related nano-bio interactions. Though the significant role of PC formation on nanoparticles fate and behavior has been revealed before in previous studies, this

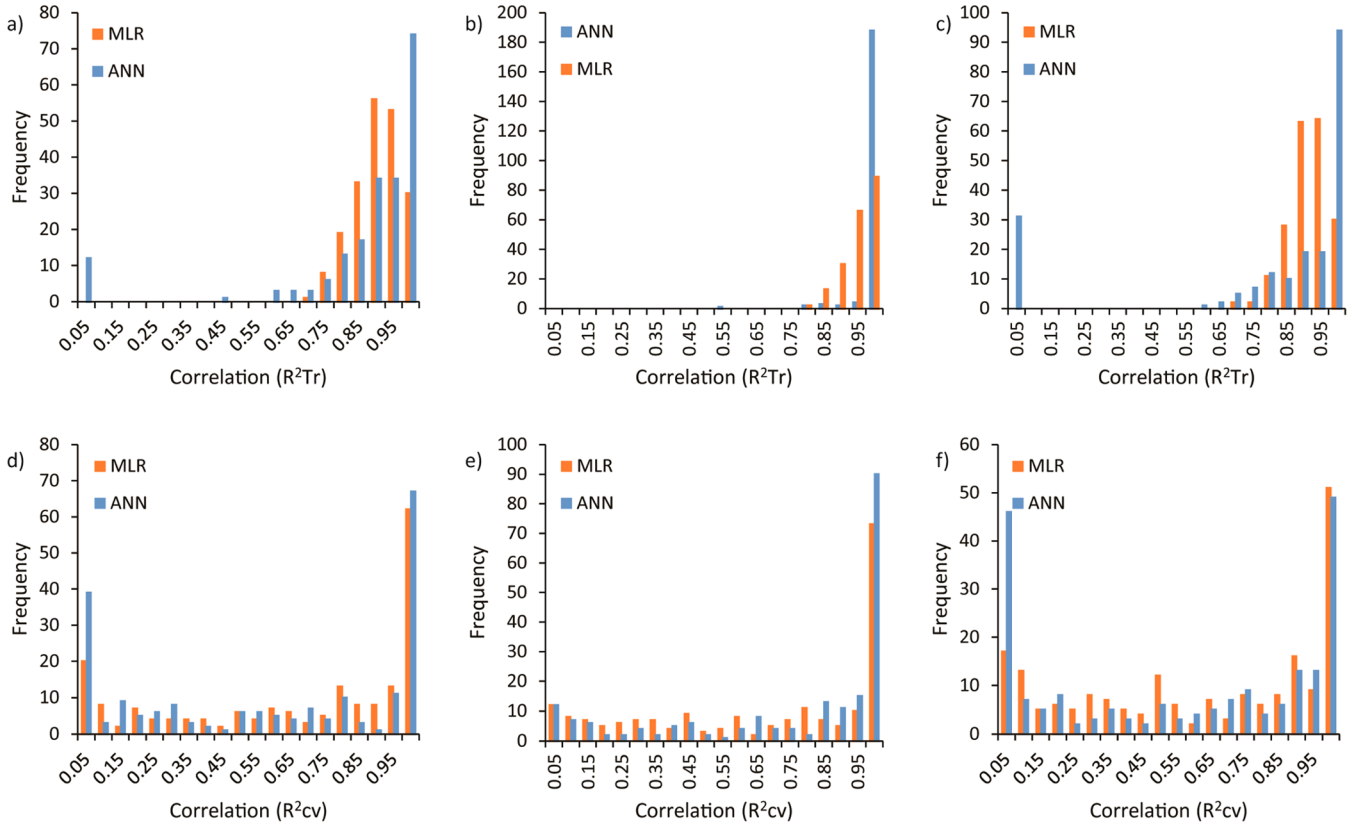

Figure 6. Histograms showing the correlation coefficients obtained from 200 repetitions of modeling on randomly prepared training (R^2_{Tr}) and prediction (R^2_{cv}) sets. (a and d) Cell viability; (b and e) cellular uptake_Fluorescence; (c and f) cellular uptake_Intensity. The histograms herein are for PC3 cells and +HP. The results for HeLa cells and −HP/+HP are in Figures S13−S15. Please note that R^2_{Tr} is another nomination for R^2_{Fit}.

