PCCP

Accepted Manuscript

This article can be cited before page numbers have been issued, to do this please use: J. I. Lopez-Ortiz, P. B. Torres, E. Quiroga, C. F. Narambuena and A. Ramirez-Pastor, *Phys. Chem. Chem. Phys.*, 2017, DOI: 10.1039/C7CP06618J.

ROYAL SOCIETY OF CHEMISTRY

rsc.li/pccp

View Article Online
DOI: 10.1039/C7CP06618J

Journal Name

ARTICLE TYPE

Cite this: DOI: 10.1039/xxxxxxxxxx

Adsorption of three-domain antifreeze proteins on ice: a study using LGMMAS theory and Monte Carlo simulations

Juan Ignacio Lopez Ortiz[a], Paola Torres[a,b], Evelina Quiroga[a,c], Claudio F. Narambuena[a,b], and Antonio J. Ramirez-Pastor[a,*]

Received Date
Accepted Date

DOI: 10.1039/xxxxxxxxxx

www.rsc.org/journalname

In the present work, the adsorption of three-domain antifreeze proteins on ice is studied by combining a statistical thermodynamics based-theory and Monte Carlo simulations. The three-domain protein is modeled by a trimer, and the ice surface is represented by a lattice of adsorption sites. The statistical theory, obtained from the exact partition function of non-interacting trimers adsorbed in one dimension and its extension to two dimensions, includes the configuration of the molecule in the adsorbed state, and allows the existence of multiple adsorption states for the protein. We called this theory "lattice-gas model of molecules with multiple adsorption states" (LGMMAS). The main thermodynamics functions (partial and total adsorption isotherms, Helmholtz free energy and configurational entropy) are obtained by solving a non-linear system of j equations, where j is the total number of possible adsorption states of the protein. The theoretical results are contrasted with Monte Carlo simulations, and a modified Langmuir model (MLM) where the arrangement of the adsorption sites in space is immaterial. The formalism introduced here provides exact results in one-dimensional lattices, and offers a very accurate description in two dimensions (2D). In addition, the scheme is capable to predict the proportion between coverage degrees corresponding to different conformations in the same energetic state. In contrast, the MLM does not distinguish between different adsorption states, and shows severe discrepancies with the 2D simulation results. This finding indicates that adsorbate structure and lattice geometry play fundamental roles in determining the statistics of multistate adsorbed molecules, and consequently, must be included in the theory.

1 Introduction

Certain organisms, such as fishes, insects, plants, fungi and bacteria, are protected against the stresses that occur during freezing at subzero temperatures, by the presence of antifreeze proteins (AFPs) in their body fluids[1–6]. These proteins have the ability to bind to ice crystals and prevent the water around from freezing[7–9]. Such proteins could have various potential applications. They could be used in medicine for cryopreservation of organs (freezing at extremely low temperatures), for transplantation or to improve the quality of sperm, ovules and embryos stored in a frozen state[10,11]. They can also have application to water-based materials, such as food and waterborne paints[12–18].

At present, it is not fully understood how AFPs interact with ice and prevent the growth of ice crystals. However, different mechanisms have been proposed, including the protein acting as an impurity, remodeling of the ice lattice, decreasing the radius of curvature of the ice surface, etc[19–25]. In the last mechanism, the adsorbed AFPs cause an increase in the micro-curvature of the ice front, and inhibit their growth[19–21]. The curved growth increases the vapor pressure of the ice, thus halting the further growth of the ice[22–24]. This protein-ice interaction lowers the freezing point of the solution. The temperature separation created by this phenomenon is called thermal hysteresis (TH) and can be defined as the gap between melting and freezing points[24]. The TH value is a non-lineal function of protein concentration in bulk (increases of a non-colligative manner).

The detailed experimental study on how AFPs and water molecules arrange themselves at the ice surface is a challenging task. In turn, the adsorption isotherms of AFPs on ice are ex-

[a] Instituto de Física Aplicada, Universidad Nacional de San Luis-CONICET, Ejército de Los Andes 950, D5700BWS San Luis, Argentina
[b] Universidad Tecnológica Nacional, Facultad Regional San Rafael, Gral. Urquiza 314, 5600, San Rafael, Mendoza, Argentina
[c] Laboratorio de Membranas y Biomateriales, Instituto de Física Aplicada, Universidad Nacional de San Luis-CONICET, San Luis, Argentina
∗ Corresponding author.
Email: antorami@unsl.edu.ar

View Article Online
DOI: 10.1039/C7CP06618J

perimentally inaccessible. Instead, numerous computer simulations have been developed. Among them, Nada and Furukawa[26] used a molecular dynamics simulation to investigate the growth kinetics at the ice prismatic plane interface to which an AFP was bound. In these surface phenomena, the coverage degree of proteins on the ice surface plays a key role, since this dictates the distance between adsorbed proteins and the final micro-curvature of ice surface.

Xu et al.[27] examined, by using all-atom molecular dynamics simulations, the dynamics of water molecules and hydrogen bonds at the protein-water interface of an AFP from the spruce budworm *Choristoneura fumiferana* and a mutant that has little antifreeze activity. The authors explored the effects of the heterogeneity of the protein surface, and how the dynamics is affected by mutation. The protein-water dynamics was also studied in Ref.[28]. By a combination of terahertz spectroscopy and molecular dynamic simulations, Meister et al. showed that long-range protein-water interactions play an important role in explaining the hyperactive antifreeze activity of insect AFPs. Based on the results of Refs.[27,28], Pandey and Leitner[29] investigated the thermodynamic properties of the antifreeze protein DAFP-1 and its hydration water by molecular simulations. At each level of hydration, the specific heats, entropy, and free energy of mixing water and protein were calculated.

There are AFPs with two or more binding domains or segments links together. In this sense, a modified Langmuir model (MLM) has been used for studying the reversible adsorption of two different AFPs (a single and two-domain protein) onto an ice crystal[30]. An alternative approach, using the statistical mechanics formalism, was developed[31]. In Ref.[31], a lattice-gas model of molecules with multiple adsorption states (LGMMAS) was applied to describe the adsorption of AFPs onto an ice crystal where the proteins were modeled as chains of n identical units (domains) connected by flexible linkers, which can be adsorbed in n different adsorption states. The adsorption mechanism of two-domain AFPs on ice was studied by using Monte Carlo (MC) simulations, MLM and LGMMAS[32]. The ice surface was modeled as a square lattice, and the protein was modeled as a dimer with two possible adsorption states: upright and flat conformations. In the upright conformation the protein is adsorbed only by one of the domains, and adopts a perpendicular orientation with respect to the ice surface. In the flat conformation the protein is adsorbed with the two domains occupying two nearest-neighbor adsorption sites on the lattice. The total and partial adsorption isotherms, and the concentration dependence of the Helmholtz free energy and the configurational entropy were obtained. Even though a good qualitative agreement was obtained between MLM and MC data, it was found that the LGMMAS offers a more accurate description of the phenomenon of adsorption of two-domain AFPs.[32]

In the line of previous work[30–32], Can and Holland[33] extended the study by designing a three-domain type III AFP, which is formed by three domains that can independently bind to ice. As a result of the increasing size of the protein (increasing number of bound domains), the three-domain AFP yielded significantly greater activity than the one and two-domain proteins. A similar effect was proposed by Kubota and Mullin[34] in the case of crystal growth from aqueous solution in the presence of impurities. The authors suggested that the crystal growth mechanism changes with the size of the impurity.

In Ref.[33], a very simple generalization of the MLM was proposed to model the adsorption of three-domain AFPs on ice. In this scheme, the arrangement of the adsorption sites in space is immaterial, and consequently, the model does not distinguish between different adsorption states. Even though interesting results were obtained on the relationship between protein surface coverage and thermal hysteresis activity, the limitations of the MLM did not allow to investigate the effect of competitive adsorption between the different configurational states of the trimers (linear and angular trimers). These considerations encourage the development of more complex analytical solutions, capable of describing the adsorption of large molecules with multiple adsorption states.

In this context, the main objective of this paper is to present a more refined model for the analysis of three-domain AFPs adsorption. For this purpose, and based on the formalism presented in Ref.[31], the main thermodynamic functions (total and partials adsorption isotherms, free energy and configurational entropy) characterizing the adsorption of three-domain AFP (with different adsorption states) on one-dimensional (1D) and two-dimensional (2D) substrates were obtained. In the case of 2D lattices, special attention was devoted to the study of the competitive adsorption between linear and angular trimers. This phenomenon is reported here for the first time in this type of systems. The validity of the analytical results was evaluated by comparing with previous results in the literature[33], and extensive Monte Carlo simulations of the system. The comparison indicates that the proposed theoretical framework represents a qualitative advance with respect to the existing models to describe the adsorption thermodynamics of structurally diverse proteins.

