Small-angle neutron scattering reveals the nanostructure of liposomes with embedded OprF porins of *Pseudomonas Aeruginosa*

Francesco Spinozzi,[†] Jean-Pierre Alcaraz,[‡] Maria Grazia Ortore,[†] Landry Gayet,[‡] Aurel Radulescu,[¶] Donald K. Martin,[‡] and Marco Maccarini[*,‡]

[†]*Department of Life and Environmental Sciences, Polytechnic University of Marche, 60131 Ancona, Italy*

[‡]*Univ. Grenoble Alpes, CNRS, UMR 5525, VetAgro Sup, Grenoble INP, TIMC, 38000 Grenoble, France*

[¶]*Jülich Centre for Neutron Science JCNS at Heinz Maier-Leibnitz Zentrum (MLZ), Forschungszentrum Jülich GmbH, Garching*

E-mail: marco.maccarini@univ-grenoble-alpes.fr

Abstract

The use of liposomes as drug delivery systems emerged in the last decades in view of their capacity and versatility to deliver a variety of therapeutic agents. By means of small angle neutron scattering (SANS), we performed a detailed characterisation of liposomes containing OprF, the main porin of the *Pseudomonas Aeruginosa* bacterium outer membrane. These OprF-liposomes are at the basis of a novel vaccine against this antibiotic resistant bacterium, which is one of the main hospital-acquired pathogens and causes each year a significant number of deaths. SANS data were analysed by a specific model we created to quantify crucial information about the structure of the liposome

containing OprF, including the lipid bilayer structure, the amount of protein in the lipid bilayer, the average protein localisation and the effect of the protein incorporation on the lipid bilayer. Quantification of such structural information is important to enhance the design of liposomal delivery systems for therapeutic applications.

Introduction

Liposomes are classes of nanoscopic materials that possess excellent capabilities to deliver drugs and a variety of therapeutic agents into the human body. An important feature of these systems is that hydrophilic molecules can be contained in the inner part of the liposome, while hydrophobic molecules can be inserted into the hydrocarbon chains, thus showing a great versatility in the type of molecules they can carry. Many classes of liposomes of different size, shape and composition as drug carriers have been developed[1] including the m-RNA vaccines used in the current Covid-19 pandemic.[2,3]

A detailed structural characterisation of liposomes containing molecular drugs or therapeutical agents is a key factor for their development and optimisation. Size and lipid composition in addition to the presence, the amount, the localisation and orientation of non-lipidic molecular components embedded in the lipidic membrane may influence the interaction of liposomes with human cells. The use of small angle neutron and X-ray scattering provides averaged structural information on length-scales ranging from micrometer to sub nanometer scales. When applied to liposomes, these scattering techniques provide information about size and shape as well as a variety of details of the lipid and of the non-lipid components, such as thickness, roughness and hydration of the head and tail regions, amount and localisation of guest molecules.[4,5]

In this article we describe a structural characterisation performed with SANS of liposomes containing the outer membrane protein F (OprF), the main porin of *Pseudomonas aeruginosa*. This gram-negative human pathogen is responsible for a significant number of nosocomial acquired infections that are resistant to most of the common antibiotics.[6] This

resistance is mainly due to the low permeability of the outer membrane that it is thought to be controlled by the OprF porin.[7] OprF is also involved in the mechanisms of adhesion with host cell and in the interaction with gamma interferon for evasion of the host immune system.[8] It was also recognised that OprF might impact the biogenesis of outer membrane vesicles used for cell to cell communication and for trafficking hydrophobic signals within a bacterial population.[9,10]

Structurally, OprF appears in two conformation states: (i) the closed conformation, formed by two domains, the N-terminal included in the outer membrane, and the C-terminal residing outside the membrane as a globular domain; and (ii) the open state conformation, fully contained in the outer membrane.[11] The closed conformation represents around 95% of OprF population.[12] This is consistent with the low permeability of *P. aeruginosa* outer membrane.

As other outer membrane proteins of *P. aeruginosa*, OprF is an ideal antigen that was considered as vaccine candidate.[13–18] OprF is a highly conserved among all the serotypes of *P. aeruginosa* and it exposes some epitopes on the surface of the bacteria that can be recognised by the immune system. It was shown by Lenormand and coworkers[19] that fully functional recombinant OprF, obtained by cell-free expression in form of proteoliposomes, generates a strong immune response against cell-associated hemolytic activity (CHA) strain of *P. aeruginosa* challenge when injected in mice. Importantly, that study showed it possible to reconstruct the entire OprF membrane antigen, and hence all the conformational epitopes of the protein, in the lipid bilayer of a synthetic liposome in large amounts. The mice vaccinated with the injection of the OprF proteoliposomes not only survived and recovered after lethal doses of a CHA strain, but showed a strong cross-protection conferred by sera from immunised mice. Inoculation with the serum provided total protection against infection.

OprF proteoliposomes produced with the same method utilised for the vaccine candidate[19] were analysed with a novel model to calculate small angle scattering form factors. The model presented in this manuscript describes the form factor of the proteoliposome

through a combination of volume fraction radial distributions of relevant molecular groups appropriately chosen. This volume fraction approach, already applied to planar or cylindrical geometry,[20,21] is here extended to particles with spherical symmetry. The analysis returned key nanostructural information of the OprF proteoliposome, including average structural information on the lipid bilayer, the localisation of the OprF and the effect of its incorporation on the lipid bilayer. Importantly, this methodology can be applied to study liposomes formed by a mixture of lipids and containing proteins or other compounds in the lipid bilayer, either in its interior or on the liposome surface.

Materials and Methods

Synthesis of OprF Proteoliposomes

OprF proteoliposomes were produced with the cell-free expression technique.[22] Briefly, liposomes were prepared starting from a lipid mixture of DOPC:DOPE:Cholesterol:DMPA (2:1:1:1) mole proportion. The mixture in chloroform was firstly dried under nitrogen during almost 2 hours. The film was then resuspended and then solution was sonicated 3 times 30 s (5 s pulses). The resulting solution was then filtered with a 0.2 μm polyethersulfone (PES) filter. Liposomes at a concentration of 3.3 g/L were mixed with the cell free lysate and the plasmids. After the expression of the membrane protein, the crude extract was centrifuged and washed with salt solution (5 M NaCl) to remove extrinsic proteins from the proteoliposomes. The resulting proteoliposomes purity was confirmed by on SDS-PAGE to be around $90 - 95\%$. After the centrifugation, the pellet containing the protein was resuspended in an adequate volume to obtain a protein concentration of 1 g/L. The concentration of lipids in the proteoliposomes was not estimated.

Sample preparation

Liposomes were prepared starting from a lipid mixture of DOPC:DOPE:Cholesterol:DMPA (2:1:1:1) mole proportion with a concentration of 5 g/L in D_2O/H_2O mixtures. The lipids were dissolved in water and sonicated. Then the resulting solution was filtered with a PES filter of 0.2 μm. Once they were filtered, their relative concentration might have slightly changed, because some lipids can be preferentially trapped in the sieve compared to others. Dynamic light scattering experiments (DLS) showed that the radius size was distributed around 150 nm.

Proteoliposomes were prepared with the cell-free expression.[22] Liposomes prepared as described above at a concentration of 3.3 g/L were mixed with the lysate and the plasmids. After the expression, the mixture was centrifuged and washed with salts solution to remove all the rest of the lysate. The resulting purity of proteoliposomes was around $90-95\%$. After the centrifugation, the pellet containing the protein was resuspended in an adequate volume to obtain a protein concentration of 1 g/L. The concentration of lipids was not estimated. 140 μL of the proteoliposome suspensions were diluted to 400 μL with proper amounts of H_2O and D_2O to obtain a total solvent deuteration grade, $x_D = [D_2O]/([D_2O] + [H_2O])$, of 0, 0.1, 0.14 and 0.4 Hence the final protein concentration in the SANS measurement cells was about 0.5 g/L, whereas the nominal calculated lipid concentration was around 1.2 g/L, with a possible variation of the molar ratios among the lipids molecules. DLS measurements showed that the proteoliposome radius is around 1000 nm with a polidispersity index (PDI) around 0.25. Salt traces in the samples are estimated to be lower than 150 mM.

SAS Model

We have developed a novel method to calculate the SAXS or SANS form factor of a spherical vesicle with polydispersed radius surrounded by an asymmetric bilayer. The method is based on the volume fraction approach, which has been widely applied in planar geometry[4,20,23–25] and, more recently, in cylindrical geometry.[21] Molecules constituting the inner or the outer

monolayer may comprise one or several kinds of lipids, as well as embedded proteins or other compounds. The solution inside the vesicle could have a composition different from the one of the bulk solution. According to the molar composition of all the molecules forming a monolayer, two *molecular units* can be identified so that the spherical shell associated to each monolayer can be seen by the self-assembling of many corresponding molecular units. Following the standard scattering density profile (SDP)[4,23–25] approach, the molecular unit is described in terms of *united groups*, each one characterized by two relevant properties from the scattering point of view: the volume and the *scattering length*.

We will now describe the details regarding the novel method.

Inner and outer monolayers are hereafter labeled by the symbol $\alpha = \text{in}$ and $\alpha = \text{out}$, respectively, and we indicate with $N_{\text{sa},\alpha}$ the corresponding number of self-assembled molecular units. For each united group of the molecular unit in the α monolayer (from now on labelled with the subscript $_{g,\alpha}$), we assume to know their number (which depends on the composition of the molecular unit), volume and scattering length, indicated with $\text{n}_{g,\alpha}$, $\nu_{g,\alpha}$ and $b_{g,\alpha}$, respectively. Hence the *scattering length density* (SLD) of the g-group in the α-monolayer will be $\rho_{g,\alpha} = b_{g,\alpha}/\nu_{g,\alpha}$. It is worth to treat in a distinct way the hydrophobic and the polar groups of the molecular units as well as the hydration molecules (water, co-solvents as well as other compounds randomly dispersed in solution) that are close to the polar groups. Let $N_{\text{h},\alpha}$ and $N_{\text{p},\alpha}$ be the number of hydrophobic and polar groups necessary to describe the molecular unit in the α monolayer, respectively. We start to label these groups from the hydrophobic to the polar ones. The group $g = 1$ is considered the one that mostly occupies the total hydrophobic domain volume: typically it is the methylene CH_2 group of the aliphatic chains. The group with the index $g = N_{\text{h},\alpha} + N_{\text{p},\alpha} + 1$ comprises all the hydration molecules: we assume that the composition and the molecular volumes of these molecules inside the vesicle ($\alpha = \text{in}$), including the one in contact with the inner polar groups, are unique and determine a SLD indicated with $\rho_{N_{\text{h,in}}+N_{\text{p,in}}+1,\text{in}}$ and that the bulk solution could have a different composition and hence a different SLD indicated with ρ_0. Moreover, accord-

ing to Berndt et al.[26], we assume that the molecular volume of the solvent (water in most cases) in the vicinity of the polar groups of the outer monolayer ($\alpha = \text{out}$) could be different from the one of the bulk solvent: we hence indicate with $\rho_{N_{\text{h,out}}+N_{\text{p,out}}+1,\text{out}}$ the SLD of the outer hydration molecules.