important aspect has not been thoroughly addressed in a "quantitative" framework. Thus, it is essential to take into account different properties of PC (*e.g.*, composition, arrangement, structure, affinity, thickness, density, and conformation) as true evidence of the biological identity of nanoparticles in a quantitative manner and try to disclose what really happens at the nano-bio interfaces. Introducing PC descriptors in this data mining enables researchers to accurately understand and control the interactions of nanoparticles at the nano-bio interfaces which is required for the efficient design of nanoparticles for various medical applications. The use of several cellular responses herein highlighted the fact that different descriptors seem to be relevant to diverse biological processes. The well-established QSARs on different biological end points supplied clues for better understanding and uncovering fundamental mechanisms beyond different liposome−cell interactions.

Identification of specific proteins involved in the related bioresponse produces vital information, offering the potential to be used for improved liposomal design, controlled targeted delivery and other therapeutic goals in which control over nano-bio interactions is highly desired. For instance, the framework presented herein can facilitate the identification of potential inhibitors of well-known targets, such as JQ1 for BRD4 or rhodium complexes as effective STAT3 inhibitors with potent antitumor activity.[25,26]

METHODS

Samples Preparation. Cationic lipids 1,2-dioleoyl-3-trimethylammonium-propane (DOTAP) and (3β-[*N*-(*N*′,*N*′-dimethylaminoethane)carbamoyl])cholesterol (DC-Chol), zwitterionic lipids dioleoylphosphocholine (DOPC), dioleoylphosphatidylethanolamine (DOPE), DOPE-polyethylene glycol (PEG)-1K, DOPE-PEG-2K, DOPE-PEG-5K, 1,2-distearoyl-*sn*-glycero-3-phosphocholine (DSPC), 1,2-dipalmitoyl-*sn*-glycero-3-phosphocholine (DPPC), 1,2-diarachidoyl-*sn*-glycero-3-phosphocholine (20:0 PC) and anionic lipids 1,2-dioleoyl-*sn*-glycero-3-phospho-(1′-*rac*-glycerol) (DOPG), and 1,2-dioleoyl-*sn*-glycero-3-phosphate (DOPA) were purchased from Avanti Polar Lipids (Alabaster, AL). Sphingosine, Cholesterol (Chol) and DOPE labeled with 7-nitrobenzofurazan (NBD) were purchased from Sigma-Aldrich (St. Louis, MO, USA). All lipids were used without further refinement. Seventeen liposomal formulations were prepared at desired molar ratios between lipid species fixing the molar ratio ϕ = neutral lipid/total lipid (mol/mol) = 0.5. In the following text, liposomal formulations will be indicated as L1−L17. Each mixture was dissolved in chloroform and the solvent was evaporated under a vacuum for at least 24 h. Lipid films were hydrated to obtain a final lipid concentration of 1 mg/mL with ultrapure water for size, zeta potential, cytotoxicity, and flow cytometry experiments. For proteomics experiments, lipid films were hydrated with a dissolving buffer (Tris-HCl, pH 7.4, 10 mmol L^{-1}; NaCl, 150 mmol L^{-1}; EDTA, 1 mmol L^{-1}) and stored at 4 °C. The obtained liposome solutions were extruded 20 times through a 0.1 μm polycarbonate carbonate filter with the Avanti Mini-Extruder (Avanti Polar Lipids, Alabaster, AL).

Size and Zeta Potential. All size and zeta potential measurements were made at 25 °C on a Zetasizer Nano ZS90 spectrometer (Malvern, U.K.) equipped with a 5 mW HeNe laser (wavelength λ = 632.8 nm) and a digital logarithmic correlator. The normalized intensity

autocorrelation functions were analyzed using the CONTIN method,[27] to obtain the distribution of the diffusion coefficient D of the particles. This coefficient is converted into an effective hydrodynamic radius R_H using the Stokes–Einstein equation $R_H = K_BT/(6\pi\eta D)$, where K_BT is the thermal energy and η is the solvent viscosity. The electrophoretic mobility measurements were performed by means of the laser Doppler electrophoresis technique, by the same apparatus used for size measurements. The mobility u was converted into the zeta potential using the Smoluchowski relation, zeta potential = $u\eta/\varepsilon$, where η and ε are, respectively, the viscosity and the permittivity of the solvent phase. A volume of 100 μL of each liposome formulation was mixed with an equal amount of HP and the mixture was incubated for 1 h at 37 °C. After the incubation in HP, the samples were centrifuged at 14 000 rpm for 15 min at 4 °C to pellet the particle–protein complexes and separate them from the supernatant. The pellet was then resuspended in 1 mL of PBS and centrifuged again for 3 min at 14 000 rpm at 4 °C. Three washing steps were made before resuspension of the final pellet. Size and zeta potential results are given as mean ± standard deviation of five replicates.