The paper is organized as follows: the exact solution for the thermodynamic functions of three-domain proteins adsorbed in a infinite 1D space is presented in Section 2. The functions are further extended to higher dimensions based upon their exact form in one dimension and a connectivity ansatz (this theoretical is called as LGMMAS). The description of the MLM developed in Ref.[33] is given in Section 3. The simulation scheme is described in Section 4. Simulation results and theoretical predictions are discussed and compared in Section 5. Finally, the conclusions are drawn in Section 6.

2 Theory: lattice-gas model and thermodynamic functions

Let us assume that the ice surface is represented by a lattice of M adsorption sites, with lattice constant a, connectivity c ($c = 2$ for 1D lattices, and $c = 4$ for 2D square lattices), and periodic boundary conditions. Thus, all lattice sites are equivalent, and these are in contact with a protein solution of concentration C_P.

The general statistical-mechanics derivation for a lattice-gas model of molecules with multiple adsorption states (LGMMAS) was presented in Ref.[31]. This theory is applied here to calculate the main thermodynamic functions of the system under study.

View Article Online
DOI: 10.1039/C7CP06618J

Thus, N protein molecules are adsorbed on the surface with the following considerations: (1) each protein molecule is constituted by three identical domains, which are linearly (and covalently) connected by a peptide segment; (2) a molecule can be adsorbed on the lattice in three states S_1, S_2 and S_3, according to the number of domains linked to the lattice (see Fig. 1). The model mimics the adsorption of a type III AFP as introduced in Ref.[33].

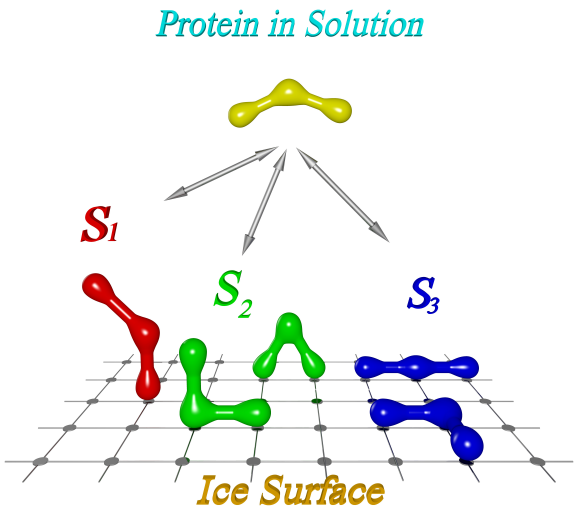

Fig. 1 Three-domain antifreeze proteins adsorbed with one, two, and three domains bound to the surface. These states are denoted as S_1, S_2 and S_3, respectively.

A molecule in the S_1 state is assumed to be a molecule adsorbed perpendicular to the surface, occupying one lattice site. Next, a molecule in the S_2 state is a molecule with two domains adsorbed on two nearest neighbor lattice sites, and the third domain unbound (non-adsorbed). As shown in Fig. 1, S_2 state includes (*i*) molecules with two consecutive domains on the lattice and the third domain in contact with the solution, and (*ii*) molecules with the two end domains bound to the surface and the central domain unbound. Lastly, in the S_3 state, the protein is adsorbed occupying three sites on the lattice; and (3) the only interaction between different proteins is hard-core exclusion: no site can be occupied by more than one protein.

According to the lattice geometry and protein molecule flexibility, LGMMAS theory allows several possible molecule configurations in the S_3 state. In the particular case of a 1D system ($c = 2$), the only option for the protein in the S_3 state is the adsorption of the three domains in a collinear form. On the other hand, in the case of square lattices ($c = 4$), two different conformations can be observed: (*i*) a linear form, the three segments are adsorbed along one of the two possible directions of the lattice (see panel A, Fig. 2), and (*ii*) an angular form, the three segments form a right angle (as shown in panel B, Fig. 2.)

Since the adsorbed proteins do not interact with each other (except the excluded volume interaction), all configurations with N_1 molecules in S_1 state, N_2 molecules in S_2 state and N_3 molecules in S_3 state have the same energy:

$$E(N_1,N_2,N_3) = \varepsilon_1 N_1 + \varepsilon_2 N_2 + \varepsilon_3 N_3, \qquad (N = N_1 + N_2 + N_3), \quad (1)$$

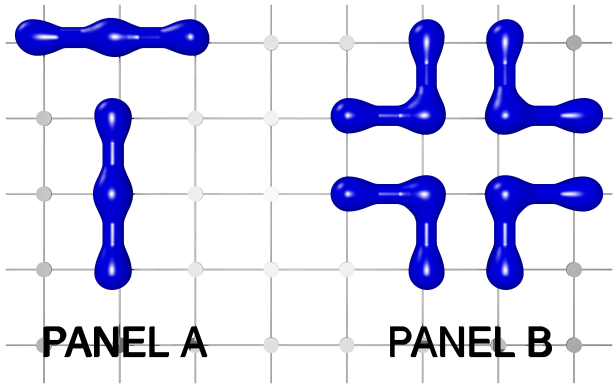

Fig. 2 Possible configurations for a trimer protein adsorbed on a 2D square lattice in S_3 state. Linear and angular forms are shown in panel A and panel B, respectively.

where ε_i is the adsorption energy for molecules adsorbed in S_i state ($i \equiv 1,2,3$). Then, the canonical partition function for this system can be written as:

$$Q(N_1,N_2,N_3,M,T) = \Omega(N_1,N_2,N_3,M)\exp\left[-\frac{E(N_1,N_2,N_3)}{k_B T}\right], \quad (2)$$

where $\Omega(N_1,N_2,N_3,M)$ is the number of configurations or ways to arrange N molecules distributed in N_1, N_2, and N_3 adsorbed proteins on a lattice of M sites. In addition, T is the absolute temperature and k_B is the Boltzmann constant. In order to simplify the calculations, the internal and vibrational contributions to the partition function are assumed to be a unitary factor in Eq. (2).

In the canonical ensemble, the Helmholtz free energy $F(N_1,N_2,N_3,M,T)$ relates to the partition function through

$$\begin{aligned}\beta F(N_1,N_2,N_3,M,T) &= -\ln Q(N_1,N_2,N_3,M,T) \\ &= -\ln\Omega(N_1,N_2,N_3,M) + \beta\sum_{i=1}^{3}\varepsilon_i N_i \qquad (3)\end{aligned}$$

where $\beta = 1/k_b T$.

The configurational entropy S, can be written as[35]

$$S(N_1,N_2,N_3,M,T) = -\left(\frac{\partial F}{\partial T}\right)_{M,N_1,N_2,N_3}. \qquad (4)$$

On the other hand, the chemical potential of the protein adsorbed in the S_j state, $\mu_{j,ads}$ can be calculated as[35]

$$\mu_{j,ads} = \left(\frac{\partial F}{\partial N_j}\right), \qquad (5)$$

where j represents each one of the different adsorbed states of the protein.

At equilibrium, the chemical potential of the adsorbed and so-

View Article Online
DOI: 10.1039/C7CP06618J

lution phase are equal. Then,

$$\mu_{j,ads} = \mu_{sol}, \tag{6}$$

where μ_{sol} represents the chemical potential corresponding to an ideal solution, at temperature T and concentration C_P. Hence,

$$\beta\mu_{sol} = \beta\mu^0 + \ln C_P, \tag{7}$$

where μ^0 is the standard chemical potential.

In the next sections, the main adsorption thermodynamic functions corresponding to three-domain antifreeze proteins adsorbed on 1D and 2D lattices will be obtained.