At the heart of the method there is the concept of *group volume fraction radial distribution*: it is the function $\varphi_{g,\alpha}(r)$ that indicates the volume fraction occupied by the g-group in the α-monolayer in the spherical shell placed at a distance from the vesicle center comprised between r and $r + dr$. Clearly, since these functions must represent fractions, they should obey to the constraint $\sum_{g=1}^{N_{\text{h},\alpha}+N_{\text{p},\alpha}+1} \varphi_{g,\alpha}(r) = 1$. Moreover, the volume integral in spherical coordinates of $\varphi_{g,\alpha}(r)$ should give the total volume occupied by the g-group, hence

$$4\pi \int_0^\infty r^2 \, \varphi_{g,\alpha}(r) \, dr = N_{\text{sa},\alpha} \text{n}_{g,\alpha} \nu_{g,\alpha}. \tag{1}$$

In the SDP method,[4,23–25,27–29] volume fraction distributions are modeled by Gaussians, whereas in the modified SDP (MSDP)[20] method they are a combination of error functions. Here, according to Berndt et al.[26], in order to simplify the Fourier transform calculation in spherical geometry, we have chosen to model the group volume fraction radial distributions as peaks obtained by a combinations of four parabolas. Before the proper normalization,

this kind of peak is defined by the expression

$$f(r) = \begin{cases} 0 & r \leq r_1 \\ \frac{(r_1-r)^2}{2(\xi_l w)^2} & r_1 < r \leq r_2 \\ 1 - \frac{(r_3-r)^2}{2(\xi_l w)^2} & r_2 < r \leq r_3 \\ 1 & r_3 < r \leq r_5 \\ 1 - \frac{(r_5-r)^2}{2(\xi_r w)^2} & r_5 < r \leq r_6 \\ \frac{(r_7-r)^2}{2(\xi_r w)^2} & r_6 \leq r \leq r_7 \\ 0 & r \geq r_7 \end{cases} , \tag{2}$$

$$r_1 = R + r_0 - w - \xi_l w,$$

$$r_2 = R + r_0 - w,$$

$$r_3 = R + r_0 - w + \xi_l w,$$

$$r_4 = R + r_0,$$

$$r_5 = R + r_0 + w - \xi_r w,$$

$$r_6 = R + r_0 + w,$$

$$r_7 = R + r_0 + w + \xi_r w.$$

A representative plot of $f(r)$ is reported in Figure 1. In this equation R is the radius of

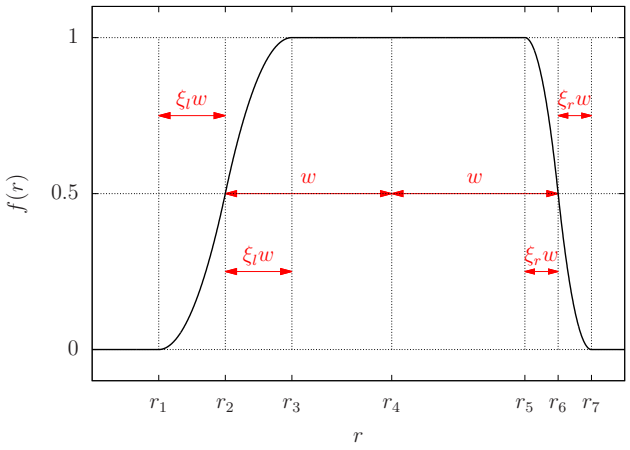

Figure 1: Plot of the function $f(r)$. See Eq. 2 for the definition of r_i with $i = 1, 7$.

the vesicle, defined by the distance from the vesicle center to the interface between inner and outer monolayers, $R + r_0 = r_4$ is the peak position, $w = r_6 - r_4 = r_4 - r_2$ is half the width at half the height (in short the peak wideness) and $\xi_l = (r_3 - r_2)/w = (r_2 - r_1)/w$ and $\xi_r = (r_6 - r_5)/w = (r_7 - r_6)/w$ (both ranging from 0 to 1) are smoothness parameters that modulate the parabolic profile on the left side and on the right side of the peak, respectively. To note, Eq. 2 shows that $f(r)$ is different from 0 only in the range $r_1 \leq r \leq r_7$. Hence, its volume integral can be easily calculated leading to

$$\begin{aligned} V_f &= 4\pi \int_{r_1}^{r_7} r^2 f(r) dr \\ &= \frac{2\pi w}{3}(12r_4^2 + w^2(4 + \xi_r^2 + \xi_l^2) + 2wr_4(\xi_r^2 - \xi_l^2)) \end{aligned} \tag{3}$$

According to Eq. 1, the group volume fraction radial distribution can be then expressed by

$$\varphi_{g,\alpha}(r) = \frac{N_{\mathrm{sa},\alpha} \mathrm{n}_{g,\alpha} \nu_{g,\alpha}}{V_f} f_{g,\alpha}(r) \tag{4}$$

where the symbol $f_{g,\alpha}(r)$ indicates the function $f(r)$ corresponding to the g-group of the α-monolayer, with $r_0 = r_{g,\alpha}$, $w = w_{g,\alpha}$ and $\xi_l = \xi_r = \xi_{g,\alpha}$. Notice that, for the sake of simplicity, we have assumed that the left and the right smoothness parameters are equal. We should consider that groups in the inner monolayer will have negative values of $r_{g,\alpha}$, whereas for the ones on the outer monolayer $r_{g,\alpha}$ is positive.

In order to avoid the presence of water molecules among the hydrophobic groups, following De Rosa et al.[20], we consider that the peak function defined in Eq. 2, whose maximum value is 1, can be well suited, without any normalization, to represent the volume fraction radial distribution of the *α-hydrophobic domain* (which include all the $N_{\mathrm{h},\alpha}$ hydrophobic groups of the α-monolayer), indicated with $\varphi_{\mathrm{h},\alpha}(r)$. Let us call $D_{\mathrm{h},\alpha}$ the thickness of the hydrophobic domain: considering the meaning of the parameters defining Eq. 2 and its behavior reported in Figure 1, for the inner monolayer we should consider $r_6 = R$, $r_6 - r_2 = D_{\mathrm{h,in}}$ and

$r_4 = R - D_{\mathrm{h,in}}/2$, whereas for the outer monolayer we have $r_2 = R$, $r_6 - r_2 = R + D_{\mathrm{h,out}}$ and $r_4 = D_{\mathrm{h,out}}/2$. Moreover, we consider that a unique smoothness length, defined as σ_{term}, could describe the junction between inner ($\sigma_{\mathrm{term}} = r_7 - r_6 = r_6 - r_5$) and outer monolayers ($\sigma_{\mathrm{term}} = r_3 - r_2 = r_2 - r_1$). Hence, on the basis of these considerations, we define $\varphi_{\mathrm{h,in}}(r)$ the hydrophobic volume fraction distribution of the inner monolayer obtained through the function $f(r)$ with $r_0 = -D_{\mathrm{h,in}}/2$, $w = D_{\mathrm{h,in}}/2$, $\xi_l = \xi_{\mathrm{h,in}}$ and $\xi_r = \sigma_{\mathrm{term}}/(D_{\mathrm{h,in}}/2))$. Likewise, the hydrophobic volume fraction distribution of the outer monolayer, $\varphi_{\mathrm{h,out}}(r)$, is defined through the function $f(r)$ with $r_0 = D_{\mathrm{h,out}}/2$, $w = D_{\mathrm{h,out}}/2$, $\xi_l = \sigma_{\mathrm{term}}/(D_{\mathrm{h,out}}/2)$ and $\xi_r = \xi_{\mathrm{h,out}}$.

The integral over the hydrophobic domain volume fraction radial distribution should be the volume $V_{\mathrm{h},\alpha}$ occupied by the hydrophobic components of molecules in the α-monolayer. Thus we derive

$$\begin{aligned} V_{\mathrm{h,in}} &= 4\pi \int_{R-D_{\mathrm{h,in}}-\xi_{\mathrm{h,in}}D_{\mathrm{h,in}}/2}^{R+\sigma_{\mathrm{term}}} \varphi_{\mathrm{h,in}}(r) r^2 dr \\ &= \frac{\pi}{6}(8D_{\mathrm{h,in}}^3 + D_{\mathrm{h,in}}^3\xi_{\mathrm{h,in}}^2 - 24D_{\mathrm{h,in}}^2 R - D_{\mathrm{h,in}}^2\xi_{\mathrm{h,in}}^2 R + 4\sigma_{\mathrm{term}}^2 R + 24D_{\mathrm{h,in}}R^2) \\ &= N_{\mathrm{sa,in}} \sum_{g=1}^{N_{\mathrm{h,in}}} \mathrm{n}_{g,\mathrm{in}}\nu_{g,\mathrm{in}} = N_{\mathrm{sa,in}}\nu_{\mathrm{h,in}}, \end{aligned} \tag{5}$$