Cell Lines. Human prostate cancer cell line (PC3), derived from human bone prostate cancer metastasis, was purchased from ATCC (Manassas, VA, USA). PC3 cells were maintained in RPMI 1640 medium supplemented with 2 mM L-glutamine, 100 IU/mL penicillin/streptomycin, 1 mM sodium pyruvate, 10 mM HEPES, 1.5 mg/L sodium bicarbonate, and 10% fetal bovine serum (FBS) (Sigma-Aldrich, St. Louis, MO, USA). Human cervical cancer cell line (HeLa), derived from human cervix adenocarcinoma, was purchase from ATCC (Manassas, VA, USA). HeLa cells were maintained in Eagle's Minimum Essential Medium (EMEM) supplemented with 2 mM L-glutamine, 100 IU/mL penicillin/streptomycin, 1 mM sodium pyruvate, 10 mM HEPES, 1.5 mg/L sodium bicarbonate, and 10% fetal bovine serum (FBS).

Protein Assay. To quantify the amount of adsorbed proteins on liposomes after incubation with HP, we used the Protein Assay reagent (Pierce, Thermo Scientific, Waltham, MA, USA), according to manufacturer's protocol. Briefly, each liposome sample was incubated with HP for 1 h at 37 °C. Then, liposome–HP complexes were centrifuged at 14 000 rpm for 15 min at 4 °C and washed 3 times with PBS, and the resulting pellet was resuspended in urea 8 mol/L NH_4CO_3 50 mmol/L. Ten microliters of each sample was placed into a 96-multiwell plate and then 50 μL of Protein Assay reagent was added. The measures were performed in triplicate. The multiwell was covered, mixed on a plate shaker, and incubated at room temperature for 5 min. The absorbance of each sample, standard and blank was measured with the Glomax Discover System (Promega, Madison, WI, USA) at 660 nm. The protein concentration was calculated using the standard curve.

Protein Identification and Quantification. A total of 100 μL of liposomes (1 mg/mL) was incubated with an equal volume of human plasma at 37 °C. Following 1-h incubation, liposome–protein complexes were centrifuged (15 min at 14 000 rpm) to remove loosely bound proteins (*i.e.*, the soft corona). Then, pellets were accurately washed three times with 100 μL of the dissolving buffer. Protein denaturation, digestion, and desalting were performed by a robust methodology that is commonly followed to separate liposome–protein complexes from unbound proteins.[11] The following step was sample lyophilization. This was achieved by means of a Speed-Vac apparatus (mod. SC 250 Express; Thermo Savant, Holbrook, NY, USA). Subsequently, samples were reconstituted with 0.1% HCOOH solution (final concentration 0.32 mg/mL) and stored at 80 °C until nano liquid chromatography (LC) tandem mass spectrometry (MS/MS) analysis. Tryptic peptides were examined by a dedicated nano-LC system (Dionex Ultimate 3000, Sunnyvale, CA, USA) connected to a hybrid mass spectrometer (Thermo Fisher Scientific Bremen, Germany), equipped with a nanoelectrospray ion source. Comprehensive experimental details can be found elsewhere.[11] Data analysis and protein validation were made according to standard procedures. Briefly, Xcalibur (v.2.07, ThermoFisher Scientific) raw data files were submitted to Proteome Discover (1.2 version, Thermo Scientific) for database search using Mascot (version 2.3.2 Matrix Science). Data were searched against SwissProt database (v 57.15, 20 266 sequences) using the decoy search option of Mascot. Final protein quantification analysis was made by Scaffold software that allows the normalization of the spectral countings (normalized spectral countings, NSCs). For each identified protein, the mean value of NSCs was normalized to the protein molecular weight (MWNSC) to obtain the relative protein abundance. Statistical significance of data was guaranteed by replicating the procedure for nine samples. Data of relative protein abundance were provided as mean ± standard deviation.