2.1 One-dimensional lattices

In the particular case of a 1D lattice, $\Omega(N_1,N_2,N_3,M)$ can be exactly calculated as the total number of permutations of N_e entities on the surface. This total number of entities takes into account the number N_1, N_2 and N_3 of indistinguishable adsorbed molecules and N_0 empty sites ($N_0 = M - \sum_{i=1}^{3} iN_i$). Then, N_e results:

$$\begin{aligned} N_e &= N_1 + N_2 + N_3 + N_0 \\ &= \sum_{i=1}^{3} N_i + M - \sum_{i=1}^{3} iN_i = M - \sum_{i=1}^{3} (i-1)N_i. \end{aligned} \tag{8}$$

Accordingly,

$$\Omega(N_1,N_2,N_3,M) = \frac{[M - \sum_{i=1}^{3}(i-1)N_i]!}{N_1!N_2!N_3![M - \sum_{i=1}^{3} iN_i]!}. \tag{9}$$

From Eqs. (3,4,9), the molar configurational entropy $s = S/M$ can be expressed as

$$\begin{aligned} \frac{s(\theta_1,\theta_2,\theta_3,T)}{k_B} &= \left(1 - \frac{\theta_2}{2} - \frac{2\theta_3}{3}\right)\ln\left(1 - \frac{\theta_2}{2} - \frac{2\theta_3}{3}\right) \\ &- \theta_1\ln(\theta_1) - \frac{\theta_2}{2}\ln\left(\frac{\theta_2}{2}\right) - \frac{\theta_3}{3}\ln\left(\frac{\theta_3}{3}\right) \\ &- (1 - \theta_1 - \theta_2 - \theta_3)\ln(1 - \theta_1 - \theta_2 - \theta_3) \end{aligned} \tag{10}$$

where $\theta_i = iN_i/M$ represents the partial coverage of the species S_i. In addition, $\theta_T = \sum_{i=1}^{3}\theta_i$, being θ_T the total surface coverage.

On the other hand, the $\mu_{j,ads}$'s can be obtained from Eqs. (3,5,9),

$$\beta\mu_{1,ads} = \ln\theta_1 - \ln(1 - \theta_1 - \theta_2 - \theta_3) + \beta\varepsilon_1, \tag{11}$$

$$\beta\mu_{2,ads} = \ln\left(1 - \frac{1}{2}\theta_2 - \frac{2}{3}\theta_3\right) + \ln\frac{\theta_2}{2} - 2\ln(1 - \theta_1 - \theta_2 - \theta_3) + \beta\varepsilon_2, \tag{12}$$

$$\beta\mu_{3,ads} = 2\ln\left(1 - \frac{1}{2}\theta_2 - \frac{2}{3}\theta_3\right) + \ln\frac{\theta_3}{3} - 3\ln(1 - \theta_1 - \theta_2 - \theta_3) + \beta\varepsilon_3 \tag{13}$$

Now, equating Eqs. (11-13) with Eq. (7), it results

$$K_1 C_P = \frac{\theta_1}{1 - \theta_1 - \theta_2 - \theta_3}, \tag{14}$$

$$K_2 C_P = \frac{\left(1 - \frac{\theta_2}{2} - \frac{2\theta_3}{3}\right)\frac{\theta_2}{2}}{(1 - \theta_1 - \theta_2 - \theta_3)^2}, \tag{15}$$

$$K_3 C_P = \frac{\left(1 - \frac{\theta_2}{2} - \frac{2\theta_3}{3}\right)^2\frac{\theta_3}{3}}{(1 - \theta_1 - \theta_2 - \theta_3)^3}, \tag{16}$$

where $K_i = \exp[\beta(\mu^0 - \varepsilon_i)]$ is the equilibrium binding constant between the protein in the solution and the molecule adsorbed in the S_i state.

Equations (14-16) represent the partial adsorption isotherms corresponding to three-domain AFPs adsorbed on a 1D lattice. There is one equation for each adsorption state. Following, these equations will be generalized to two-dimensional lattices.

2.2 Two-dimensional square lattices

Hereafter, we address the calculation of approximated thermodynamical functions of complex molecules adsorbed on lattices with connectivity c higher than 2 (i.e., dimensions higher than one).

In general, the number of states $\Omega(N,M)$ for a system of N molecules on M sites will be a function of the lattice connectivity; henceforth $\Omega_c(N,M)$. This quantity can be calculated considering that the molecules are distributed completely at random on the lattice, and assuming the arguments given by different authors[36–39] to relate the configurational factor $\Omega_c(N,M)$ for any c, with the same quantity in one dimension $\Omega_{c=2}(N,M)$,

$$\Omega_c(N,M) \approx m(c,k)^N \Omega_{c=2}(N,M), \tag{17}$$

where $m(c,k)$ represents the degrees of freedom gained by an adsorbed molecule as the dimensionality of the space increases. $m(c,k)$ is, in general, a function of the connectivity and the size/shape of the adsorbate.

For the case of three-domain molecules (modeled as trimers) on a lattice with connectivity c, Eq. (9) can be generalized as follows:

$$\Omega_c(M,N_1,N_2,N_3) \approx \prod_{j=1}^{3} m_j(c,j)^{N_j}\ \Omega_{c=2}(M,N_1,N_2,N_3), \tag{18}$$

where $\Omega_{c=2}(M,N_1,N_2,N_3)$ can be obtained from Eq. (9) and $m_j(c,j)$ represents the number of available configurations for a molecule in the S_j state at zero coverage.

From Eqs. (3,4,18), the molar configurational entropy $s_c = S_c/M$ results

View Article Online
DOI: 10.1039/C7CP06618J

$$\frac{s_c(\theta_1,\theta_2,\theta_3,T)}{k_B} = \theta_1 \ln[m_1(c,1)] + \frac{\theta_2}{2}\ln[m_2(c,2)] + \frac{\theta_3}{3}\ln[m_3(c,3)] + \left(1-\frac{\theta_2}{2}-\frac{2\theta_3}{3}\right)\ln\left(1-\frac{\theta_2}{2}-\frac{2\theta_3}{3}\right) - \theta_1\ln(\theta_1) - \frac{\theta_2}{2}\ln\left(\frac{\theta_2}{2}\right) - \frac{\theta_3}{3}\ln\left(\frac{\theta_3}{3}\right) - (1-\theta_1-\theta_2-\theta_3)\ln(1-\theta_1-\theta_2-\theta_3) \quad (19)$$

Operating as in Subsec. 2.1 [see Eqs. (11-16)], the partial adsorption isotherms corresponding to trimers adsorbed on a lattice with connectivity c can obtained are as follows:

$$m_1(c,1)K_1C_P = \frac{\theta_1}{1-\theta_1-\theta_2-\theta_3}, \quad (20)$$

$$m_2(c,2)K_2C_P = \frac{\left(1-\frac{\theta_2}{2}-\frac{2\theta_3}{3}\right)\frac{\theta_2}{2}}{(1-\theta_1-\theta_2-\theta_3)^2}, \quad (21)$$

$$m_3(c,3)K_3C_P = \frac{\left(1-\frac{\theta_2}{2}-\frac{2\theta_3}{3}\right)^2\frac{\theta_3}{3}}{(1-\theta_1-\theta_2-\theta_3)^3}. \quad (22)$$

As mentioned in the beginning of this section, we focus on the case of square lattices ($c=4$). In order to rationalize our analysis, three different cases have been considered, according to the different adsorption states involved in the process:

- Case A: Proteins in S_3 state adsorb occupying three consecutive sites along one of the two possible directions of the lattice (see panel A, Fig. 2). Then, $m_1(c=4,1)=1$, $m_2(c=4,2)=2$ and $m_3(c=4,3)=2$.
- Case B: Proteins in S_3 state adsorb in one of the four possible angular forms, as shown in panel B, Fig. 2. Then, $m_1(c=4,1)=1$, $m_2(c=4,2)=2$ and $m_3(c=4,3)=4$.
- Case C: This case is the combination of Case A and Case B, where the proteins in S_3 state can adsorb in both linear and angular forms. Then, $m_1(c=4,1)=1$, $m_2(c=4,2)=2$ and $m_3(c=4,3)=6$.

3 Modified Langmuir Model

A modified Langmuir model (MLM) was developed by Can et. al. to account for the adsorption process of a three-domain protein onto an ice crystal.[33] The three-domain protein was modeled as three identical units (or domains) connected by a flexible linker and three possible adsorption states: with one, two, or three domains bound to the surface. These states are denoted as S_1^*, S_2^* and S_3^*, respectively.

As is expected for a simple Langmuir-type model, the arrangement of the adsorption sites in space is immaterial and, consequently, the MLM does not allow to distinguish between different adsorbate structures or different lattice geometries. For this reason, in this section we will use the asterisk symbol ($*$) to distinguish states without a defined structure S_j^*'s, from real states S_j's. In the following, we will reproduce the calculations of Can and Holland[33].

