Electronic Supplementary Information

Effects of surface ligand density on lipid-monolayer-mediated 2D assembly of proteins

Masafumi Fukuto, Suntao Wang, Matthew A. Lohr, Sumit Kewalramani and Lin Yang

1. Analysis of x-ray reflectivity data

The fitting of the XR data was carried out by using the standard "*n*-box" model for the average electron density profile $\langle\rho(z)\rangle$,[1] which is based on the combination of $n + 1$ error functions. In this model, each of the n layers assumed between the aqueous subphase ($\langle\rho\rangle = \rho_{buffer} = 0.339$ electrons/Å^3) and the vapor above ($\langle\rho\rangle = 0$) is represented by a box of height ρ_i and thickness l_i, and each of the $n + 1$ interfaces is smeared out by a Gaussian roughness σ_i. In the analysis, the box-model profile $\langle\rho(z)\rangle$ was first divided into a stack of constant-density, 0.5-Å-thick slabs, and the exact reflectivity was calculated for this set of slabs by using the matrix method of the Parratt formalism.[2, 3] The theoretical reflectivity curve thus calculated was fitted to the data for $q_z \geq 0.05$ Å^{-1} by varying the box-model parameters (ρ_i, l_i, σ_i). It should be noted that XR is sensitive only to the profile $\langle\rho(z)\rangle$ itself, i.e., not to exactly how $\langle\rho(z)\rangle$ is constructed. For most of the data, the use of the 4-box model was necessary (and sufficient) to produce good fits, where the four layers correspond to the protein layer ($i = 1$), a low-density region between the protein and the lipid ($i = 2$), the lipid head group ($i = 3$), and the lipid tails ($i = 4$) (see Fig. 5C,D). The box parameters describing the lipid part of $\langle\rho(z)\rangle$ were coupled such that different sets of these parameters could be found that produced nearly

the same lipid-layer profile. However, the protein layer's electron density ρ_1 and thickness l_1 were well defined and could be determined with relatively small uncertainties, as they were largely independent of the lipid box parameters.

2. Model adsorption isotherms

The model adsorption curves $\Gamma(a_b)$ shown in Fig. 4B are based on the following general formula:

$$\Gamma(a_b) = \begin{cases} \dfrac{2-\langle x_2 \rangle}{a_b} & \text{if } a_b > 2-\langle x_2 \rangle \\ 1 & \text{if } a_b \le 2-\langle x_2 \rangle \end{cases}. \qquad (1)$$

Here, the relative protein adsorption Γ measures the amount of interface-bound proteins and is normalized to be unity for the full surface coverage by a monolayer of the 2D crystal of SA; $a_b = A_b/A_s$ is the ratio of the area/biotin, A_b, to the area per lipid-facing ligand-binding site in the 2D crystal of SA, A_s; and $<x_2>$ is the expectation value for the fraction of biotinylated lipids that are bound to SA-b_2 below saturation, i.e., $\Gamma < 1$. The three model curves in Fig. 4B differ in the values of $<x_2>$.

In deriving Eq. (1), we have made only two assumptions: (i) nonspecific binding of SA to the interface is negligible, and (ii) all the biotin-lipids on the surface will eventually be bound to SA so long as the resulting surface density of bound-SA is less than that in 2D crystals, i.e., $\Gamma < 1$. These assumptions describe the case in which the energetic gain of protein adsorption is dominated by the specific binding of SA to biotin and remains independent of the lateral protein packing density until the crystalline

density is reached. Assumption (i) above is supported by the study of the adsorption of SA onto lipid monolayers by Lösche et al.,[4] whose neutron-reflectivity and fluorescence-microscopy results "gave no indication for the presence of proteins at the interface" in the absence of biotinylated lipids. Assumption (ii) is equivalent to approximating the strong binding affinity between SA and biotin (dissociation constant $K_d \sim 10^{-14}$ M) with the limit of infinitely high affinity ($K_d \rightarrow 0$). The plausibility of these assumptions is also supported by our XR observation that a doubling of the biotin density, or decrease in a_b from 2 to 1, is accompanied by an increase in protein adsorption Γ by a factor of roughly two (symbols in Fig. 4B).

According to the above assumptions, the number of the doubly-bound protein SA-b_2 (see Fig. 1) below saturation ($\Gamma < 1$) is equal to $n\langle x_2\rangle/2$ where n denotes the total number of the biotinylated lipids. Similarly, the number of the singly-bound SA-b_1 is equal to $n(1 - \langle x_2\rangle)$. Combining the two contributions, the number of bound proteins (N_p) per unit area (A) for $\Gamma < 1$ is given by:

$$N_p/A = (1/A_b)[(1 - \langle x_2\rangle) + \langle x_2\rangle/2], \quad (2)$$

where $A_b = A/n$ is the area/biotin defined earlier. In the 2D crystals of SA, each protein takes up an area of $2A_s$ by definition, since two of its ligand-binding sites point toward the lipid monolayer. Thus, the surface density of proteins in the 2D crystal is given by

$$(N_p/A)_x = 1/(2A_s). \quad (3)$$

The normalized adsorption is equal to the ratio between the two protein densities above, i.e., $\Gamma = (N_p/A)/(N_p/A)_x$. This reduces to Eq. (1) via $a_b = A_b/A_s$ and also shows that the saturation $\Gamma = 1$ occurs at $a_b = 2 - \langle x_2\rangle$.