Analysis of Cell Viability. To investigate the potential toxicity arising from bare NPs, cell viability of prostate cancer cells PC3 and cervical cancer cells HeLa was assessed by 3-(4,5-dymethyl thiazol 2-y1)-2,5-diphenyl tetrazolium bromide (MTT, mitochondrial respiration analysis; Sigma-Aldrich), according to Mosmann protocol. Briefly, PC3 and HeLa cells were seeded on 96-wells plate (10 000 cells/well). The day after, cells were treated with 10 μg/mL of each formulation in Optimem medium (Life Technologies, Carlsbad, CA) for 24 h. Then, MTT was added to each well at the final concentration of 0.5 mg/mL, and after 4 h of incubation at 37 °C, the formazan salt was dissolved with 100 μL of isopropylic alcohol. The absorbance of each well was measured with Glomax Discover System (Promega, Madison, WI, USA), a ready-to-use high-performance multimode detection instrument. The viability was calculated for each treatment as (OD of treated cells/OD of control cells) × 100.

Flow Cytometry. To investigate cellular uptake of nanoparticles in PC3 and HeLa cell lines, each of the 17 liposomal formulations was synthesized with the addition of DOPE-NBD. In all the formulations, the concentration of fluorescently labeled NBD-DOPE was 7×10^{-3} mg/mL (fluorescent lipid/total lipid molar ratio = 5/1000). Bare liposomes and liposome–HP complexes were administered to cells cultured with serum-free medium. PC3 and HeLa cells were plated at 200 000 cells/mL in 12-well dishes. After 24 h, cells were incubated for 3 h with 10 μg/mL of NBD-labeled liposomes. After the treatment, cells were detached with trypsin/ethylenediaminetetraacetic acid (EDTA), washed twice with cold PBS, and run on a BD LSFORTESSA (BD Biosciences, San Jose, CA, USA). Cells were gated using forward *vs* side scatter to exclude debris. Data were analyzed using FlowJo software (FlowJo LLC data analysis software, Ashland, OR, USA).

Descriptor Generation. A set of 656 descriptors was collected for 17 liposome formulations, consisting of 633 PC fingerprints (PCF) and 23 liposome physicochemical properties (LPP). The full descriptions of all the 656 descriptors together with their values are given in Table S1 of Supporting Information. The first action taken on the data set was initial screening for removing highly correlated descriptors. Descriptors with zero or constant values and also the ones with correlation coefficients higher than 80% with each other (multicollinear) were removed from the initial pool of variables. Thus, the data set consisting of 17 liposome formulations and 148 descriptors was prepared for the QSAR studies. Table 2 represents the two classes of descriptors used in the QSAR analysis.

QSAR Modeling. Linear and nonlinear QSAR models were used to predict 12 different sets of biological end points for 17 liposome formulations. The response vectors were $\log_2$ transformed and included cell viability, cellular uptake_fluorescent, and cellular uptake_intensity, each measured for two cell types (PC3 and HeLa) and with/without HP incubation (= 3 × 2 × 2).

In Multiple Linear Regression (MLR),[28,29] a simple linear regression was used to fit the data and minimize the root-mean-square error between the measured (y_i) and predicted ($\hat{y}_i$) responses for each liposome:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

Another parameter that is widely used for evaluating the performance of a QSAR model is the coefficient of determination:

$$R^2 = 1 - \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 / \sum_{i=1}^{n} (y_i - \bar{y}_i)^2$$

$\bar{y}_i$ is the mean response measured for each liposome. Depending on the n samples coming from the training set or the prediction set, for which $\hat{y}_i$ is calculated, the coefficient of determination would be referred to as R^2_{fit} or R^2_{cv}, respectively.

As a nonlinear model, Artificial Neural Networks (ANN) was utilized to correlate the LPP and PCF descriptors to different bioresponse profiles. ANN with a layered structure is a mathematical system stimulating the biological neural network, consisting of neurons and synapses. The measured variables are presented to an input layer and are processed, by one or more intermediate ("hidden") layers, to produce one or more outputs. Synapses are responsible for the connections between these input, hidden and output neurons. The more the strength of the synapses, the heavier the weight of the connections between different neuron layers. The weights and biases are two internal parameters in a neural network that iteratively change during training in order to minimize a performance function which is usually the discrepancy between the observed and predicted values (*i.e.*, error between network inputs and target outputs). The performance function herein was set to the root minimum square error, as shown in the above formula. Hence, the neural network will be trained until the desired degree of accuracy is achieved. There are several algorithms for training multilayer feed forward neural networks.

