Spiers Memorial Lecture: Effect of interaction specificity on the phase behaviour of patchy particles

Nicolas Dorsaz,[a] Laura Filion,[a] Frank Smallenburg[b] and Daan Frenkel[a]

We report a numerical study on the phase behaviour of a 'patch–anti-patch' model for particles with tetrahedrally arranged attractive spots. In particular, we compute the phase equilibria between the fluid and a low density diamond cubic (DC) crystal for different realizations of the patch–anti-patch interaction. By increasing the 'specificity' of the patches, *i.e.* lowering the number of corresponding attractive 'anti-patches' to a given patch, we find that the metastability gap between the DC freezing boundary and the liquid–gas critical point widens considerably. We argue that this effect of interaction specificity is relevant for the description of protein phase diagrams, as patch–anti-patch interactions can stabilise relatively open, ordered structures.

1 Introduction

Growing sufficiently high-quality protein crystals for use in X-ray crystallography is still a bottleneck to structure determination in protein science. Indeed, only a small fraction of all globular proteins crystallize readily and most that do, do so under conditions that are non-physiological: at high salt concentrations or in presence of a high depletant concentration.[1]

The poor crystallizability of proteins is usually attributed to two main causes. First of all, proteins might have directional interactions that are not compatible with simple crystal lattices.[2] In this context, it is interesting to note that the typical space groups of protein crystals are rather different form those of simple molecular crystals. There are 65 space groups compatible with crystal structures of chiral macromolecules, yet one-third of all protein crystals form in a single space group, $P2_12_12_1$. Conversely, many space groups that are common for simple molecular crystals have never been observed in protein crystals. The number of degrees of freedom for packing a low-symmetry molecule in different space groups seems to be the key: proteins crystallize preferentially in space groups where it is easier to achieve connectivity.[3] Secondly, kinetically trapped phases such as aggregates and gels that are frequently seen experimentally in protein suspensions, are enhanced by anisotropic, specific interactions and the formation of such structurally arrested states is likely to suppress the nucleation of protein crystals.[4-6]

Until recently, the search for suitable crystallization conditions was mainly based on trial and error.[7] In the last two decades there have been several attempts to tackle the problem of protein crystallization in a more systematic manner, trying to

[a]*Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge, CB2 1EW, UK*
[b]*Soft Condensed Matter, Deybe Institute for NanoMaterials Science, Utrecht University, Princetonplein 5, 3584CC, The Netherlands*

understand the subtle interplay between protein interactions and their equilibrium phase diagram. In a seminal paper, George and Wilson[8] were able to rationalize the observed crystallization region for proteins: solvent conditions that were known to promote crystallization could be related to a particular range of the second osmotic virial coefficients B_2 of the protein solution. Hence, interactions should be attractive enough to promote crystallization, while not being so large that they result in disordered aggregation.[9] Colloidal models for globular proteins have also revealed that the fluid–fluid coexistence line might become metastable with respect to the solid–fluid boundary for a sufficiently small range of the attraction and that the presence of a metastable fluid–fluid critical point can lower substantially the free-energy barrier for nucleation of a crystalline germ and, thus, indirectly promote crystallization.[10] At the same time, the competition between glassy and crystal phases has been extensively investigated both in systems of purely hard spheres[11] and in systems of particles with short-ranged attraction.[12]

These investigations of the sphero-symmetric potential properties suggest that good knowledge of the equilibrium phase diagram is a prerequisite for materials design or protein crystal predictions. Such information is essential to predict not only if, but also how rapidly, a given crystal structure will assemble. A particular crystal structure may look promising on paper, but such information is irrelevant if it is not stable with respect to the liquid phase in the experimentally accessible temperature-density range. Similarly, even when an optimized interaction has been identified to produce a given crystal structure that is thermodynamically stable, the nucleation rate of the crystal might be so small that effective production of the crystal is impossible. To take a simple example, a suspension of mono-disperse colloids brought to a sufficiently high density will form beautiful, irides-cent crystals but the rate at which these crystals form has been shown to depend strongly on the steepness of the repulsive forces that act between the colloids: charged colloids with a soft, long-ranged repulsion tend for instance to crystallize much faster than hard-sphere colloids at the same supersaturation.[13] Kinetic effects are even more dramatic when it comes to proteins and their highly directional interactions.