and

$$\begin{aligned} V_{\mathrm{h,out}} &= 4\pi \int_{R-\sigma_{\mathrm{term}}}^{R+D_{\mathrm{h,out}}+\xi_{\mathrm{h,out}}D_{\mathrm{h,out}}} \varphi_{\mathrm{h,out}}(r) r^2 dr \\ &= \frac{\pi}{6}(8D_{\mathrm{h,out}}^3 + D_{\mathrm{h,out}}^3\xi_{\mathrm{h,out}}^2 + 24D_{\mathrm{h,out}}^2 R + D_{\mathrm{h,out}}^2\xi_{\mathrm{h,out}}^2 R - 4\sigma_{\mathrm{term}}^2 R + 24D_{\mathrm{h,out}}R^2) \\ &= N_{\mathrm{sa,out}} \sum_{g=1}^{N_{\mathrm{h,out}}} \mathrm{n}_{g,\mathrm{out}}\nu_{g,\mathrm{out}} = N_{\mathrm{sa,out}}\nu_{\mathrm{h,out}}, \end{aligned} \tag{6}$$

where we have introduced the volume $\nu_{\mathrm{h},\alpha} = \sum_{g=1}^{N_{\mathrm{h},\alpha}} \mathrm{n}_{g,\alpha}\nu_{g,\alpha}$ occupied by all the hydrophobic groups of the molecular unit of the α-monolayer. To note, Eqs. 5-6 can be used to calculate the self-assembling number $N_{\mathrm{sa},\alpha}$ (which we have seen enters in Eq. 4) as a function of the peak parameters defining $\varphi_{\mathrm{h},\alpha}(r)$. Considering the Fourier transform and the need to deal

with the polydispersion on R, as we will discuss in the next paragraph, it is useful to develop in power series of $1/R$ the normalization factor introduced in Eq. 4, according to

$$\frac{N_{\mathrm{sa},\alpha}\mathrm{n}_{g,\alpha}\nu_{g,\alpha}}{V_f} = \sum_{k=0}^{K}\frac{a_{k,g,\alpha}}{R^k} \tag{7}$$

where K is the maximum value of the power (we have used $K = 3$) and the expansion coefficients $a_{k,g,\alpha}$ will depends on all the parameters, apart R, that define the hydrophobic volume fraction distribution of the α-monolayer, $\varphi_{\mathrm{h},\alpha}(r)$, and the g-group volume fraction distribution, $\varphi_{g,\alpha}(r)$. The analytic values of $a_{k,g,\alpha}$, that have been obtained on the basis of Eqs. 3, 5 and 6, are reported in supporting information (SI), Eq. S1. The volume fraction radial distribution of the dominant hydrophobic group ($g = 1$, typically CH_2) is obtained by subtracting from the overall volume fraction of the hydrophobic domain the volume fractions of all the other hydrophobic groups,[4,23–25]

$$\varphi_{1,\alpha}(r) = \varphi_{\mathrm{h},\alpha}(r) - \sum_{g=2}^{N_{\mathrm{h},\alpha}}\varphi_{g,\alpha}(r). \tag{8}$$

To avoid nonphysical results, it should be ensured that this subtraction is never negative.

We turn now to describe the volume fraction radial distribution of hydration molecules in the inner and in the outer monolayer, which we have indexed with $g = N_{\mathrm{h},\alpha} + N_{\mathrm{p},\alpha} + 1$. We firstly observe that the minimum radius of the inner region is $R_{\mathrm{in}} = R - D_{\mathrm{in}}$, where $D_{\mathrm{in}} = -\min\{r_{g,\mathrm{in}} - w_{g,\mathrm{in}}(1+\xi_{g,\mathrm{in}})\}$, with $g = 1, N_{\mathrm{h,in}} + N_{\mathrm{p,in}}$, defined as the thickness of the hydrated inner monolayer, whereas the maximum radius of the solvated outer monolayer is $R_{\mathrm{out}} = R + D_{\mathrm{out}}$, where $D_{\mathrm{out}} = \max\{r_{g,\mathrm{out}} + w_{g,\mathrm{out}}(1+\xi_{g,\mathrm{out}})\}$, with $g = 1, N_{\mathrm{h,out}} + N_{\mathrm{p,out}}$, is the thickness of the hydrated outer monolayer. Secondly, we observe that, since hydration molecules cannot enter in the hydrophobic domains, their volume fraction radial distribution can be different from 0 only in the range $0 \le r \le R$, for the inner monolayer, and only in the range $R \le r \le R_{\mathrm{out}}$, for the outer monolayer. Third, since hydration molecules cannot enter

in the junction between the two monolayers, for calculating the hydration volume fraction due to, respectively, the inner monolayer in the range $0 \leq r \leq R$ and to the outer monolayer in the range $R \leq r \leq R_{\text{out}}$, we need to set $\sigma_{\text{term}} = 0$. In these conditions, the hydration volume fraction radial distribution can be expressed as the complement to 1 of all the other group volume fraction radial distributions, according to

$$\varphi_{N_{\text{h,in}}+N_{\text{p,in}}+1,\text{in}}(r) = \text{rect}\left(\frac{r-R}{R}+\frac{1}{2}\right) - \bar{\varphi}_{\text{h,in}}(r) - \sum_{g=N_{\text{h,in}}+1}^{N_{\text{h,in}}+N_{\text{p,in}}} \varphi_{g,\text{in}}(r), \tag{9}$$

and

$$\varphi_{N_{\text{h,out}}+N_{\text{p,out}}+1,\text{out}}(r) = \text{rect}\left(\frac{r-R}{D_{\text{out}}}-\frac{1}{2}\right) - \bar{\varphi}_{\text{h,out}}(r) - \sum_{g=N_{\text{h,out}}+1}^{N_{\text{h,out}}+N_{\text{p,out}}} \varphi_{g,\text{out}}(r), \tag{10}$$

In these equations $\text{rect}(x)$ is the rectangle function ($\text{rect}(x) = 1$ if $|x| \leq 1/2$ and 0 otherwise): it ensures that the complement to 1 is calculated only for the proper values of r. Notice that $\bar{\varphi}_{\text{h},\alpha}(r)$ is the hydration volume fraction of the α-monolayer calculated with $\sigma_{\text{term}} = 0$.

The radial excess SLD profile with respect to the bulk SLD, ρ_0, is

$$\varrho(r) - \rho_0 = \sum_{\alpha=\text{in,out}} \sum_{g=1}^{N_{\text{h},\alpha}+N_{\text{p},\alpha}+1} (\rho_{g,\alpha} - \rho_0)\varphi_{g,\alpha}(r) \tag{11}$$

This equation can be transformed according to Eqs. 8, 9 and 10, as shown in SI, Eq. S3.

The vesicle form factor is the so-called isotropic Fourier transform o the excess SLD, according to

$$F(q) = 4\pi \int_0^\infty [\varrho(r) - \rho_0] \frac{\sin(qr)}{qr} r^2 dr \tag{12}$$

On the basis of Eq. 11, $F(q)$ will be a linear combination of the Fourier transforms of the volume fraction distributions $\varphi_{g,\alpha}(r)$, which, as previously discussed, are both expressed by the peak $f(r)$ defined in Eq. 2. As shown in Eq. S5 of the SI, the parabolic nature of $f(r)$

allows the easy derivation of the analytic solution of its Fourier transform, $\tilde{f}(q)$.

As shown in detail in the SI, we have been able to express both $F(q)$ and its square in a compact form, prone to be averaged under the polydispersion distribution function $p(R)$ of the vesicle radius R. Indeed, we have obtained Eqs. S15 and S20 here reported,

$$F(q) = \frac{4\pi}{q} \sum_{k=-1}^{K} \Im \left\{ e^{\imath qR} \frac{F_k(q)}{R^k} \right\} \tag{13}$$

$$F^2(q) = \frac{8\pi^2}{q^2} \sum_{k_1,k_2=-1}^{K} \Re \left\{ \frac{F_{k_1}(q) F^*_{k_2}(q)}{R^{k_1+k_2}} - e^{\imath 2qR} \frac{F_{k_1}(q) F_{k_2}(q)}{R^{k_1+k_2}} \right\} \tag{14}$$

where all the features of the volume distribution functions and SLDs are contained in the analytic functions $F_k(q)$, reported in Eqs. S16-S19.

According to Pencer et al.[30], we have chosen the Schulz distribution function as a model for the $p(R)$,

$$p(R) = \frac{s^{sR_0}}{\Gamma(sR_0)} R^{sR_0-1} e^{-sR} \tag{15}$$

being $\Gamma(x)$ the so-called Gamma function. This distribution is characterized by two parameters, the average radius R_0,

$$R_0 = < R > = \int_0^\infty R\, p(R)\, dR, \tag{16}$$

and $s = \frac{1}{R_0 \xi_R^2}$, where ξ_R is the dispersion of R,

$$\xi_R^2 = (< R^2 > - < R >^2) / < R >^2 \, . \tag{17}$$

The expressions that we have derived for the form factor and its square (Eqs. 13 and 14), which show a simple dependency on R due to the power expansions of the normalization factors (Eq. 7), allow to derive analytic equations for their polydispersion averages, $< F(q) >$ and $< F^2(q) >$ over $p(R)$. They are fully reported in the SI, Eqs. S26 and S27. We wish to

emphasize that this feature allows to apply this model in a rapid and efficient data fitting procedure.