In this study, we used Bayesian Regularization (BR)[30] in combination to Levenberg–Marquardt training algorithm to develop nonlinear ANN models. BR is a powerful optimization procedure in ANN modeling and is a beneficial approach to deal with the overfitting (*i.e.*, occurs when the model has memorized the training samples but has not learned to generalize to new samples, leading to too high degree of accuracy for training samples but poor prediction for new samples). BR as a solution to this regularization problem modifies the performance function so that the network will have smaller weights and biases and its response will be smoother and less likely to overfit. Unlike most of the linear models, ANN does not start by assuming a particular type of mathematical relationship between the input and output variables. This makes ANN particularly useful when the underlying model is unknown. However, the operation of ANN is like a black box, comprising complicated mathematical and systematic relations. Therefore, in QSAR models, ANNs are generally used as purely predictive tools rather than as an aid in understanding structure–property trends, although great prediction ability, high reproducibility, and generalization power of BRANN modeling makes it superior and offers the opportunity to explore nonlinear phenomena and curved manifold. Since biological subjects may have nonlinear characteristics, ANN techniques are beneficial for discovering the possible relationship between the input descriptors and the output bioresponses.

For each liposome, the biological responses where predicted using both MLR and ANN models. The QSAR analysis began by randomly splitting the data set into two subsets, called prediction set (20% of the raw data) and training set (the remaining 80% of the raw data). The training set is used to build the model which is then validated by the prediction set. To achieve a good QSAR model, a minimal set of information-rich descriptors is needed, requiring a suitable variable selection step to be applied.[31] Stepwise regression as a simple yet powerful variable selection method is an iterative selection procedure which starts from a variable with the largest empirical correlation with the dependent variable (response). Each iteration of stepwise regression includes two phases: the inclusion phase in which each remaining variable is subjected to a partial F-test, and the variable with the F-value larger than a critical 'F-to-enter' value is inserted in the model; and the exclusion phase, in which the variable with the F-value smaller than a critical 'F-to-remove' value will be removed from the model and returned to the pool of variables still available for the selection. Thus, in each iteration of the initial splitting, the training set was applied to stepwise regression variable selection method to choose the best descriptors among the large number of descriptors.

To deal with overfitting in both variable selection and modeling steps, Monte Carlo cross validation was used.[32−34] This technique involves a large number of random splits of the data set repeatedly, in each of which the available data is divided into two groups to be used for the training and testing. The criterion (*e.g.*, root-mean-square error or coefficient of determination) is averaged over all the repeated splits, so as to not tie the measure to one particular division of the data and to be representative of the whole data. For this reason, the training set was itself randomly divided into two subsets (train and test) for 100 times and the most frequent variables were chosen afterward. In each iteration of the outer loop (200 runs), after 100 times of repeating variable selection, the prediction set with the best selected descriptors was then applied to the linear MLR or nonlinear ANN model to assess its prediction ability. The algorithm followed this procedure until 200 runs of the outer loop were completed. The results were averaged over the corresponding repetitions. It should be noted that the selected descriptors from each iteration were recorded and the best ones where subsequently chosen based on the variables frequency plot which was illustrated after completing the whole iterations. The descriptors with the highest frequency having reasonable difference with others were represented as the best of the best selected descriptors, most suitable for correlation.

Sensitivity Analysis and Descriptor Importance. To assess the importance of the selected descriptors (which were identified based on their higher frequency in incorporating in the model development), sensitivity analysis was used to measure their relative importance in the predictive ability of the model.[35] In this approach, each descriptor was in turn randomly scrambled and the RMSE of the resulting model was calculated and compared with that when all the descriptors were available. Therefore, an increase in RMSE (or decrease in the model performance) according to scrambling a selected descriptor represents the relative importance of the corresponding descriptor. In other words, the importance priority of the selected descriptors is achieved based on the amount of decrease in the model performance. In addition, the frequency plot of descriptors illustrates how many times each descriptor has participated in the QSAR procedure. The higher a variables' frequency, the more it has been displayed in variable selection and the better it correlates to activities.