It is clear that the model adsorption isotherm $\Gamma(a_b)$ depends on the binding state(s) of the bound proteins through the parameter $\langle x_2 \rangle$. The extreme case of $\langle x_2 \rangle = 1$ (solid curve in Fig. 4B) describes the situation in which the adsorption is dominated by the formation of SA-b_2. The other extreme at $\langle x_2 \rangle = 0$ (dotted curve in Fig. 4B) would correspond to the case in which the adsorption resulted primarily from the formation of SA-b_1. In the case of random binding (dashed curve in Fig. 4B) where there is no preference for the formation of SA-b_2 over SA-b_1 and vice versa, we estimate $\langle x_2 \rangle \sim 0.6$ from the following simple statistical considerations.

If the energetics of the protein adsorption were dominated simply by the number of SA-biotin bonds and were independent of whether the binding resulted in the formation of SA-b_2 or SA-b_1, then, the equilibrium distribution between the two binding states would be dictated by the configurational entropy. Suppose that a total of n biotinylated lipids form a 2D lattice with each having z nearest neighbors. Then, the total number of the nearest-neighbor "bonds" between biotins is equal to

$$M = nz/2. \qquad (4)$$

Each SA-b_2 could be considered to occupy one of these M "bonds" in the 2D plane. With x_2 denoting the fraction of biotin lipids that contribute to the formation of SA-b_2, the number of SA-b_2 on the surface is equal to $x_2 n/2$.

If the same biotin could be shared by two adjacent SA-b_2 proteins, the total number of configurations in this fictitious case would be given by

$$\Omega_0 = \frac{M!}{\left(\frac{1}{2} x_2 n\right)! \left(M - \frac{1}{2} x_2 n\right)!}. \qquad (5)$$

In reality, two adjacent SA-b_2 proteins *cannot* share a biotin. Thus, for each occupied bond, all of its $2(z-1)$ neighboring bonds must be unoccupied. The probability that this holds for any one of these M bonds is

$$p_1 = \left(\frac{M - \frac{1}{2} x_2 n}{M} \right)^{2(z-1)} . \qquad (6)$$

The probability that the above condition holds for all of the $x_2 n/2$ bonds to be occupied is

$$p = (p_1)^{\frac{1}{2} x_2 n} = \left(\frac{M - \frac{1}{2} x_2 n}{M} \right)^{x_2 n (z-1)} . \qquad (7)$$

The total number of configurations when the "biotin sharing" is forbidden is then given by

$$\Omega = \Omega_0 p. \qquad (8)$$

Substituting Eqs. (4)-(7) into Eq. (8) and applying the Stirling's approximation for large n, one obtains

$$\ln(\Omega) = \tfrac{1}{2} z n \left[\left\{ \frac{x_2}{z} (2z-1) - 1 \right\} \ln\left(1 - \frac{x_2}{z} \right) - \left(\frac{x_2}{z} \right) \ln\left(\frac{x_2}{z} \right) \right]. \qquad (9)$$

The expectation value $\langle x_2 \rangle$ is given by the value of x_2 that maximizes Ω. It can be easily shown that the larger n is, the more sharply peaked Ω is at $x_2 = \langle x_2 \rangle$. The dependence of $\langle x_2 \rangle$ on the coordination number z is rather weak; Eq. (9) gives: $\langle x_2 \rangle = 0.55$ for $z = 4$, $\langle x_2 \rangle = 0.58$ for $z = 5$, and $\langle x_2 \rangle = 0.61$ for $z = 6$. The model adsorption curve for the random-binding case (dashed curve in Fig. 4B) is obtained by setting $\langle x_2 \rangle = 0.6$ in Eq. (1).

References

1. J. Als-Nielsen and D. McMorrow, *Elements of Modern X-ray Physics*, Wiley, Chichester, 2001.
2. L. G. Parratt, *Phys. Rev.*, 1954, **95**, 359-369.
3. J. Lekner, *Theory of Reflection*, Martin Nijhoff, Dordrecht, 1987.
4. M. Lösche, M. Piepenstock, A. Diederich, T. Grunewald, K. Kjaer and D. Vaknin, *Biophys. J.*, 1993, **65**, 2160-2177.