In the MLM, the surface coverage is described at equilibrium using the rate constants for adsorption and desorption of the individual domains as was introduced previously in Ref.[30]. This modified Langmuir approach assumes monolayer adsorption and that each adsorbed domain occupies an equivalent surface area (adsorption site). Under these considerations, the kinetic equations for S_1^*, S_2^* and S_3^* states are

$$\frac{d\theta_1}{dt} = k_{a1}C_P(1-\theta_1-\theta_2-\theta_3) - k_d\theta_1 - k_{a2}\theta_1(1-\theta_1-\theta_2-\theta_3) + k_d\theta_2 \quad (23)$$

$$\frac{d\theta_2}{dt} = 2k_{a2}\theta_1(1-\theta_1-\theta_2-\theta_3) - 2k_d\theta_2 - 2k_{a3}\theta_2(1-\theta_1-\theta_2-\theta_3) + 2k_d\theta_3 \quad (24)$$

$$\frac{d\theta_3}{dt} = 3k_{a3}\theta_2(1-\theta_1-\theta_2-\theta_3) - 3k_d\theta_3 \quad (25)$$

where θ_1, θ_2, and θ_3 are the surface coverage values for S_1^* state, S_2^* state, and S_3^* state, respectively. k_{a1} is the adsorption rate constant from solution to S_1^* state; k_{a2} is the adsorption rate constant from S_1^* state to S_2^* state; and k_{a3} is the adsorption rate constant from S_2^* state to S_3^* state. In addition, it is assumed that the desorption rate does not depend on the number of domains bound to the surface. Accordingly, the desorption rate constants for the domains in each state were taken as k_d. Interested readers are referred to Refs.[30,33] for a more complete description of Eqs. (23-25).

By equating to zero Eqs. (23-25) (equilibrium conditions), θ_1, θ_2, and θ_3 can be obtained:

$$\theta_1 = K_{s1}C_P(1-\theta_1-\theta_2-\theta_3), \quad (26)$$

$$\theta_2 = K_{12}\theta_1(1-\theta_1-\theta_2-\theta_3), \quad (27)$$

$$\theta_3 = K_{23}\theta_2(1-\theta_1-\theta_2-\theta_3), \quad (28)$$

where $K_{s1}=k_{a1}/k_d$ ($K_{12}=k_{a2}/k_d$) [$K_{23}=k_{a3}/k_d$] is the equilibrium binding constant between solution and S_1^* state (S_1^* state and S_2^* state) [S_2^* state and S_3^* state]. Finally,

$$\theta_1 = K_{s1}C_P(1-A), \quad (29)$$

View Article Online
DOI: 10.1039/C7CP06618J

$$\theta_2 = K_{s1}K_{12}C_P(1-A)^2, \tag{30}$$

and

$$\theta_3 = K_{s1}K_{12}K_{23}C_P(1-A)^3, \tag{31}$$

where

$$A = 1 + \frac{1}{a} - \sqrt[3]{\frac{-q+\sqrt{q^2+\frac{4p^3}{27}}}{2}} - \sqrt[3]{\frac{-q-\sqrt{q^2+\frac{4p^3}{27}}}{2}}, \tag{32}$$

with a, p and q being:

$$a = \frac{1}{K_{23}}; \tag{33}$$

$$p = \frac{K_{s1}C_P+1}{K_{s1}K_{12}K_{23}C_P} - \frac{1}{3K_{23}^2}; \tag{34}$$

$$q = \frac{1}{27K_{23}^3} - \frac{K_{s1}C+1}{81K_{s1}K_{12}K_{23}^2C_P} - \frac{1}{K_{s1}K_{12}K_{23}C_P}. \tag{35}$$

Equations (29-31) were numerically solved through a standard computing procedure.

4 Monte Carlo simulation details

The theoretical predictions for the adsorption of three-domain AFPs on an ice surface are compared with MC computational simulations in the grand canonical ensemble. As mentioned in Section 2, the ice surface is represented by an L-one-dimensional lattice or an $L\times L$-square lattice of M equivalent adsorption sites and periodic boundary conditions. The protein is modeled as a trimer with one, two or three domains adsorbed on the surface, each domain occupying one lattice site, see Fig. 1. The corresponding adsorption energies are ε_1, ε_2, and ε_3.

The ice surface is in contact with a protein solution at temperature T and concentration C_P. Then, the adsorption process is simulated, according to the following elementary Monte Carlo step (MCS):

1. A lattice site i is chosen at random.
2. If the site i is empty, one of the three possible adsorption states (S_1, S_2 and S_3) is selected with equal probability $\rho = 1/3$.
   - 2.1. If the selected state (in the incise 2) is S_1, then an attempt is made to adsorb a protein in an upright orientation with probability
   $$\min\left\{1, C_P\exp\left(-\frac{\Delta E}{k_BT}\right)\frac{1}{\rho}\right\}, \tag{36}$$
   where ΔE is the difference between the energies of the final (new) and initial (old) states, $\Delta E = \varepsilon_1$.
   - 2.2. If the selected state (in the incise 2) is S_2, a site j is randomly chosen among the c nearest neighbours of the site i.
     - 2.2.1. If the site j is empty, then an attempt is made to adsorb a protein with probability according to Eq. (36) and $\Delta E = \varepsilon_2$.
     - 2.2.2. If the site j is occupied, then the attempt is rejected.
   - 2.3. If the selected state (in the incise 2) is S_3 and $c = 2$, then two consecutive sites j and k are randomly chosen. In the case of square lattices, three different situations can occur according to the studied case: Case A, then j and k are randomly chosen along one of the two possible lattice directions (see panel A, Fig. 2); Case B, then j and k are randomly chosen among one of the four possible angular forms (see panel B, Fig. 2); and Case C, then j and k are randomly chosen among one of the six possible linear and angular forms.
     - 2.3.1. If the sites j and k are empty, then an attempt is made to adsorb a protein with probability according to Eq. (36) and $\Delta E = \varepsilon_3$.
     - 2.3.2. Otherwise, the attempt is rejected.
3. If the site i is occupied by a protein molecule, we try to desorb.
   - 3.1. If the selected molecule (in the incise 3) is in S_1 state, then an attempt to desorb with probability
   $$\min\left\{1, \frac{1}{C_P}\exp\left(-\frac{\Delta E}{k_BT}\right)\rho\right\}, \tag{37}$$
   where $\Delta E = -\varepsilon_1$.
   - 3.2. If the adsorbed domain in the site i belongs to a protein molecule in S_2 state, then a site j is randomly chosen among the c nearest neighbours of the site i.
     - 3.2.1. If the sites i and j are occupied by segments belonging to the same protein molecule, then an attempt is made to desorb the molecule with probability depicted in Eq. (37) and $\Delta E = -\varepsilon_2$.
     - 3.2.2. Otherwise, the attempt is rejected.
   - 3.3. If the adsorbed domain in the site i belongs to a protein molecule adsorbed in S_3 state and $c = 2$, then two consecutive sites j and k are randomly chosen. In the case of square lattices, three different situations can occur according to the studied case: Case A, then j and k are randomly chosen along one of the two possible lattice directions; Case B, then j and k are randomly chosen among one of the four possible angular forms; and Case C, then j and k are randomly chosen among one of the six possible linear and angular forms.
     - 3.3.1. If the sites i, j and k are occupied by domains belonging to the same molecule, then an attempt is made to desorb with probability according to Eq. (37) and $\Delta E = -\varepsilon_3$.
     - 3.3.2. Otherwise, the attempt is rejected.

View Article Online
DOI: 10.1039/C7CP06618J

4. Repeat from step 1. M times.

For simplicity, the elementary MCS takes into account only the adsorption or desorption of protein molecules, and does not include transitions between different adsorption states. The inclusion of transitions between different states of the adsorbed proteins does not affect the equilibrium properties of the system, and would slow down the simulations. Interested readers are referred to Ref.[32] for a more complete discussion on this topic.

In our MC simulations, the equilibrium state can be well reproduced after discarding the first $m_0 = 10^6$ MCS. Then, averages are taken over $m = 10^6$ MCS successive configurations. The initial configuration of the system is an empty lattice, and the final configuration obtained for a given concentration is used as the initial configuration for the next (higher) concentration.

Thermodynamic quantities, such as total and partial isotherms and adsorption energy per site $u = E/M$, are obtained as simple averages

$$\theta_1 = \frac{\langle N_1 \rangle}{M}, \tag{38}$$

$$\theta_2 = \frac{2\langle N_2 \rangle}{M}, \tag{39}$$

$$\theta_3 = \frac{3\langle N_3 \rangle}{M}, \tag{40}$$

$$\theta_T = \theta_1 + \theta_2 + \theta_3 = \frac{\langle N_1 \rangle + 2\langle N_2 \rangle + 3\langle N_3 \rangle}{M}, \tag{41}$$

and

$$u = \frac{\varepsilon_1 \langle N_1 \rangle + \varepsilon_2 \langle N_2 \rangle + \varepsilon_3 \langle N_3 \rangle}{M}, \tag{42}$$

where $\langle \dots \rangle$ means the average over the MC simulation runs.