Model Robustness. In addition to Monte Carlo cross validation which was used to quantify the prediction abilities of the developed QSAR models (the histogram of Monte Carlo cross validation correlation coefficients shows the distribution of R^2_{cv} in 200 repetitions), the robustness of the models was also further examined. The statistical significance of QSAR models was checked by comparing its measure of fit to the average measure of fit when the response vector is completely randomized.[25,26] This approach is called Y-randomization or Y-scrambling, and is used to check if the model is built upon chance correlations. By scrambling the response vector, it is expected to get much lower R^2 values, indicating the significance of the original QSAR model which has been now destroyed by this shuffling. Thus, a 100-round Y-randomization was applied to the data to confirm that the developed models were not chance correlations.

Applicability Domain. William's graph as the most common method was applied to define the applicability domain of each QSAR model.[34,36] The applicability domain represents the theoretical descriptor space in which reliable QSAR predictions can be made. William's graph depicts a QSAR's applicability domain with a two-dimensional scatter plot. The first dimension which reflects the structural similarity between one sample and all other samples in the training set is the "leverage" which is calculated for each sample (x_i) in the data set (X) by

$$h_i = x_i (X^{\mathrm{T}} X)^{-1} x_i^{\mathrm{T}}$$

The leverage quantitatively expresses the distance of a given liposome to the center of all liposome samples used for QSAR development in the descriptor space. Liposomes with smaller h value are more similar to the data set used for QSAR modeling, meaning that they are within the descriptor space in which the bioresponse can be predicted more

reliably. A threshold leverage value, which covers about 99% of normally distributed training samples, is usually calculated by

$$h^* = \frac{3(m+1)}{n}$$

where m and n are the number of descriptors in the model and number of training samples, respectively. Liposomes with leverage values higher than this warning threshold value are defined as outliers, meaning that they do not fit well in the current model, as others.

The second dimension of William's graph is the standardized prediction error:

$$\varepsilon_i = \frac{e_i - \bar{e}}{\sigma_e}$$

where $\bar{e}$ and σ_e are the average and standard deviation of the residual error ($=y_i - \hat{y}_i$), respectively. A sample is considered as outlier if its absolute ε_i value is greater than 3.

AUTHOR INFORMATION

Corresponding Authors

*E-mail: giulio.caracciolo@uniroma1.it.

*E-mail: mahmoudi-m@tums.ac.ir, Mahmoudi@stanford.edu.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

M.M. would like to thank Tehran University of Medical Sciences for supporting this work (Grant No. 94-04-33-31032). G.C. and D.P. acknowledge support by the Italian Minister for University and Research (MIUR) (Futuro in Ricerca 2008, Grant No. RBFR08TLPO).

REFERENCES

(1) Allen, T. M.; Cullis, P. R. Liposomal Drug Delivery Systems: From Concept to Clinical Applications. *Adv. Drug Delivery Rev.* **2013**, *65*, 36−48.

(2) Lasic, D. D.; Papahadjopoulos, D. *Medical Applications of Liposomes*; Elsevier Science: Amsterdam, 1998; pp 1−13.

(3) Al-Jamal, W.; Kostarelos, K. Liposomes: From a Clinically Established Drug Delivery System to a Nanoparticle Platform for Theranostic Nanomedicine. *Acc. Chem. Res.* **2011**, *44*, 1094−1104.

(4) Lynch, I.; Salvati, A.; Dawson, K. A. Protein-nanoparticle Interactions: What Does the Cell See? *Nat. Nanotechnol.* **2009**, *4*, 546−547.

(5) Walczyk, D.; Bombelli, F. B.; Monopoli, M. P.; Lynch, I.; Dawson, K. A. What the Cell "sees" in Bionanoscience. *J. Am. Chem. Soc.* **2010**, *132*, 5761−5768.

(6) Lundqvist, M.; Stigler, J.; Elia, G.; Lynch, I.; Cedervall, T.; Dawson, K. A. Nanoparticle Size and Surface Properties Determine the Protein Corona with Possible Implications for Biological Impacts. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 14265−14270.

(7) Walkey, C. D.; Chan, W. C. W. Understanding and Controlling the Interaction of Nanomaterials with Proteins in a Physiological Environment. *Chem. Soc. Rev.* **2012**, *41*, 2780−2799.

(8) Caracciolo, G. Liposome−Protein Corona in a Physiological Environment: Challenges and Opportunities for Targeted Delivery of Nanomedicines. *Nanomedicine* **2015**, *11*, 543−557.