The macroscopic differential scattering cross section (SCS) provided by a SAXS or a SANS experiment of polydispersed vesicles can be written in the following form

$$\frac{d\Sigma}{d\Omega}(q) \quad = \quad N_A C_{\text{ves}} < F^2(q) > S_M(q) + B, \tag{18}$$

where N_A is Avogadro's number, C_{ves} is the molar concentration of polydispersed vesicles (calculated as shown in the SI, Eq. S32) and B is a flat residual background mostly due to the inchoerent scattering contribution of the non acidic protons of the sample. It's worth remarking that the decoupling approximation is poor for systems with high polydispersities such the proteoliposomes used in this study, and the local monodisperse approximation used by Pedersen[31] might work better. However, in view of the low concentration of the proteoliposomes in solution this does not acctually make a substantial difference. The term $S_M(q)$ in Eq. 18 is called the "effective" structure factor: it takes into account all the interference effects coming from distinct vesicles. In the case of a polydispersed system, it can be approximated by the following equation

$$S_M(q) = 1 + \frac{< F(q) >^2}{< F^2(q) >}[S(q) - 1] \tag{19}$$

where $S(q)$ is the vesicle-vesicle structure factor, which can be described by different approximations as discussed in Spinozzi et al.[32]. In this case the $S(q)$ was derived by the perturbation of the Perkus-Yevick structure factor due to the presence of a double Yukawian potential within the random phase approximation.[32]

In the case of SANS, we have also to consider the non negligible experimental uncertainty on q. To deal with this effect, we have to convolute the model SCS with the instrumental

resolution function $R_{\mathrm{av}}(q', q)$, according to

$$\left\langle \frac{d\Sigma}{d\Omega}(q) \right\rangle = \int_{q-3\sigma_q}^{q+3\sigma_q} \frac{d\Sigma}{d\Omega}(q')\, R_{\mathrm{av}}(q', q)\, dq', \tag{20}$$

where σ_q is the standard deviation of the q, an information that should be experimentally available. Analytic approximation of $R_{\mathrm{av}}(q', q)$ are available in literature.[33]

SANS experiments

SANS measurements were performed on KWS-2 SANS diffractometer operated by Jülich Centre for Neutron Science at the neutron source Heinz Maier–Leibnitz (FRM II reactor) in Garching, Germany.[34] The incident neutron beam at KWS-2 was monochromatized with a velocity selector to have an average wavelength λ with a wavelength distribution of $\delta\lambda/\lambda = 20\%$. The scattering patterns were collected using a two-dimensional (2D) scintillation detector which was placed at 2 m, 8m and 20m after the sample. The measurements at 2m and 8m detection distance were carried out with a wavelenegth $\lambda = 5$ Å while for the sample-to-detector distance of 20m a wavelength $\lambda = 10$ Å was used. The scattering patterns were corrected for the background of the sample quartz cuvette, the electronic noise of the detector and the detector sensitivity and circularly averaged to obtain scattering intensity profiles as a function of Q, where Q is the scattering vector and defined as $Q = (4\pi/\lambda)\sin(\theta/2)$, with θ the scattering angle. The final intensity profiles obtained were corrected for the solvent buffer contribution. All measurements were performed at 25° C.

Results and Discussion

SANS curves recorded at KWS2 for samples of liposomes (L) and proteoliposomes (PL) in different conditions of the solvent deuteration grades (x_D) are shown in Figure 2. All curves have been analysed with the model introduced in the section SAS Model, which has been included in the GENFIT software.[32]

Since, the composition of the mixture DOPC:DOPE:CHOL:DMPA constituting the liposomes (L samples) is not exactly known, and considering that it could change in the presence of OprF (PL samples), we have introduced two sets of molar ratios $x_{\mathrm{S,DOPC}}$:$x_{\mathrm{S,DOPE}}$:$x_{\mathrm{S,CHOL}}$:$x_{\mathrm{S,DMPA}}$ (where S stands for L or PL samples) and leaving each value of $x_{\mathrm{S},k}$ ($k = 1, 4$) to vary within a range of $\pm 20\%$ around the nominal value ($x_{\mathrm{S,DOPC}} = 2$, $x_{\mathrm{S,DOPE}} = 1$, $x_{\mathrm{S,CHOL}} = 1$, $x_{\mathrm{S,DMPA}} = 1$). Accordingly, the average molecular weight of all the lipid molecules constituting the molecular unit is $< M_w >_{\mathrm{S}} = \sum_{k=1}^{4} y_{\mathrm{S},k} M_{w,k}$, $y_{\mathrm{S},k}$ being the lipid normalized ratio, $y_{\mathrm{S},k} = x_{\mathrm{S},k} / \sum_{k'=1}^{4} x_{\mathrm{S},k'}$. The total w/v lipid concentrations for the two series of samples, c_{S}, are also fitting parameters that cannot be greater than the nominal values ($c_{\mathrm{L}} \leq 5$ g/L and $c_{\mathrm{PL}} \leq 1.2$ g/L). Conversely, the w/v protein concentration in the PL samples, c_{OprF}, is refined in a narrow range $0.4 - 0.6$ g/L around the nominal value. To note, for PL samples, the number of protein molecules in each molecular unit is $y_{\mathrm{PL,OprF}} = c_{\mathrm{OprF}} / M_{w,\mathrm{OprF}} / (c_{\mathrm{PL}} / < M_w >_{\mathrm{PL}})$.

The list of the j-chemical groups constituting all the k-molecules in the samples is reported in Table 1, together with their nominal molecular volume (ν_j), molecular weight ($M_{w,j}$), scattering length calculated in pure light water (where all exchangeable hydrogens atoms are protons, $b_{\mathrm{H},j}$) and in pure heavy water (with all exchangeable hydrogens are substituted with deuterons, $b_{\mathrm{D},j}$) and their abundance in each molecule ($\mathrm{m}_{j,k}$).

In order to simplify the description of the system and to reduce the number of fitting parameters, for L samples the molecular units of both inner and outer monolayer are defined by only one united hydrophobic group, which includes all the lipid hydrophobic j-chemical groups (first block of Table 1), and one polar united group, referring to the polar j-chemical groups (second block of Table 1). For PL samples we add a second hydrophobic united group for each monolayer, which is half the OprF barrel ($\mathrm{OprF_{HB}}$), and a second polar united group. The scattering length of the united groups belonging to the OprF depends on the isotopic content of the aqueous solution due to the presence of labile hydrogens that can exchange with the solution. Accordingly, the scattering length of the g-united group, in the presence of a solvent with deuteration grade x_D, is calculated by $b_{g,\alpha} = x_D b_{\mathrm{D},g,\alpha} + (1 - x_D) b_{\mathrm{H},g,\alpha}$,

Table 1: List of all chemical groups and their molecular volumes, molecular weights, number of exchangeable hydrogens, scattering length in H_2O, scattering length in D_2O, and abundance in the four lipid molecules investigated and in OprF protein. [(a)] data calculated according to Marsh [35]. Neutron scattering length and molecular volumes of protein groups have been calculated according to Jacrot [36]. The first and the last block of the table refer to hydrophobic and polar groups, respectively.

						DOPC	DOPE	CHOL	DMPA	OprF
j-group	$\nu_j^{(a)}$	$M_{w,j}$	$n_{H,exg}$	$b_{H,j}$	$b_{D,j}$			$m_{j,k}$		
	(Å^3)	(g/mol)		(10^{-12} cm)		$k=1$	$k=2$	$k=3$	$k=4$	$k=5$
Hydrophobic chemical groups										
CH_2	27.7	14.03	0	−0.083	−0.083	28	28		24	
CH_3	52.9	15.04	0	−0.458	−0.458	2	2		2	
CH	21.5	13.02	0	0.291	0.291	4	4			
$CHOL_h$	608.3	369.66	0	1.119	1.119			1		
$OprF_{HB}$	12662.2	10004.95	159	238.321	404.148					2
Polar chemical groups										
PCN	87.8	137.03	0	3.612	3.612	1				
CG	143.3	129.09	0	3.776	3.776	1	1		1	
$ColCH_3$	99.9	45.11	0	−1.373	−1.373	1				
OH	14.3	17.01	1	0.206	1.248			1		
$PCNH_3$	103.3	140.06	3	2.489	5.613		1			
P	31.9	94.97	1	2.839	3.880				1	
$OprF_C$	21631.4	17595.69	290	418.199	720.697					1
$OprF_N$	3187.8	2575.34	44	68.751	114.048					1
Water										
H_2O	29.9	18.02	2	−0.168	1.915					

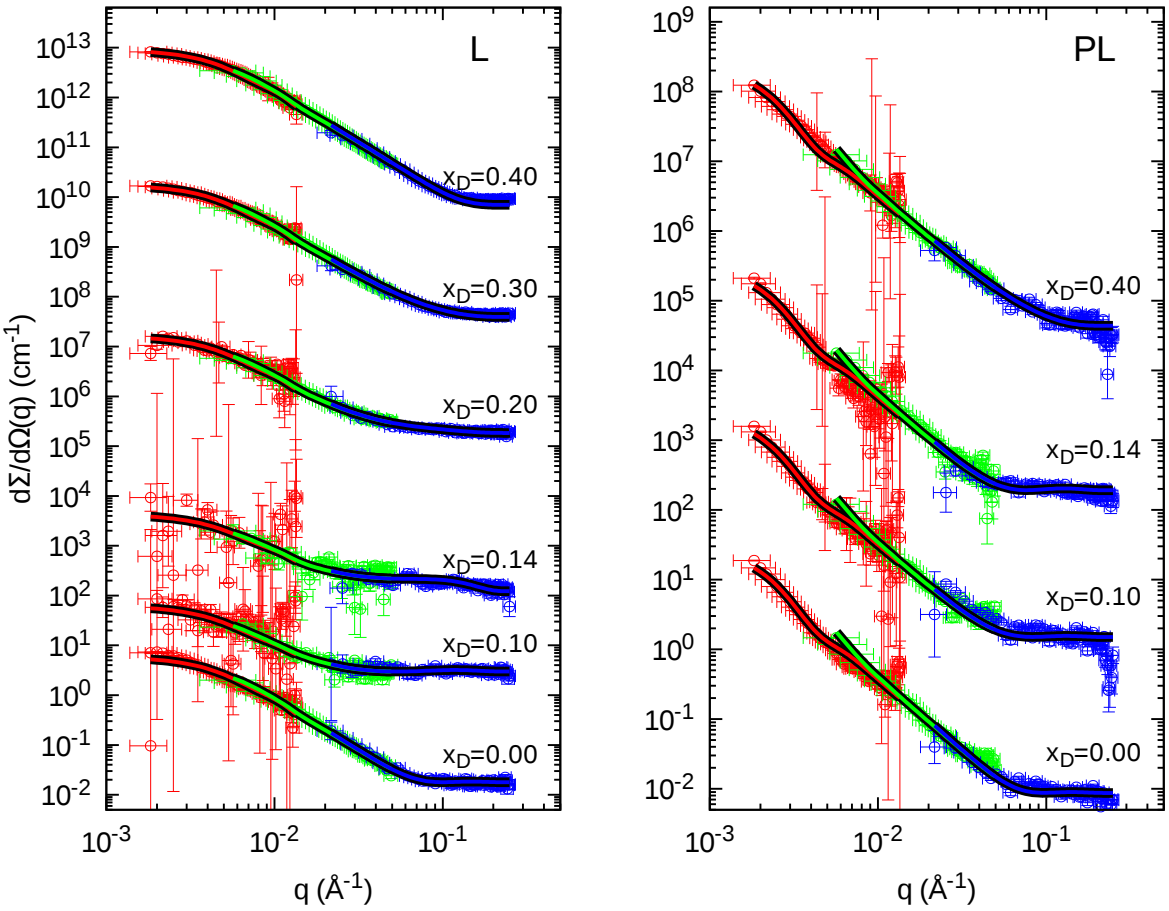

Figure 2: SANS experimental data of liposomes (L, left panel) and proteoliposomes (PL, right panel) at different solvent deuteration grades (x_D). Red, green and blue points refer to a sample-to-detector distance of 20, 8 and 2 m, respectively. Solid black/colored lines are the best fits obtained by GENFIT software.[32] Subsequent curves are multiplied by a factor 50.0 for clarity.