The Helmholtz free energy per site $f = F/M$ is calculated by using the well-known thermodynamic integration method[40]. The method in the grand canonical ensemble relies upon integration of the chemical potential μ on coverage along a reversible path between an arbitrary reference state and the desired state of the system. This calculation also requires the knowledge of the Helmholtz free energy per site in the reference state f_0. Thus,

$$f = f_0 + \frac{\int_0^N \mu dN}{M}. \tag{43}$$

The determination of f_0 is trivial [$f_0 = F(M, N_1 = 0, N_2 = 0, N_3 = 0, T)/M = 0$]. Finally, the entropy per site s is calculated as the difference between internal and free energy[35]:

$$\frac{s}{k_B} = \frac{u}{k_B T} - \frac{f}{k_B T}. \tag{44}$$

5 Results and discussion

We will consider first the adsorption on a 1D system, since the LGMMAS theory give exact results for this geometry, and then we will address the adsorption problem in the two-dimensional square lattice. In addition, in this work we consider the case in which the adsorption energies are accumulative and directly proportional to the amount of adsorbed domains is considered: $\varepsilon_2 = 2\varepsilon_1$, and $\varepsilon_3 = 3\varepsilon_1$.

Figure 3a) shows the total and partial adsorption isotherms for a 1D lattice with $\varepsilon_1/k_BT = -2$, $\varepsilon_2/k_BT = -4$ and $\varepsilon_3/k_BT = -6$.

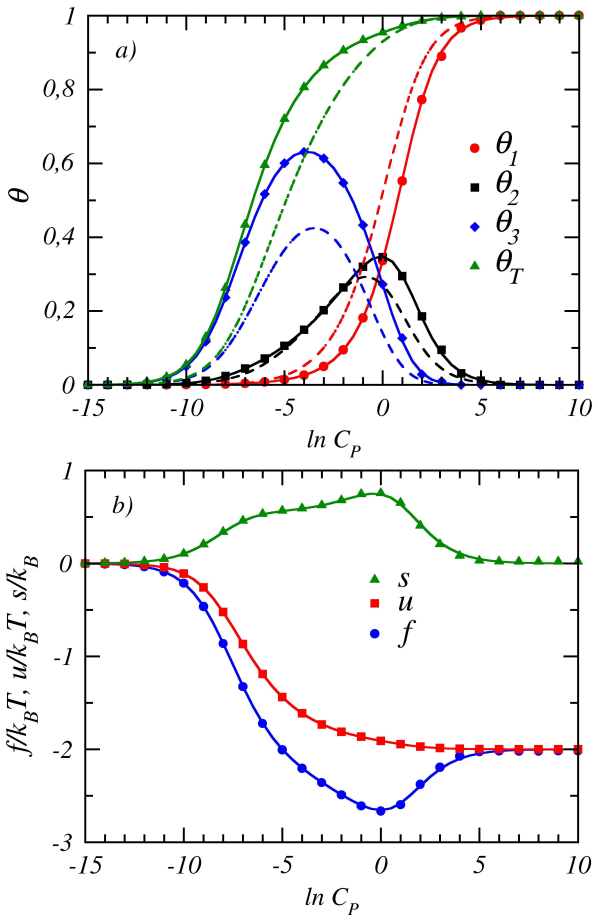

Fig. 3 a) Total and partial adsorption isotherms Antifreeze trimer adsorbed on a 1D lattice with $\varepsilon_1/k_BT = -2$, $\varepsilon_2/k_BT = -4$ and $\varepsilon_3/k_BT = -6$. Symbols, solid lines and dashed lines represent Monte Carlo simulation, LGMMAS [Eqs. (14-16)] and MLM [Eqs. (29-31)] data, respectively. b) Adsorption energy per site (in k_BT units), configurational entropy per site (in k_B units) and Helmholtz free energy per site (in k_BT) for the case studied in Fig. 3a).

Symbols represent simulation data for L-lattices with $L = 1200$ and periodic boundary conditions. Solid and dashed lines correspond to theoretical results from LGMMAS [Eqs. (14-16)] and MLM [Eqs. (29-31)], respectively.

The adsorption process starts at very low bulk protein concentration ($\ln C_P \approx -12$). Under these conditions, the protein is adsorbed at the most energetically stable state (S_3 state). The increase in the bulk protein concentration is followed by an increase in the amount of protein adsorbed in S_3 state, reaching a maximum coverage of 0.62 at ($\ln C_P \approx -4.5$). Then, S_3 coverage decreases following a symmetric profile. The S_2 and S_1 partial densities start to increase at $\ln C_P \approx -9$ and $\ln C_P = -6$, respectively. The population in S_2 state reaches a maximum of 0.35 at $\ln C_P \approx -0.5$ and then decays. On the contrary, the amount of protein in S_1 state increases monotonically until the surface is completely covered at high bulk protein concentrations. The displacement of molecules in S_2 and S_3 states by the molecules in $S1$ state is called adsorption preference reversal (APR) phenomenon, and it has been previously observed[41,42]. The behavior of the total surface coverage can be understood from the sum of the indi-

View Article Online
DOI: 10.1039/C7CP06618J

vidual behaviors for S_1, S_2 and S_3.

In Fig. 3b), the adsorption energy per site u, configurational entropy per site s, and Helmholtz free energy per site f were calculated as a function of the protein concentration C_P. The energy parameters are the same as in Fig. 3a). Symbols and solid lines represent MC simulation and LGMMAS data, respectively. The calculations with MLM theory were limited to the total and partial adsorption isotherms.[33] Consequently, MLM curves are not included in the thermodynamic analysis shown in Fig. 3b). The total adsorption energy per site decreases monotonically as C_P increases, tending asymptotically to $u/k_BT = \varepsilon_1/k_BT$ for higher concentrations. In this limit, the lattice is completely filled by proteins in the S_1 state.

Regarding the configurational entropy per site, the overall behavior can be explained as follows. For low protein concentrations ($\ln C_P < -15$), the surface is empty and the entropy is zero. For concentrations higher than $\ln C_P \approx -15$, proteins in S_1, S_2 and S_3 states begin to adsorb, and the configurational entropy per site is an increasing function of C_P. This process continues until $\theta_1 \approx \theta_2 \approx \theta_3 \approx 1/3$ [see Fig. 3a)], the mixture effect is maximum, and the entropy shows a peak around $\ln C_P \approx 0$. Then, s/kB decreases monotonically to zero for high values of C_P, where the whole surface is covered by proteins in the S_1 state. This limit value of entropy is zero because there is only one available configuration at high concentrations. The behavior of the Helmholtz free energy per site can be understood from the analysis of the curves of u and s.

The LGMMAS and MLM results can be quantitatively evaluated with respect to MC results using the percentage relative discrepancy, D_{θ_j}, defined as:

$$D_{\theta_j}(\%) = 100\frac{\int \left|\theta_j^{MC}(\ln C_P) - \theta_j^{th}(\ln C_P)\right| d\ln C_P}{\int \theta_j^{MC}(\ln C_P)\, d\ln C_P} \qquad \{j = 1,2,3,T\} \tag{45}$$

where θ_j^{MC} and θ_j^{th} represent the coverage obtained by using MC simulation and theoretical approximations (LGMMAS or MLM theories) respectively. Each pair of values (θ_j^{MC}, θ_j^{th}) is obtained at fixed $\ln C_P$.

Table 1 Percentage relative discrepancy [Eq. (45)] for the theoretical curves in Fig. 3a).

	$D_{\theta_j}(\%)$ 1D	
Coverage	LGMMAS	MLM
θ_1	0.4	9.1
θ_2	0.2	19.0
θ_3	0.1	42.3
θ_T	0.01	9.2

The values of $D_{\theta_j}(\%)$ for the 1D case [Fig. 3a)] are compiled in Table 1. The results obtained with LGMMAS theory are exact, then the small discrepancy values reported in Table 1 (less than 0.5%) correspond to MC simulation errors. However, significant discrepancy are observed for MLM results, especially for the partial coverage corresponding to S_2 state [$D_{\theta_2}(\%) \approx 19\%$] and S_3 state [$D_{\theta_3}(\%) \approx 42\%$]. This finding indicates that MLM has serious limitations for studying the adsorption of n-domains molecules with $n > 2$. As discussed below, these limitations are even more evident in 2D systems.

Hereafter, we present the analysis of the adsorption of proteins on 2D substrates and Cases A, B and C. For this purpose, square lattices of $M = 120 \times 120$ sites and periodic boundary conditions were simulated. With this lattice size we verified that finite size effects are negligible. As in Fig. 3, we set $\varepsilon_1/k_BT = -2$, $\varepsilon_2/k_BT = -4$ and $\varepsilon_3/k_BT = -6$.