(9) Pozzi, D.; Colapicchioni, V.; Caracciolo, G.; Piovesana, S.; Capriotti, A. L.; Palchetti, S.; Grossi, S. D.; Riccioli, A.; Amenitsch, H.; Lagana, A. Effect of Polyethyleneglycol (PEG) Chain Length on the Bio−nano-interactions Between PEGylated Lipid Nanoparticles and Biological Fluids: From Nanostructure to Uptake in Cancer Cells. *Nanoscale* **2014**, *6*, 2782−2792.

(10) Capriotti, A. L.; Caracciolo, G.; Cavaliere, C.; Foglia, P.; Pozzi, D.; Samperi, R.; Lagana, A. Do Plasma Proteins Distinguish Between Liposomes of Varying Charge Density? *J. Proteomics* **2012**, *75*, 1924−1932.

(11) Barran Berdon, A. L.; Pozzi, D.; Caracciolo, G.; Capriotti, A. L.; Caruso, G.; Cavaliere, C.; Riccioli, A.; Palchetti, S.; Lagana, A. Time Evolution of Nanoparticle-Protein Corona in Human Plasma: Relevance for Targeted Drug Delivery. *Langmuir* **2013**, *29*, 6485−6494.

(12) Lundqvist, M. Nanoparticles: Tracking Protein Corona over Time. *Nat. Nanotechnol.* **2013**, *8*, 701−702.

(13) Mahmoudi, M.; Abdelmonem, A. M.; Behzadi, S.; Clement, J. H.; Dutz, S.; Ejtehadi, M. R.; Hartmann, R.; Kantner, K.; Linne, U.; Maffre, P.; Metzler, S.; Moghadam, M. K.; Pfeiffer, C.; Rezaei, M.; Ruiz-Lozano, P.; Serpooshan, V.; Shokrgozar, M. A.; Nienhaus, G.; Ulrich; Parak, W. J. Temperature: The "Ignored" Factor at the NanoBio Interface. *ACS Nano* **2013**, *7*, 6555−6562.

(14) Caracciolo, G.; Pozzi, D.; Capriotti, A. L.; Cavaliere, C.; Foglia, P.; Amenitsch, H.; Lagana, A. Evolution of the Protein Corona of Lipid Gene Vectors as a Function of Plasma Concentration. *Langmuir* **2011**, *27*, 15048−15053.

(15) Hajipour, M. J.; Laurent, S.; Aghaie, A.; Rezaee, F.; Mahmoudi, M. Personalized Protein Coronas: A "key" Factor at the Nano-biointerface. *Biomater. Sci.* **2014**, *2*, 1210−1221.

(16) Caracciolo, G.; Pozzi, D.; Capriotti, A. L.; Cavaliere, C.; Piovesana, S.; La Barbera, G.; Amici, A.; Lagana, A. The Liposome−Protein Corona in Mice and Humans and Its Implications for *in vivo* Delivery. *J. Mater. Chem. B* **2014**, *2*, 7419−7428.

(17) Walkey, C. D.; Olsen, J. B.; Song, F.; Liu, R.; Guo, H.; Olsen, D. W. H.; Cohen, Y.; Emili, A.; Chan, W. C. W. Protein Corona Fingerprinting Predicts the Cellular Interaction of Gold and Silver Nanoparticles. *ACS Nano* **2014**, *8*, 2439−2455.

(18) Liu, K.; Jiang, W.; Walkey, C. D.; Chan, W. C. W.; Cohen, Y. Prediction of Nanoparticles-Cell association based on Corona Proteins and Physicochemical Properties. *Nanoscale* **2015**, *7*, 9664−9675.

(19) Pozzi, D.; Caracciolo, G.; Digiacomo, L.; Colapicchioni, V.; Palchetti, S.; Capriotti, A. L.; Cavaliere, C.; Zenezini Chiozzi, R.; Puglisi, A.; Lagana, A. The biomolecular corona of nanoparticles in circulating biological media. *Nanoscale* **2015**, *7*, 13958−13966.

(20) Tenzer, S.; Docter, D.; Kuharev, J.; Musyanovych, A.; Fetz, V.; Hecht, R.; Schlenk, F.; Fischer, D.; Kiouptsi, L.; Reinhardt, C.; Landfester, K.; Schild, H.; Maskos, M.; Knauer, S. K.; Stauber, R. H. Rapid Formation of Plasma Protein Corona Critically Affects Nanoparticle Pathophysiology. *Nat. Nanotechnol.* **2013**, *8*, 772−781.