with $b_{\mathrm{I},g,\alpha} = \sum_{k=1}^{5} y_{\mathrm{S},k} \sum_j \delta_{j,g} \mathrm{m}_{j,k} b_{\mathrm{I},j}$ with $\mathrm{I} = \mathrm{D}, \mathrm{H}$, $\delta_{j,g}$ being the Kronecker δ function which is 1 only if the united group g includes the chemical group j and 0 otherwise (Table 1). The molecular volume of the united group g is $\nu_{g,\alpha} = \sum_{k=1}^{5} y_{\mathrm{S},k} \sum_j \delta_{j,g} \mathrm{m}_{j,k} \nu_j$. On this basis, for L samples we have $N_{\mathrm{h},\alpha} = 1$ and $N_{\mathrm{p},\alpha} = 1$, as well as $\mathrm{n}_{1,\alpha} = 1$ and $\mathrm{n}_{2,\alpha} = 1$ for both $\alpha = \mathrm{in}$ and $\alpha = \mathrm{out}$. On the other hand, for PL samples, we have $N_{\mathrm{h},\alpha} = 2$ and $N_{\mathrm{p},\alpha} = 2$, $\mathrm{n}_{1,\alpha} = 1$, $\mathrm{n}_{2,\alpha} = y_{\mathrm{PL,OprF}}$, $\mathrm{n}_{3,\alpha} = 1$, $\mathrm{n}_{4,\alpha} = y_{\mathrm{PL,OprF}}$, again for both $\alpha = \mathrm{in}$ and $\alpha = \mathrm{out}$. We have to pay attention to the $4, \alpha$-united groups, which represent the polar protein domains. Indeed, it is not certain if the OprF C-terminal domain points toward the

inner or the outer vesicle layers. Thus, we introduce an adjustable parameter, $x_{\mathrm{C},\in}$, which indicates the fraction of the OprF proteins with the C-terminal domain pointing toward the inner vesicle. Hence, the volume and the scattering length of the second inner polar group are $\nu_{4,\mathrm{in}} = x_{\mathrm{C},\in}\nu_{\mathrm{OprF_C}} + (1 - x_{\mathrm{C},\in})\nu_{\mathrm{OprF_N}}$ and $b_{\mathrm{I},4,\mathrm{in}} = x_{\mathrm{C},\in}b_{\mathrm{I},\mathrm{OprF_C}} + (1 - x_{\mathrm{C},\in})b_{\mathrm{I},\mathrm{OprF_N}}$ (with $\mathrm{I} = \mathrm{D}, \mathrm{H}$). Conversely, the volume and the scattering length of the second outer polar group are $\nu_{4,\mathrm{out}} = (1 - x_{\mathrm{C},\in})\nu_{\mathrm{OprF_C}} + x_{\mathrm{C},\in}\nu_{\mathrm{OprF_N}}$ and $b_{\mathrm{I},4,\mathrm{out}} = (1 - x_{\mathrm{C},\in})b_{\mathrm{I},\mathrm{OprF_C}} + x_{\mathrm{C},\in}b_{\mathrm{I},\mathrm{OprF_N}}$.

The set of all the $N_{\mathrm{C}} = 30$ experimental SANS curves shown in Figure 2 (3 curves for each samples, recorded at three different sample-to-detector distances) have been simultaneously co-refined by means of Eq. 20 with a unique optimum fit. The merit function that has been minimized is the overall reduced chi-square,

$$\chi^2 = \frac{1}{N_{\mathrm{C}}} \sum_{\mathrm{c}=1}^{N_{\mathrm{C}}} \frac{1}{N_{q,\mathrm{C}}} \sum_{k=1}^{N_{q,\mathrm{C}}} \left(\frac{\left\langle \frac{d\Sigma}{d\Omega}(q_k) \right\rangle_{\mathrm{c}} - \frac{d\Sigma}{d\Omega}_{\mathrm{c,exp}}(q_k)}{\sigma_{\mathrm{c,exp}}(q_k)} \right)^2 \tag{21}$$

where $\left\langle \frac{d\Sigma}{d\Omega}(q_k) \right\rangle_{\mathrm{c}}$, $\frac{d\Sigma}{d\Omega}_{\mathrm{c,exp}}(q_k)$ and $\sigma_{\mathrm{c,exp}}(q_k)$ are theoretical, experimental and standard deviation values of the c-th SANS curve measured at the k-th of the $N_{q,\mathrm{C}}$ values of q.

Fitting parameters have been classified in three classes. In the first class we have parameters that should have a unique value for all the 30 SANS curves. They are: the volume of the methylene group, $\nu_{\mathrm{CH_2}}$; the ratios $x_{12} = \nu_{\mathrm{CH_3}}/\nu_{\mathrm{CH_2}}$ and $x_{13} = \nu_{\mathrm{CH}}/\nu_{\mathrm{CH_2}}$; the volumes of the lipid polar chemical groups, $\nu_{\mathrm{CHOL_h}}$, ν_{PCN}, ν_{CG}, ν_{OH}, $\nu_{\mathrm{PCNH_3}}$, and ν_{P}; the relative mass density of the hydration water, d_{hw}.

In the second class there are fitting parameters shared by the all the curves within each of the two series of samples, S = L and S = PL. This class contains the largest number of parameters. They are: the composition of the molecular unit, $x_{\mathrm{S},k}$, with $k = 1, 4$; the total w/v lipid concentrations, c_{S}; the protein concentration, c_{OprF} (only for S = PL samples); the average radius of the vesicles, R_0, and its dispersion, ξ_R; the smoothness length at the monolayers' junction, σ_{term}; the thickness of the hydrophobic domain of each monolayers, $D_{\mathrm{h},\alpha}$; a unique hydrophobic domain smoothness parameter, $\xi_{\mathrm{h}} = \xi_{\mathrm{h,in}} = \xi_{\mathrm{h,out}}$, a choice we

made in order to reduce at the minimum the number of fitting parameters; the wideness parameters of the lipid polar group of each monolayers, $w_{\mathrm{p},\alpha}$ (corresponding to $w_{N_{\mathrm{h},\alpha}+1,\alpha}$, according to the label assignment rules); a unique lipid polar group smoothness parameter, $\xi_{\mathrm{p}} = \xi_{N_{\mathrm{h},\alpha}+1,\mathrm{in}} = \xi_{N_{\mathrm{h},\alpha}+1,\mathrm{out}}$; a unique protein hydrophobic group smoothness parameter, $\xi_{\mathrm{OprF,h}} = \xi_{N_{\mathrm{h},\alpha},\mathrm{in}} = \xi_{N_{\mathrm{h},\alpha},\mathrm{out}}$ (only for S = PL samples); the wideness parameters of the protein polar group of each monolayers, $w_{\mathrm{OprF,p},\alpha}$ (only for S = PL samples, corresponding to $w_{N_{\mathrm{h},\alpha}+N_{\mathrm{p},\alpha},\alpha}$); the protein polar group smoothness parameters of each monolayers, $\xi_{\mathrm{OprF,p},\alpha}$ (only for S = PL samples, corresponding to $\xi_{N_{\mathrm{h},\alpha}+N_{\mathrm{p},\alpha},\alpha}$); the structure factor parameters J, d, Z as well as the ionic strength I_S.[32]

The third and last class of parameters include those parameters that are shared by samples that only differ for the sample-to-detector distance. They are: the refined values of the deuteration grade, x_D, that could be slightly different form the nominal value, due to the contribution of the exchangeable protons coming from all the compounds except light water; the flat background B due to the incoherent scattering contribution to the SANS curves.

It is worth noting that there are no fitting parameters that refer to only one of the 30 experimental curves and that the values of the parameters were restricted to physically relevant intervals.

It is also worth to underline that, in order to set at the minimum possible value the number of the fitting parameters, and also to control the continuity of the protein domains, for PL samples we have assumed, from one hand, that the wideness parameter of the two protein hydrophobic groups ($\mathrm{OprF_{HB}}$) should be half the thickness of the hydrophobic domain, $w_{N_{\mathrm{h},\alpha},\alpha} = D_{\mathrm{h},\alpha}/2$, and, from the other hand, that the positions of these two groups should be $r_{N_{\mathrm{h,in}},\mathrm{in}} = -D_{\mathrm{h,in}}/2$ and $r_{N_{\mathrm{h,out}},\mathrm{out}} = D_{\mathrm{h,out}}/2$. This choice ensures that the hydrophobic protein groups have the same thickness of the hydrophobic domain of the molecular units. Moreover, the positions of the protein polar domains have been written as $r_{N_{\mathrm{h,in}}+N_{\mathrm{p,in}},\mathrm{in}} = -D_{\mathrm{h,in}} - w_{\mathrm{OprF,p,in}} - l_{\mathrm{OprF,p,in}}$ and $r_{N_{\mathrm{h,out}}+N_{\mathrm{p,out}},\mathrm{out}} = D_{\mathrm{h,out}} + w_{\mathrm{OprF,p,out}} + l_{\mathrm{OprF,p,out}}$, where $l_{\mathrm{OprF,p},\alpha}$ are taken as positive second class fitting parameters.

Best fitting curves are reported in Figure 2 with the same colour of the experimental points. The quality of the fit, within the experimental uncertainty, is, in all cases, high. Looking at the three fitting curves related to the three sample-to-detector distances, we notice that they do not always overlap each other: this effect is due to the very different q resolution of at the three configurations, which, according to Eq. 20, differently affects the unique model curve $\frac{d\Sigma}{d\Omega}(q)$ behind the three curves.