Figure 4 shows the total and partial adsorption isotherms for a 2D-square lattice for Cases A, B and C, as indicated. The results are qualitatively similar to those reported for the one-dimensional system [Fig. 3a)]. However, some details may be pointed out. When the protein is adsorbed in the S_3 state occupying three consecutive sites along one of the two possible directions of the lattice [Case A, Fig. 2a)] the maximum coverage reaches a value of $\theta_3 \approx 0.65$ [Fig. 4a)]. If the protein is adsorbed in angular form [Case B, Fig. 2b)], the coverage reaches a maximum of $\theta_3 \approx 0.75$ see Fig. 4b). Finally, when the protein is adsorbed in both linear and angular form (Case C), a maximum coverage value of $\theta_3 \approx 0.8$ is reached [Fig. 4c)]. This tendency reflects the number of conformations in which the protein may be adsorbed in the S_3 state; in Case A, B and C there are two (Panel A, Fig 2), four (Panel B, ref) and six (Panel A and B, ref) possible configurations respectively for the AFP. It is important to note that the maximum values mentioned above were obtained at a similar protein concentration ($\ln C_P \approx -4$) for the studied cases.

Analyzing now the partial coverage of the protein in the S_2 state, the maximum values obtained for θ_2 are $\theta_2 \approx 0.43$, $\theta_2 \approx 0.35$ and $\theta_2 \approx 0.31$ (at $\ln C_P \approx 0$) for Cases A, B and C, respectively. An opposite tendency is shown by these values with respect to the maximum values in the S_3 state. This behavior is due to the competition between the proteins to adsorb in the S_2 and S_3 states. In fact, since the number of forms to adsorb the molecules in the S_2 state is the same for Cases A, B and C, then, the increasing of the number of possible configurations for the molecules in the S_3 state gives rise to a decrease in θ_2.

Finally, when the protein is adsorbed in the S_1 state, the coverage profile is similar in all cases; specifically, it prefers to adsorb in the S_1 state starting from $\ln C_P \approx -5$, increasing monotonically with C_P, and fully covering the lattice, displacing the S_2 and S_3 states of the protein.

The evaluation of the performance of LGMMAS and MLM is carried out using the discrepancy criteria [Eq. 45]. The results are shown in Table 2.

Table 2 Percentage relative discrepancy [Eq. (45)] for the theoretical curves in Fig. 4.

	Case A		Case B		Case C	
Coverage	LGMMAS	MLM	LGMMAS	MLM	LGMMAS	MLM
θ_1	1.3	15.5	2.3	13.8	2.4	16.3
θ_2	0.9	39.9	2.3	25.2	3.4	13.9
θ_3	9.5	42.8	12.2	54.0	11.5	59.4
θ_T	1.8	11.4	2.6	14.5	2.7	15.7

Significant differences between the values of $D_{\theta_j}(\%)$ for LGM-

View Article Online
DOI: 10.1039/C7CP06618J

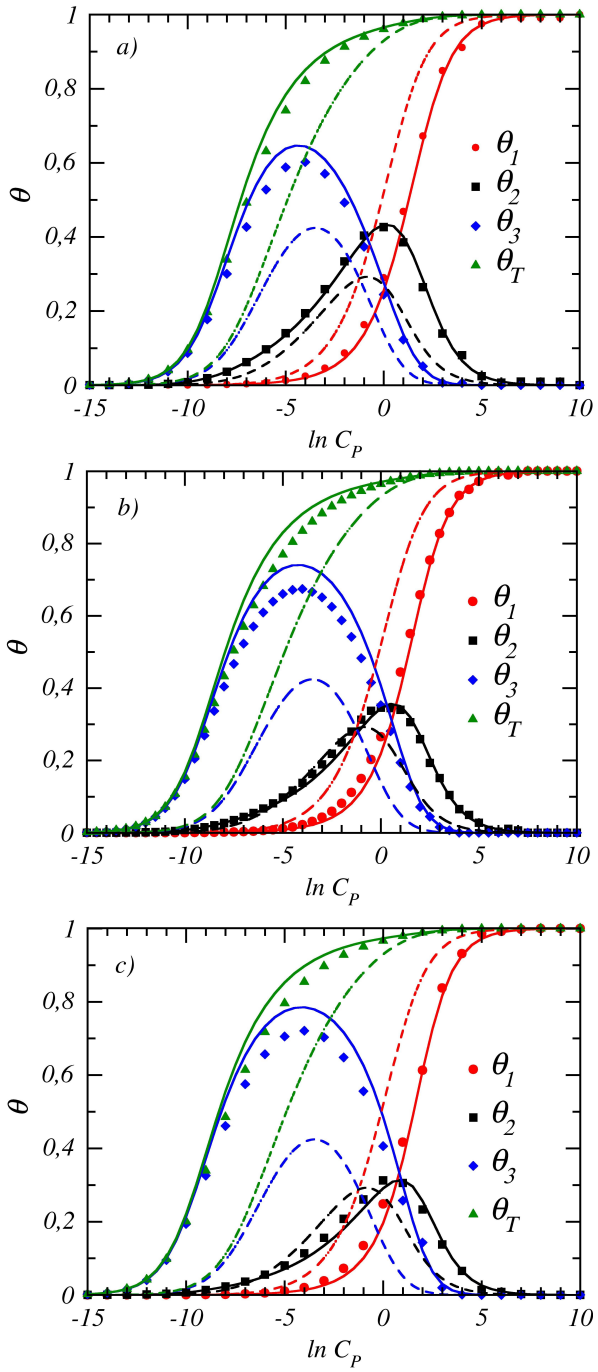

Fig. 4 Total and partial adsorption isotherms (coverage vs $\ln C_P$) for three-domain AFPs adsorbed on square lattices with $\varepsilon_1/k_BT = -2$, $\varepsilon_2/k_BT = -4$ and $\varepsilon_3/k_BT = -6$. Symbols, solid lines and dashed lines represent simulation, LGMMAS [Eqs. (20-22)] and MLM [Eqs. (29-31)] data, respectively. a) Case A: $m_1(c=4,1)=1$, $m_2(c=4,2)=2$ and $m_3(c=4,3)=2$. b) Case B: $m_1(c=4,1)=1$, $m_2(c=4,2)=2$ and $m_3(c=4,3)=4$. c) Case C: $m_1(c=4,1)=1$, $m_2(c=4,2)=2$ and $m_3(c=4,3)=6$.

MAS and MLM are observed. Regarding to LGMMAS theory, $D_{\theta_j}(\%)$ varies between $\approx 1\%$ and $\approx 12\%$. These values are of the order or smaller than the expected experimental errors. In addition, from a simple inspection of Table 2, it is observed that the values of $D_{\theta_j}(\%)$ for LGMMAS and Case A are smaller than those corresponding to Cases B and C. This better performance for Case A is due to the proteins have two possible lineal conformations in the S_3 state, which is closest to the exact 1D solution*. Cases B and C have four and six possible conformations for S_3 state, respectively.

Very different is the situation in the case of MLM, where $D_{\theta_j}(\%)$ ranges between $\approx 10\%$ and $\approx 60\%$. MLM does not distinguish between different adsorption states and, consequently, the accuracy of the model diminishes as the number of possible conformations of the adsorbed proteins increases. Such discrepancies had not been observed in previous studies[30], where two-domains AFPs were modeled as simple dimers with $S_1 = 1$ and $S_2 = 2$. This finding reinforces the concept that MLM has serious limitations for studying the adsorption of n-domains molecules with $n > 2$.

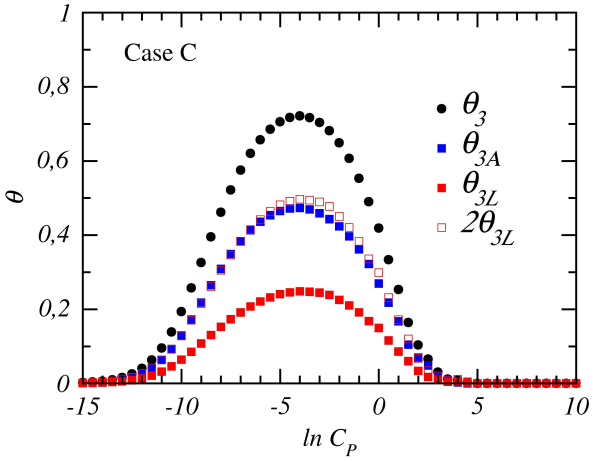

Fig. 5 Relative proportion of each type of protein conformation (linear or angular) in the S_3 state for the case studied in Fig. 4c). Solid (open) symbols correspond to θ_{3A} ($2\theta_{3L}$).