(21) Doorley, G. W.; Payne, C. K. Cellular Binding of Nanoparticles in the Presence of Serum Proteins. *Chem. Commun.* **2011**, *47*, 466−468.

(22) Huhn, D.; Kantner, K.; Geidel, C.; Brandholt, S.; Cock, I. D.; Soenen, S. J. H.; Rivera_Gil, P.; Montenegro, J. M.; Braeckmans, K.; Mullen, K.; Nienhaus, G. U.; Klapper, M.; Parak, W. J. Polymer-Coated Nanoparticles Interacting with Proteins and Cells: Focusing on the Sign of the Net Charge. *ACS Nano* **2013**, *7*, 3253−3263.

(23) Lynch, I.; Dawson, K. A. Protein-Nanoparticle Interactions. *Nano Today* **2008**, *3*, 40−47.

(24) Tenzer, S.; Docter, D.; Rosfa, S.; Wlodarski, A.; Kuharev, J.; Rekik, A.; Knauer, S. K.; Bantz, C.; Nawroth, T.; Bier, C.; Sirirattanapan, J.; Wolf, M.; Lennart, T.; Reinhard, Z.; Michael, M.; Hansjorg, S.; Stauber, R. H. Nanoparticle Size Is a Critical Physicochemical Determinant of the Human Blood Plasma Corona: A Comprehensive Quantitative Proteomic Analysis. *ACS Nano* **2011**, *5*, 7155−7167.

(25) Filippakopoulos, P.; Qi, J.; Picaud, S.; Shen, Y.; Smith, W. B.; Fedorov, O.; Morse, E. M.; Keates, T.; Hickman, T. T.; Felletar, I.;

Phipott, M.; Munro, S.; McKeown, M. R.; Wang, Y.; Christie, A. L.; West, N.; Cameron, M. J.; Schwartz, B.; Heightman, T. D.; Thangue, N. L.; French, C. A.; Wiest, O.; Kung, A. L.; Knapp, S.; Bradner, J. E. Selective Inhibition of BET Bromodomains. *Nature* **2010**, *468*, 1067−1073.
(26) Ma, D. L.; Liu, L. J.; Leung, K. H.; Chen, Y. T.; Zhong, H. J.; Chan, D. S. H.; Wang, H. M. D.; Leung, C. H. Antagonizing STAT3 Dimerization with a Rhodium(III) Complex. *Angew. Chem., Int. Ed.* **2014**, *53*, 9178−9182.
(27) Provencher, S. W. A Constrained Regularization Method for Inverting Data Represented by Linear Algebraic or Integral Equations. *Comput. Phys. Commun.* **1982**, *27*, 213−227.
(28) Gemperline, P. In *Practical Guide to Chemometrics*, 2nd ed.; CRC Press, Taylor and Francis, LLC, 2006; pp 111−116.
(29) Miller, J. N.; Miller, J. C. In *Statistics and Chemometrics for Analytical Chemistry*, 5th ed.; Pearson: Harlow, U.K., 2005; pp 228−237.
(30) MacKay, D. J. C. A Practical Bayesian Framework for Backpropagation Networks. *Neural Comput.* **1992**, *4*, 448−472.
(31) Anderson, C. M.; Bro, R. Variable Selection in Regression. *J. Chemom.* **2010**, *24*, 728−737.
(32) Gramatica, P. Principles of QSAR Models Validation: Internal and External. *QSAR Comb. Sci.* **2007**, *26*, 694−701.
(33) Eriksson, L.; Jaworska, J.; Worth, A. P.; Cronin, M. T. D.; McDowell, R. M.; Gramatica, P. Methods for Reliability, Uncertainty Assessment, and Applicability Evaluations of Classification and Regression Based QSARs. *Environ. Health Perspect.* **2003**, *111*, 1361−1375.
(34) Puzyn, T.; Leszczynska, D.; Leszczynski, J. Toward the Development of 'Nano-QSARs': Advances and Challenges. *Small* **2009**, *5*, 2494−2509.
(35) Guha, R.; Jurs, P. C. Interpreting Computational Neural Network QSAR models: A Measure of Descriptor Importance. *J. Chem. Inf. Model.* **2005**, *45*, 800−806.
(36) Tropsha, A.; Gramatica, P.; Gombar, P. V. The Importance of Being Earnest: Validation is the Absolute Essential for Successful Application and Interpretation of QSPR Models. *QSAR Comb. Sci.* **2003**, *22*, 69−77.