Fitting parameters of the first and the third class are reported in Table 2. The ones less relevant, belonging to the third class, are shown in the SI, Table S1. Other significant parameters, derived by the fitting ones, are shown in Table 3. The molecular volumes of the different chemical groups was allowed to slightly vary in the fit with respect to the their nominal values designated in Table 1, in order to account from the differences on molecular packing. Their values obtained from the fit are in general consistent within the errors of the nominal values.

The lipid w/v concentration for the L series of samples is 4.6 ± 0.2 g/L, quite close to the nominal value. For PL samples, the lipid and protein w/v concentration found by the SANS analysis are 0.56 ± 0.07 and 0.46 ± 0.07 g/L, respectively. The estimate of the lipid component in a mixed lipid protein system such the proteoliposome is a quantity subject to a significant uncertainty in the preparation phase due to the measurement uncertainties at each step, in particular in the filtration step. Our SANS analysis could provide a more precise estimation of this number thanks to the significant difference in the scattering length densities of the protein and of the lipid component and to how these densities are differently affected by the isotopic substitution of the water molecules in the different contrasts used in the experiments.

To note, for the two series of samples, L and PL, the molar concentration of vesicles has resulted to be very low (390 ± 30 nM and 6.0 ± 0.1 nM, respectively, see Table S1), so that the structure factor $S_M(q)$ is almost 1 and does not modify the form factor. The vesicle average radius R_0 is quite different for the two series of samples (200 ± 10 Å and 586 ± 7 Å

Table 2: First and second class of fitting parameters obtained by the analysis of SANS curves. The unit of length is Å. Validity ranges of fitting parameters: [a] $26.5 - 28.1$; [b] $1.92 - 1.98$; [c] $0.76 - 0.84$; [d] $626.1 - 590.5$; [e] $90.4 - 85.2$; [f] $147.5 - 139.1$; [g] $14.7 - 13.9$; [h] $106.3 - 100.2$; [i] $32.8 - 30.9$; [j] $1.00 - 1.10$; [k] $1.6 - 2.4$; [l] $0.8 - 1.2$; [m] $3.0 - 5.0$ g/L; [n] $0.5 - 1.2$ g/L; [o] $0.4 - 0.6$ g/L; [p] $100 - 2000$; [q] $0.01 - 3$; [r] $1 - 5$; [s] $12 - 17$; [t] $0.1 - 1$; [u] $4 - 7$; [v] $0 - 1$; [w] $4 - 25$; [x] $0 - 20$.

First class fitting parameters			
ν_{CH_2}	[a]	28.1±0.2	
x_{12}	[b]	1.92±0.03	
x_{13}	[c]	0.83±0.02	
ν_{CHOL_h}	[d]	620±8	
ν_{PCN}	[e]	88±2	
ν_{CG}	[f]	145.8±0.8	
ν_{OH}	[g]	14.2±0.2	
ν_{PCNH_3}	[h]	103.8±0.7	
ν_{P}	[i]	31.6±0.3	
d_{hw}	[j]	1.00±0.01	
Second class fitting parameters			
		L	PL
$x_{S,DOPC}$	[k]	2.1±0.2	2.40±0.04
$x_{S,DOPE}$	[l]	0.80±0.07	0.89±0.08
$x_{S,CHOL}$	[l]	1.1±0.1	0.8±0.2
$x_{S,DMPA}$	[l]	0.8±0.1	0.8±0.2
c_S	[m]	4.6±0.2	0.56±0.07 [n]
c_{OprF}			0.46±0.07 [o]
R_0	[p]	200±10	586±7
ξ_R	[q]	0.36±0.02	0.33±0.01
σ_{term}	[r]	1.0±0.1	4.6±0.6
$D_{h,\in}$	[s]	13.9±0.4	14.8±0.1
$D_{h,out}$	[s]	14.8±0.4	15.5±0.2
ξ_h	[t]	0.999±0.002	1.0±0.1
$w_{p,\in}$	[u]	4.6±0.6	5.5±0.3
$w_{p,out}$	[u]	7.0±0.8	6.3±0.8
ξ_p	[t]	0.5±0.4	0.1±0.4
$x_{C,\in}$			0.5±0.2 [v]
$\xi_{OprF,h}$			0.1±0.4 [t]
$w_{OprF,p,\in}$			6±2 [w]
$w_{OprF,p,out}$			6.4±0.6 [w]
$\xi_{OprF,p,\in}$			0.9±0.5 [t]
$\xi_{OprF,p,out}$			0.7±0.4 [t]
$l_{OprF,p,\in}$			0±1 [x]
$l_{OprF,p,out}$			0±7 [x]

Table 3: Derived parameters obtained by the analysis of SANS curves. The unit of length and surface area are Å and Å^2, respectively. [a] nM unit.

		L	PL
$a_{\in}$		51±1	72±6
a_{out}		66±1	74±6
$D_{\in}$		27.7±0.8	33.8±0.3
D_{out}		32.6±0.8	32.8±0.3
$n_{\text{wat},\in}$		17±2	29±2
$n_{\text{wat,out}}$		31±4	27±3
C_{ves}	[a]	390±30	6.0±0.1
$a_{\text{OprF},\in}$			4800±400
$a_{\text{OprF,out}}$			5400±900
$d_{\text{OprF},\in}$			96±2
$d_{\text{OprF,out}}$			101±4

for L and PL, respectively), as already suggested by DLS. Distribution functions, reported in Figure 3, show the effect of the quite high level of dispersion ξ_R (0.36 ± 0.02 and 0.33 ± 0.01 for L and PL, respectively). The radius R_0 of the PL obtained from the analysis of SANS is significantly smaller than that obtained from the DLS experiment. This is probably due to the presence of some larger proteoliposomes that, even in small number, can mask the DLS signal of the smaller ones. However, that the number of these larger proteoliposomes is low, can be seen from the SANS data at low Q. Larger proteoliposomes in higher number would be seen by SANS like vesicles with infinite radius, giving a Q^{-2} trend at low Q that is not present in our data. In terms of PDI, the SANS analysis returned a PDI of 0.11, which is smaller compared to that obtained with DLS.

The united group volume fraction distributions, obtained with the set of fitting parameters, are reported in Figure 4, left and right panel for L and PL samples, respectively. Regarding liposomes, the results show an high level of symmetry of the bilayer, with similar values of the inner and the outer hydrophobic thickness ($D_{\text{h,in}} = 13.9 \pm 0.4$ Å and $D_{\text{h,out}} = 14.8 \pm 0.4$ Å) and polar wideness ($w_{\text{p,in}} = 4.6 \pm 0.6$ Å and $w_{\text{p,out}} = 7.0 \pm 0.8$ Å). The different heights of the polar group (blue curves) are the effect of the difference in curvature between the two monolayers, which is not negligible being the vesicle radius R_0 of a

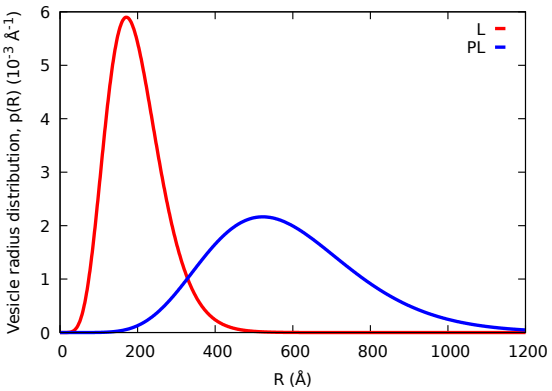

Figure 3: Vesicle radius distributions derived from the analysis of SANS data for liposomes (L, red line) and proteoliposomes (PL, blue line). The vescicle radius distribution is defined in equation S21.

size comparable with that of the overall bilayer thickness. Indeed, this difference is clearly reflected by the areas of molecular unit, which have been calculated with Eqs. S35-S36 and reported in Table 3, which have resulted $a_{\text{in}} = 51 \pm 1$ Å^2 and $a_{\text{out}} = 66 \pm 1$ Å^2. Consequently, also the number of water molecules in the outer polar regions has turned out to be higher than the one in the inner counterpart ($n_{\text{wat,in}} = 17 \pm 2$ and $n_{\text{wat,out}} = 31 \pm 4$).

Regarding PL samples, we first notice that the fraction of the OprF proteins with the C-terminal domain pointing toward the inner vesicle is $x_{\text{C},\in} = 0.5 \pm 0.2$, suggesting that, in the limit of the SANS techniques, protein molecules have not a preferential orientation within the bilayers. Indeed, Figure 4, right-PL panel, shows an asymmetric behaviour quite similar to the one of samples without proteins. The tiny asymmetry of the bilayer is well reflected by the average area of the molecular unit, which includes also protein: they are $a_{\text{in}} = 72 \pm 6$ Å^2 and $a_{\text{out}} = 74 \pm 6$ Å^2. By the obtained value of the number of protein molecules in the molecular unit (indicated with $y_{\text{PL,OprF}}$) we have calculated both the average area of the monolayer polar surface that can be assigned at each protein molecule, $a_{\text{OprF},\alpha} = a_\alpha / y_{\text{PL,OprF}}$, and, assuming that this area is a circle, the circle's diameter, $d_{\text{OprF},\alpha}$, which may represent an estimation of the average protein-protein distance along the α-monolayer surface. Results, reported in Table 3, show that this distance, measured along the polar surface of the inner monolayer, is 96 ± 2 Å, very similar to the one in the outer monolayer, 101 ± 4 Å, corresponding to a