Concerning the adsorption of molecules with $m_3(c=4,3)=6$ (Case C), simple statistical arguments can be used to obtain the relative proportion of each type of protein conformation (linear or angular) in the S_3 state. In fact, linear proteins have two possible adsorption states on a square lattice (panel A, Fig. 2), while angular proteins have four possible adsorption states (panel B, Fig. 2). Accordingly, and given that all the conformations have the same probability, it is expected that the population of proteins adsorbed in an angular conformation is twice the amount of adsorbed proteins in a linear conformation,

$$\theta_{3A} \approx 2\theta_{3L}. \tag{46}$$

In addition, $\theta_3 = \theta_{3L} + \theta_{3A}$, where θ_{3L} (θ_{3A}) denotes the partial coverage corresponding to molecules adsorbed in the S_3 state and

* If the adsorption of molecules in S_2 and S_3 states is restricted to only one lattice direction, the problem reduces to one dimension and the solution is exact.

View Article Online
DOI: 10.1039/C7CP06618J

linear (angular) conformation.

In order to verify the proposed relationship between θ_{3L} and θ_{3A}, these quantities were calculated by MC simulations for the case in Fig. 4c). The results are shown in Fig. 5. As it can be observed, the curve corresponding to θ_{3A} (solid symbols) shows an excellent agreement with the one corresponding to $2\theta_{3L}$ (open symbols), validating the expression in Eq. (46).

Continuing with the analysis of the 2D case, the energy per site, configurational entropy per site and Helmholtz free energy per site were calculated as a function of the bulk protein concentration for the same cases in Fig. 4. The results are shown in Fig. 6: a) Case A; b) Case B and c) Case C.

The energy per site is very similar to that in 1D [Fig. 3b)] because the adsorption energy is the same, and there are not lateral interactions between particles. The entropy is the amount in which one can observe appreciable differences between 1D and 2D. In 2D the entropy is higher than in 1D due to the number of available states increases when the connectivity increases. The entropy increases also with the number of possible adsorption states, and small differences are observed between the curves corresponding to Cases A, B and C. Finally, both $c = 2$ and $c = 4$ have the same limit in the configurational entropy per site, because at high bulk protein concentration there are just molecules in S_1 state (monomers) on the lattice. The degeneracy of this structure is equal to one at full coverage, and consequently, $s(\theta_T \to 1)/k_B \to 0$.

Finally, operating with Eqs. (20-22), the following relationship between the partial densities θ_1, θ_2 and θ_3 can be obtained:

$$m(c,3)K_3{\theta_2}^2 = \frac{4\theta_1\theta_3(K_2)^2[m(c,2)]^2}{3K_1 m(c,1)}. \qquad (47)$$

In this paper, we restrict the treatment to the additive case, where $\varepsilon_2 = 2\varepsilon_1$ and $\varepsilon_3 = 3\varepsilon_1$. Then, $K_1 = \exp[\beta(\mu^0 - \varepsilon_1)]$, $K_2 = \exp[\beta(\mu^0 - 2\varepsilon_1)]$, $K_3 = \exp[\beta(\mu^0 - 3\varepsilon_1)]$, and Eq. (47) reduces to

$$m(c,3){\theta_2}^2 = \frac{4\theta_1\theta_3[m(c,2)]^2}{3m(c,1)}. \qquad (48)$$

Now, considering the values of $m(c,i)$ for Cases A, B and C, the following expressions can be obtained:

$$\theta_3 = \frac{3}{8}\frac{{\theta_2}^2}{\theta_1} \qquad \text{Case A,} \qquad (49)$$

$$\theta_3 = \frac{3}{4}\frac{{\theta_2}^2}{\theta_1} \qquad \text{Case B,} \qquad (50)$$

and

$$\theta_3 = \frac{9}{8}\frac{{\theta_2}^2}{\theta_1} \qquad \text{Case C.} \qquad (51)$$

Equations (49-51) account for the main features of the process of adsorption of three-domain molecules.

Figure 7 shows the profiles for θ_3 (solid symbols) obtained from MC simulation for the Cases A, B, and C, which are equivalent to the ones in Fig. 4a), 4b), and 4c), respectively. In the figure are also shown the values of θ_3 (open symbols) calculated using Eqs. (49-51) (where θ_1 and θ_2 correspond to the values in Fig. 4). These results validate the Eqs. (49-51), which demonstrate

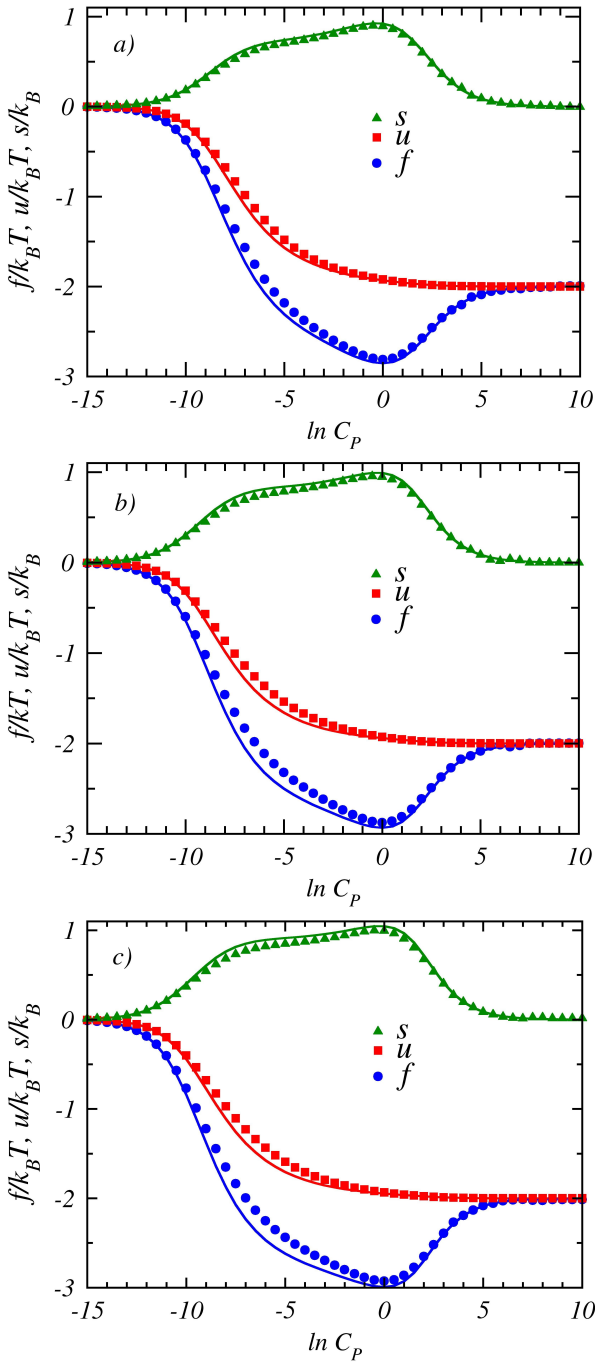

Fig. 6 Adsorption energy per site (in k_BT units), configurational entropy per site (in k_B units) and Helmholtz free energy per site (in k_BT) for a) Case A, b) Case B, and c) Case C.

View Article Online
DOI: 10.1039/C7CP06618J

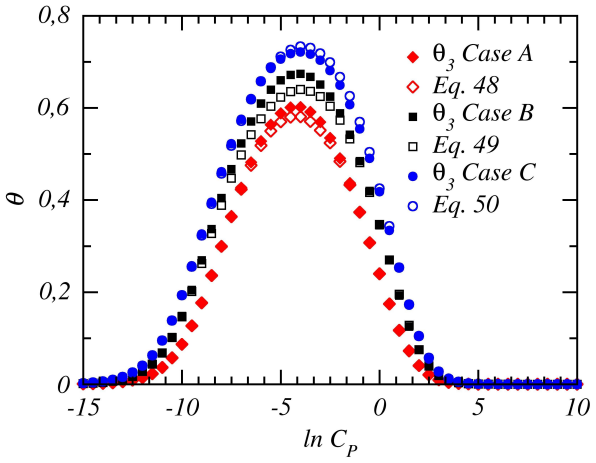

Fig. 7 Coverage degree θ_3 (solid symbols) as a function of C_P obtained from MC simulation for the cases A, B, and C. θ_3 (open symbols) obtained using θ_1 and θ_2 from MC simulation and Eqs. (49-51)

the consistency and robustness of the expressions derived from LGMMAS.

The study presented in this section indicates that the LGMMAS is, to the best of our knowledge, the most accurate approximation to the problem of adsorption of three-domain antifreeze proteins. The theory allows to identify different conformations of the molecules in the adsorbed state, and provides extremely accurate relationships between the densities of each adsorption state.