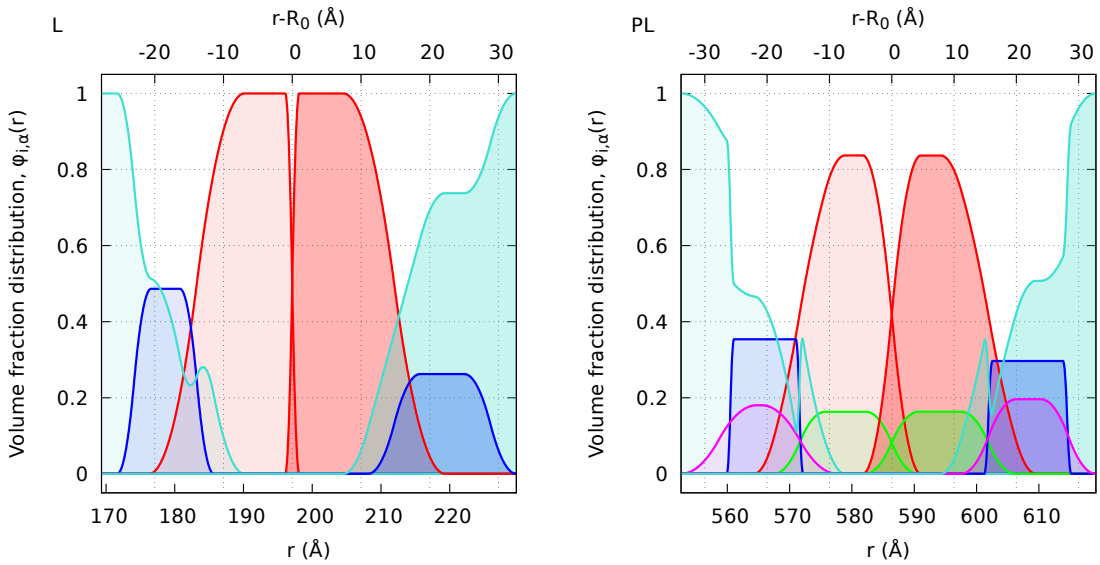

Figure 4: Volume fraction distributions of the united groups derived from the analysis of SANS data for liposomes (L, left panel) and proteoliposomes (PL, right panel) reported as a function of the radial distance r from the vesicle center (bottom horizontal axis) and as a function of the radial distance from the average vesicle radius R_0 (top horizontal axis). Filled areas with lower and higher transparency refer to inner ($\alpha = \text{in}$) and outer ($\alpha = \text{out}$) monolayer, respectively. In both panels, red, blue and turquoise distributions refer to lipid hydrophobic group, lipid polar group and water, respectively. In the right panel, green and magenta colors refer to hydrophobic protein group and polar protein group, respectively.

inner and outer surface per protein in the bilayer of 7240 and 8010 Å^2, respectively. Despite an average distance between proteins around 100 Å, no peak of the SANS intensity was detected at $q \sim 2\pi/100 = 0.06$ Å^{-1}. This means that no significant correlation in the lateral arrangement of the protein is present in the membranes, or, if a correlation exists, is not enough to modify the SANS profiles.

For the sake of completeness, we report in the Figure S1 of the SI the SLD profile of each group and their sum that have been used by the method for fitting the SANS experimental data. Their dependency on the solvent deuteration grade x_D is clearly seen, together with their significant asymmetry in the presence of protein.

The analysis of the SANS data provided a detailed structural description of liposomes containing the OprF porin on a wide range of lenghtscales. First of all, the size of the liposomes, once exposed to the cell free protein production and purification, have a significant

larger size compared to the original liposomes. This could be due to reshuffle of the liposomes during the exposure to the lysate that might induce fusion between them. The analysis returned the concentration of protein and lipids in the proteoliposomes performed on the final product that gives a more precise estimation compared to that in the phase of sample preparation. With the exception of the differences induced by the curvature of the liposomes that involve mainly the thickness and hydration of the lipid head region, the lipid bilayers show a high degree of symmetry between the inner and outer leaflet. The symmetry includes also the orientation of the OprF in the liposome at least within the simplified model used in our analysis based on the OprF closed conformer. This conformation is characterised by the presence of a globular C-terminal domain that in the real bacteria is located in the periplasm. Our analysis shows that the orientation of this C-terminal has an equal probability of being on the inner or the outer side of the liposomes. The results obtained by Lenormand et coworkers[19] on the same type or proteoliposomes, using AFM with modified tips binding the N-terminal of OprF and trypsin digestion of OprF membrane protected liposomes, suggested instead a preferential orientation of the OprF with the the globular C-terminal domain pointing toward the interior of the proteoliposome. The two experimental determinations are however not in conflict since both studies provide evidence of the presence of OprF on both orientations and the indetermination of this parameter is significantly high. The result of the SANS analysis returned the value of $x_{C,\in} = 0.5 \pm 0.2$ with an indetermination of around 40%. The indetermination of AFM results is difficult to quantify since Lenormand et coworkers[19] provide results relative to single liposomes, not an average over the whole sample of liposomes present in the solution. So within the limit of the sensitivity the two results are compatible and the different results can be ascribed to different sensitivities of the techniques used, which are subject in any case to a significant degree of uncertainty. On the other hand, our previous study performed on planar tethered lipid bilayers[22] showed that the protein was more extended in the liquid sub-phase above the distal leaflet suggesting a preferred orientation of the globular C-terminus toward this region. In that case however,

this was probably due to the steric hindrance of the tethering molecules that bound the lipid bilayer to the solid substrate. The tough competition for space in this region crowded with the tethering molecules might induce the voluminous globular part of the C-terminal away from this region and more prone to be located in the liquid sub-phase above the lipid bilayer. This steric hindrance is instead absent in the case of the protoliposomes.

Our SANS analysis provides further structural details of the protein within the bilayer, including the average volume fraction of protein in the liposome, the average area per molecule and the average distance of proteins that assume very similar values in the inner and outer leaflet. This information, very difficult to obtain with other techniques, is important for quantification of the average number of antigens present in the liposomes and could provide an indication of how to modulate the dose of a future potential vaccine.

We have also seen that the hydrophobic tail region of the liposomes is affected by the presence of the protein. After the incorporation of OprF the analysis reveals an increase in its thickness of almost 2 Å (Table 2). This behaviour is different with respect to OprF incorporated in planar tethered lipid bilayer, where we observed more pronounced reduction of the thickness of the tail region of around 6 Å.[22] Changes in the thickness of the hydrophobic region due to the presence of an integral membrane protein are generally attributed to the hydrophobic mismatch, a principle at the basis of the interaction between membrane proteins and lipid membranes that we observed experimentally also in another similar study.[37] These changes are originated from the difference between the thickness of the hydrophobic region of the lipid bilayer and of the protein contained in the membrane. The hydrophobic mismatch strongly impacts the organisation of lipids and proteins although the precise effect on the structure and function strongly depends on the individual nature of the protein-lipids systems. Several different structural rearrangements can be induced, including protein oligomerisation, protein and lipid conformational changes and membrane curvature. The different effect of protein inclusion obtained in this work on liposomes with respect to the case of planar lipid bilayers[22] might be due to a combination of factors. On

one hand, the curvature of the two surfaces are significantly different. On the other hand, the planar tethered lipid bilayer was composed by a single lipid, whereas the lipid composition of the liposomes of the present study was of four lipids. The larger variety of lipids allowed probably a more pronounced structural flexibility to preferentially accomodate some lipids to better accomodate the protein and minimise the mismatch.

Conclusions

We performed a detailed nanostructural characterisation of proteoliposomes containing OprF, the main porin of the *P. aeruginosa*, which has recently been identified as a potential candidate for a vaccine against this deadly pathogen. We described a method that we developed to analyse SANS data from spherical lipid vesicles based on the volume fraction radial distribution of the relevant molecular groups to compute the form factor. One of the main novelties and advantage of our approach is that the SLD profile is computed in terms of the volume distribution of the chemical groups directly in spherical geometry. In previous models the volume distribution of the chemical groups was realized in planar geometry,[20,23–25] as the radius of the vesicles was supposed so large (infinite) that the curvature could be neglected. For smaller vesicles, models invoking spherical shells at constant density were used that were not constrained to the volume distribution of the components inside the membrane.[30] Hence, our model extends the volume fraction approach to compute the SLD profiles also to vesicles with finite radius. Each group, including those of the protein, contributes to the total SLD profile according on how it is distributed along the radial coordinate. The fundamental constrain that the sum of the volume fractions equals 1 implies that the presence of a g-group at a given position (as for example the hydrophobic protein domain) replaces the space that would have been occupied by the other group in the absence of the g-group. Even the solvent molecules among the polar groups are described with a volume fraction distribution obtained by complementing to 1 the sum of the volume fraction of the other groups (see the turquoise

profiles in Fig. 4)

The methodology used in this study is very flexible and can be easily adapted to different classes of liposome systems containing different host molecules in the bilayer, in the interior or on the surface of the liposome, providing a useful tool to obtain a detailed average structural characterization that could be difficult to obtain with other techniques.

The analysis provided detailed information on the lipid bilayer structure, on the amount of OprF in the lipid bilayer, their average localisation, orientation and the effect of the protein incorporation on the lipid bilayer. This information is valuable for fine-tuning the proteoliposome delivery system, for example in quantifying the dosage of proteins delivered for therapy.

Supporting Information

Power expansion coefficients of normalisation factors; excess SLD profile; Fourier transform of $f(r)$; calculation of $F(q)$ and $F^2(q)$; volume of the solution inside the vesicle; area per molecular unit; third class fitting parameters; SLD profiles obtained from SANS data

Acknowledgement

The work reported here was supported financially by grants from the French Investissements d'Avenir (ANR-10-NANO-03-01, 2012–2016) and partially funded from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 823780. This work is based upon experiments performed at the KWS-2 instrument operated by Jülich Centre for Neutron Science at the Heinz Maier-Leibnitz Zentrum (MLZ), Garching, Germany.

References

(1) Tenchov, R.; Bird, R.; Curtze, A. E.; Zhou, Q. Lipid Nanoparticles - From Liposomes to mRNA Vaccine Delivery, a Landscape of Research Diversity and Advancement. *ACS Nano* **2021**, *15*, 16982–17015.

(2) Emergency Use Authorization (Eua) of the Pfizer-Biontech Covid-19 Vaccine to Prevent Coronavirus Disease 2019 (COVID-19) in Individuals 16 Years of Age and Older. `https://www.fda.gov/media/144414/download`, (accessed 2020-12-22).

(3) Emergency Use Authorization (Eua) of the Moderna Covid-19 Vaccine to Prevent Coronavirus Disease 2019 (Covid-19) in Individuals 18 Years of Age and Older. `https://www.fda.gov/media/144638/download`, (accessed 2020-12-22).

(4) Kučerka, N.; Nagle, J. F.; Sachs, J. N.; Feller, S. E.; Pencer, J.; Jackson, A.; Katsaras, J. Lipid Bilayer Structure Determined by the Simultaneous Analysis of Neutron and X-Ray Scattering Data. *Biophys. J.* **2008**, *95*, 2356–2367.

(5) Di Cola, E.; Grillo, I.; Ristori, S. Small Angle X-ray and Neutron Scattering: Powerful Tools for Studying the Structure of Drug-Loaded Liposomes. *Pharmaceutics* **2016**, *8*.

(6) Hassett, D. J.; Borchers, M. T.; Panos, R. J. Chronic obstructive pulmonary disease (COPD): Evaluation from clinical, immunological and bacterial pathogenesis perspectives. *Journal of Microbiology* **2014**, *52*, 211–226.

(7) Benz, R.; Hancock, R. E. Properties of the large ion-permeable pores formed from protein F of Pseudomonas aeruginosa in lipid bilayer membranes. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1981**, *646*, 298–308.

(8) Wu, L.; Estrada, O.; Zaborina, O.; Bains, M.; Shen, L.; Kohler, J. E.; Patel, N.; Musch, M. W.; Chang, E. B.; Fu, Y.-X.; Jacobs, M. A.; Nishimura, M. I.; Han-

cock, R. E. W.; Turner, J. R.; Alverdy, J. C. Recognition of Host Immune Activation by ¡i¿Pseudomonas aeruginosa¡/i¿. *Science* **2005**, *309*, 774–777.

(9) Wessel, A. K.; Liew, J.; Kwon, T.; Marcotte, E. M.; Whiteley, M. Role of Pseudomonas aeruginosa Peptidoglycan-Associated Outer Membrane Proteins in Vesicle Formation. *Journal of Bacteriology* **2013**, *195*, 213–219.

(10) Ballok, A. E.; Filkins, L. M.; Bomberger, J. M.; Stanton, B. A.; O'Toole, G. A. Epoxide-Mediated Differential Packaging of Cif and Other Virulence Factors into Outer Membrane Vesicles. *Journal of Bacteriology* **2014**, *196*, 3633–3642.

(11) Sugawara, E.; Nagano, K.; Nikaido, H. Alternative folding pathways of the major porin OprF of Pseudomona aeruginosa. *The FEBS Journal* **2012**, *279*, 910–918.

(12) Sugawara, E.; Nestorovich, E. M.; Bezrukov, S. M.; Nikaido, H. Pseudomonas aeruginosa Porin OprF Exists in Two Different Conformations. *Journal of Biological Chemistry* **2006**, *281*, 16220–16229.

(13) Brennan, F. R.; Jones, T. D.; Gilleland, L. B.; Bellaby, T.; Xu, F.; North, P. C.; Thompson, A.; Staczek, J.; Lin, T.; Johnson, J. E.; Hamilton, W. D. O.; Gilleland, H. E. Pseudomonas aeruginosa outer-membrane protein F epitopes are highly immunogenic in mice when expressed on a plant virus. *Microbiology* **1999**, *145*, 211–220.

(14) McGhee, J. R.; Mansouri, E.; Gabelsberger, J.; Knapp, B.; Hundt, E.; Lenz, U.; Hungerer, K.-D.; Gilleland, H. E.; Staczek, J.; Domdey, H.; von Specht, B.-U. Safety and Immunogenicity of a ¡i¿Pseudomonas aeruginosa¡/i¿ Hybrid Outer Membrane Protein F-I Vaccine in Human Volunteers. *Infection and Immunity* **1999**, *67*, 1461–1470.

(15) Matthews-Greer, J. M.; Robertson, D. E.; Gilleland, L. B.; Gilleland, H. E. Pseudomonas aeruginosa outer membrane protein F produced inEscherichia coli retains vaccine efficacy. *Current Microbiology* **1990**, *20*, 171–175.

(16) Krause, A.; Whu, W. Z.; Qiu, J.; Wafadari, D.; Hackett, N. R.; Sharma, A.; Crystal, R. G.; Worgall, S. RGD capsid modification enhances mucosal protective immunity of a non-human primate adenovirus vector expressing Pseudomona aeruginosa OprF. *Clinical & Experimental Immunology* **2013**, *173*, 230–241.

(17) Sharma, A.; Krause, A.; Xu, Y.; Sung, B.; Wu, W.; Worgall, S. Adenovirus-Based Vaccine with Epitopes Incorporated in Novel Fiber Sites to Induce Protective Immunity against Pseudomonas aeruginosa. *PLOS ONE* **2013**, *8*, 1–9.

(18) Bahey-El-Din, M.; Mohamed, S. A.; Sheweita, S. A.; Haroun, M.; Zaghloul, T. I. Recombinant N-terminal outer membrane porin (OprF) of Pseudomonas aeruginosa is a promising vaccine candidate against both P. aeruginosa and some strains of Acinetobacter baumannii. *International Journal of Medical Microbiology* **2020**, *310*, 151415.

(19) Mayeux, G.; Gayet, L.; Liguori, L.; Odier, M.; Martin, D. K.; Cortès, S.; Schaack, B.; Lenormand, J.-L. Cell-free expression of the outer membrane protein OprF of Pseudomonas aeruginosa for vaccine purposes. *Life Science Alliance* **2021**, *4*, e202000958, 1–16.

(20) De Rosa, R.; Spinozzi, F.; Itri, R. Hydroperoxide and carboxyl groups preferential location in oxidized biomembranes experimentally determined by small angle X-ray scattering: Implications in membrane structure. *BBA - Biomembranes* **2018**, *1860*, 2299–2307.

(21) Yoneda, J. S.; de Araujo, D. R.; Sella, F.; Liguori, G. R.; Liguori, T. T.; Moreira, L. F. P.; Spinozzi, F.; Mariani, P.; Itri, R. Self-assembled guanosine-hydrogels for drug-delivery application: Structural and mechanical characterization, methylene blue loading and controlled release. *Materials Science & Engineering C* **2021**, *121*, 111834.

(22) Maccarini, M.; Gayet, L.; Alcaraz, J.-P.; Liguori, L.; Stidder, B.; Watkins, E. B.; Lenormand, J.-L.; Martin, D. K. Functional Characterization of Cell-Free Expressed OprF

Porin from Pseudomonas aeruginosa Stably Incorporated in Tethered Lipid Bilayers. *Langmuir* **2017**, *33*, 9988–9996.

(23) Petrache, H. I.; Feller, S. E.; Nagle, J. F. Determination of component volumes of lipid bilayers from simulations. *Biophys. J.* **1997**, *72*, 2237–2242.

(24) Klauda, J. B.; Kučerka, N.; Brooks, B. R.; Pastor, R. W.; Nagle, J. F. Simulation-Based Methods for Interpreting X-Ray Data from Lipid Bilayers. *Biophys. J.* **2006**, *90*, 2796–2807.

(25) Pan, J.; Cheng, X.; Sharp, M.; Ho, C.-S.; Khadka, N.; Katsaras, J. Structural and mechanical properties of cardiolipin lipid bilayers determined using neutron spin echo, small angle neutron and X-ray scattering, and molecular dynamics simulations. *Soft Matter* **2015**, *11*, 130–138.

(26) Berndt, I.; Pedersen, J. S.; Richtering, W. Structure of Multiresponsive Intelligent Core-Shell Microgels. *JACS* **2005**, *127*, 9372–9373.

(27) Pabst, G.; Kučerka, N.; Nieh, M.-P.; Rheinstädter, M.; Katsaras, J. Applications of neutron and X-ray scattering to the study of biologically relevant model membranes. *Chemistry and Physics of Lipids* **2010**, *163*, 460–479.

(28) Nielsen, J. E.; Bjørnestad, V. A.; Lund, R. Resolving the structural interactions between antimicrobial peptides and lipid membranes using small-angle scattering methods: the case of indolicidin. *Soft Matter* **2018**, *14*, 8750–8763.

(29) Pachler, M.; Kabelka, I.; Appavou, M.-S.; Lohner, K.; Vácha, R.; Pabst, G. Magainin 2 and PGLa in Bacterial Membrane Mimics I: Peptide-Peptide and Lipid-Peptide Interactions. *Biophysical Journal* **2019**, *117*, 1858–1869.

(30) Pencer, J.; Krueger, S.; Adams, C. P.; Katsaras, J. Method of separated form factors for polydisperse vesicles. *Journal of Applied Crystallography* **2006**, *39*, 293–303.

(31) Pedersen, J. S. Determination of size distribution from small-angle scattering data for systems with effective hard-sphere interactions. *Journal of Applied Crystallography* **1994**, *27*, 595–608.

(32) Spinozzi, F.; Ferrero, C.; Ortore, M. G.; Antolinos, A. D. M.; Mariani, P. GENFIT: software for the analysis of small-angle X-ray and neutron scattering data of macromolecules in-solution. *J. App. Cryst.* **2014**, *47*, 1132–1139.

(33) Pedersen, J. S.; Posselt, D.; Mortensen, K. Analytical Treatment of the Resolution Function for Small-Angle Scattering. *J. Appl. Cryst.* **1990**, *23*, 321–333.

(34) Radulescu, A.; Pipich, V.; Frielinghaus, H.; Appavou, M. S. KWS-2, the high intensity / wide Q-range small-angle neutron diffractometer for soft-matter and biology at FRM II. *Journal of Physics: Conference Series* **2012**, *351*, 012026.

(35) Marsh, D. Molecular volumes of phospholipids and glycolipids in membranes. *Chem. Phys. Lipids* **2010**, *163*, 667 – 677.

(36) Jacrot, B. The Study of Biological Structures by Neutron Scattering From Solution. *Reg. Prog. Phys.* **1976**, *39*, 911.

(37) Köhler, S.; Fragneto, G.; Alcaraz, J.-P.; Nelson, A.; Martin, D. K.; Maccarini, M. Nanostructural Characterization of Cardiolipin-Containing Tethered Lipid Bilayers Adsorbed on Gold and Silicon Substrates for Protein Incorporation. *Langmuir* **2021**, *37*, 8908–8923, PMID: 34286589.