6 Conclusions

In the present work, a new theoretical approach to the adsorption of three-domain antifreeze proteins on ice has been presented on the basis of a generalization of the lattice-gas model of molecules with multiple adsorption states[31]. The statistical theory, obtained from the exact partition function of non-interacting trimers adsorbed in one dimension and its extension to two dimensions, takes into account not just the protein topology, but also how it adsorbs on the lattice.

The three-domain protein was modeled as three identical units (or domains) connected by a flexible linker and three possible adsorption states: S_1, S_2 and S_3, with one, two, or three domains bound to the surface, respectively. In addition, four different cases were considered, according to the different configurations of the proteins in the S_3 state: (*i*) the molecules adsorb occupying three consecutive sites on a 1D lattice; (*ii*) the molecules adsorb occupying three consecutive sites along one of the two possible directions of a 2D square lattice; (*iii*) the molecules adsorb in one of the four possible angular forms on a 2D square lattice; and (*iv*) the molecules adsorb in both linear and angular forms.

The results of the theory were contrasted with MC simulations, and a modified Langmuir model developed by Can and Holland[33], where the arrangement of the adsorption sites in space is immaterial. While the new scheme provides exact results in one-dimensional lattices and offers a very accurate description in two dimensions, the MLM does not distinguish between different adsorption states, and shows severe discrepancies with the two-dimensional simulation results. Such discrepancies had not been observed in previous studies with simpler molecules (monomers and dimers)[30]. In other words, MLM has serious limitations for studying the adsorption of *n*-domains molecules with $n > 2$.

Summarizing, the theory presented here (1) represents a significant q ualitative a dvance i n o ur u nderstanding o f multistate adsorbed molecules, (2) shows that adsorbate structure and lattice geometry play fundamental roles in determining the statistics of this class of systems, and (3) is the most accurate and complete approximation to this complex problem, and (4) could be very useful in interpreting experimental data. In this line, future efforts will be devoted to (*i*) extend the calculations to larger molecules and non-additive adsorption energies, and (*ii*) analyse the applicability of the model to study protein activity in some well-known systems[27–29].

7 Acknowledgments

This work was supported in part by CONICET (Argentina) under project number PIP 112-201101-00615; Universidad Nacional de San Luis (Argentina) under project 03-0816; Universidad Tecnológica Nacional, Facultad Regional San Rafael (Argentina) under project PID UTN 3542 Disp. 284/12; and the National Agency of Scientific and Technological Promotion (Argentina) under project PICT-2013-1678. The numerical work were done using the BACO parallel cluster (composed by 70 PCs each with an Intel i7-3370 / 2600 processor) located at Instituto de Física Aplicada, Universidad Nacional de San Luis - CONICET, San Luis, Argentina.

References

1 A.L. DeVries and D.E. Wohlschlag, *Science* , 1969, **163**, 1073.
2 G.L. Fletcher, C.L. Hew and P.L. Davies, *.Annu. Rev. Physiol.* , 2001, **63**, 359.
3 J.G. Duman, *Annu. Rev. Physiol.*, 2001, **63**, 327.
4 P.L. Davies, *Trends Biochem. Sci.*, 2014, **39**, 548.
5 H. Nada and Y. Furukawa, *Polym. J.*, 2012, **44**, 690.
6 C.P. Garnham, R.L. Campbell and P.L. Davies, *PNAS*, 2011, **108**, 7363.
7 J.A. Raymond, C. Fritsen and K. Shen, *FEMS Microbiol. Ecol.*, 2007, **61**, 214.
8 J.A. Raymond, B.C. Christner, S.C. Schuster, *Extremophiles*, 2008, **12**, 713.
9 C.P. Garnham, J.A. Gilbert, C.P. Hartman, R.L. Campbell, J. Laybourn-Parry and P.L. Davies, *Biochem. J.*, 2008, **411**, 171.
10 T. Kamijima, M. Sakashita, A. Miura, Y. Nishimiya and S. Tsuda, *PloS ONE*, 2013, **8**, e73643.
11 A. Ideta, Y. Aoyagi, K. Tsuchiya, Y. Nakamura, K. Hayama, A. Shirasawa, K. Sakaguchi, N. Tominaga. Y. Nishimiya and S. Tsuda, *J. Reprod. Dev.*, 2015, **61**, 1.
12 W. Boonsupthip and T.C. Lee, *J. Food Sci.*, 2003, **68**, 1804.
13 C.L. Zhao, S. Porzio, A. Smith, H. Ge, H. Davis and L. Scriven, *J. Coat. Technol. Res.*, 2006, **3**, 109.
14 G. Petzold and J.M. Aguilera, *Food Biophys.*, 2009, **4**, 378.
15 T.A. Berendsen, B.G. Bruinsma, C.F. Puts, N. Saeidi, O.B. Usta, B.E. Uygun, M.L. Izamis, M. Toner, M.L. Yarmush and K. Uygun, *Nat. Med.*, 2014, **20**, 790.

View Article Online
DOI: 10.1039/C7CP06618J

16 R.C. Deller, M. Vatish, D.A Mitchell and M.I. Gibson, *Nat. Commun.*, 2014, **5**, 3244.
17 G. Amir, B. Rubinsky, Y. Kassif, L. Horowitz, A.K. Smolinsky and J. Lavee, *Eur. J. Cardio Thorac. Surg.*, 2003, **24**, 292.
18 A. Regand and H.D. Goff, *J. Food Sci.*, 2005, **70**, E552.
19 J.A. Raymond and A.L. Devries, *Cryobiology*, 1972, **9**, 541.
20 J.A. Raymond and A.L. Devries, *Proc. Natl. Acad. Sci. USA*, 1977, **74**, 2589.
21 S. Venketesh and C. Dayananda, *Critical Rev. Biotechnol.*, 2008, **28**, 57.
22 P. Wilson, D. Beaglehole and A. DeVries, *Biophys. J.*, 1993, **64**, 1878.
23 P. Wilson and J. Leader, *Biophys. J.*, 1995, **68**, 2098.
24 E. Kristiansen and K.E. Zachariassen, *Cryobiology*, 2005, **51**, 262.
25 K.A. Sharp, *J. Chem. Phys.*, 2014, **144**, 22D510.
26 H. Nada and Y. Furukawa, *Phys. Chem. Chem. Phys.*, 2011, **13**, 19936.
27 Y. Xu, R. Gnanasekaran and D.M. Leitner, *J. At. Mol. Op. Phys.*, 2012, , 125071.
28 K. Meister, S. Ebbinghaus, Y. Xu, J.G. Duman, A. DeVries, M.
Gruebele, D.M. Leitner, *PNAS*, 2013, **110**, 1617.
29 H.D. Pandey and D.M. Leitner, *J. Phys. Chem. B*, 2017, **121**, 9498.
30 Ö. Can and N.B. Holland, *J. Colloid Interface Sci.*, 2009, **329**, 24.
31 E. Quiroga and A.J. Ramirez-Pastor, *Chem. Phys. Lett.*, 2013, **556**, 330.
32 C.F. Narambuena, F.O. Sanchez Varretti and A.J. Ramirez-Pastor, *Phys. Chem. Chem. Phys.*, 2016, **18**, 24549.
33 Ö. Can and N.B. Holland, *Biochemistry*, 2013, **52**, 8745.
34 N. Kubota and J.W. Mullin, *J. Cryst. Growth*, 1995, **152**, 203.
35 T.L. Hill, *An Introduction to Statistical Thermodynamics*, Addison-Wesley Publishing Company, Reading, MA, 1960.
36 P. Flory, *J. Chem. Phys.*, 1942, **10**, 10.
37 T. Nitta, M. Kuro-Oka and T. Katayama, *J. Chem. Eng. Jpn.*, 1984, **17**, 45.
38 T. Nitta and A.J. Yamaguchi, *J. Chem. Eng. Jpn.*, 1992, **25**, 420.
39 A.J. Ramirez-Pastor, T.P. Eggarter, V. Pereyra and J.L. Riccardo, *Phys. Rev. B*, 1999, **59**, 11027.
40 K. Binder, *Applications of the Monte Carlo Method in Statistical Physics. Topics in Current Physics*, Springer, Berlin, 1984; Vol. 36.
41 A. Khettar, S.E. Jalili, L.J. Dunne, G. Manos and Z. Du, *Chem. Phys. Lett.*, 2002, **362**, 414.
42 L.J. Dunne, G. Manos and Z. Du, *Chem. Phys. Lett.*, 2003, **377**, 551.