Approaches based on orientationally averaged isotropic potentials have enriched our understanding of the general features of protein phase diagrams, and addressed, *via* ad-hoc coarse-grained models, the fluid phase behaviour of solutions of globular proteins more quantitatively.[14–16] However, they fail to predict quantitatively the fluid–solid equilibrium. In fact, isotropic interactions favor densely packed crystal phases of high symmetry such as face-centered cubic (FCC), body-centered cubic (BCC) or hexagonal close-packed (HCP), whereas protein crystals are typically much more open, with around 7 contacts on average per protein.[4] Clearly, where crystal structures are concerned, the patchiness of protein surfaces that is essential for their biological role, leads to strong directionality of the associated interactions that cannot simply be averaged out.

Patchy—or *aeolotopic*—models, which can be used to describe the directional nature of protein interactions have now been studied for a decade.[2,5,17] The first numerical studies of the Kern-Frenkel "patchy" potential have revealed that the fluid–fluid coexistence curve can be shifted to lower temperature by making the patches smaller or by decreasing the number of patches, and that anisotropy can stabilize multiple solid phases.[18–20] Investigations of a slightly different model in which each attractive spot is involved in not more than a single-bonded inter-action have demonstrated that, by diminishing the number of bonded nearest neighbors, it is possible to generate liquid states (*i.e.*, states with temperature T lower than the liquid–gas critical temperature) with a very low, and eventually vanishing, packing fractions; a situation that cannot be realized with spherically interacting particles.[21] The resulting increase of the region of stability of the liquid phase is expected to favor considerably the formation of stable equilibrium gels at low densities.

This journal is © The Royal Society of Chemistry 2012

Here, we study a slightly different class of patchy models in which each patch of a given particle interacts with a subset N_a of the N_p complementary anti-patches on the other particles. The introduction of such 'patch–anti-patch' (p-ap) models[2,20,22] was motivated by studies that compute the different contributions to the protein's osmotic second virial coefficients based on atomistic structural information[23,24] and has been used in other systems, such as supramolecular polymers.[25] The main contribution to the overall protein–protein interaction can be attributed to a small number of complementary, highly attractive configurations. The high-affinity configurations, denoted as 'patch–anti-patch' pairs,[22] result from complementary, opposing surface regions of the proteins in a particular arrangement.[26] Since the highest-energy bonds are formed by patch–anti-patch contacts, they are expected to play a fundamental role in protein aggregation and crystallization.

For a given protein structure, one can construct a coarse-grained patch–anti-patch model starting from atomistic structural information.[22] This approach would make it possible to arrive at a fair prediction of the full phase diagram of that specific protein. However, the objective of the present work is to derive some generic features of the p-ap model—in particular the similarities and differences with the better-known p-p models. We therefore focus on the p-ap version of the tetrahedral Kern–Frenkel patchy-sphere model. The phase diagrams of tetrahedral patchy particles have been determined by computer simulations.[27–29] These studies revealed the existence of an open, low density, diamond cubic (DC) and a diamond hexagonal (DH) crystals were found. The free-energy barriers to nucleate DC and/or DH crystals were calculated for different widths of the tetrahedral patches.

Knowledge of the phase behaviour of this reference system allows us to focus on the effects of higher patch specificity on the phase behaviour. From a geometric viewpoint, a fully bonded DC crystal can be satisfied by the patch-patch and the patch–anti-patch tetrahedral models, therefore both will have the same potential energy at zero temperature. However, from an entropic perspective, the two models are very different: N^2 different bonds can be formed between a pair of particles with N patches, whilst there are only N possible bonds between a pair of patch–anti-patch particles. The question is how highly specific, directional interactions modify the stability of the various crystal phases with respect to the liquid and the vapor.

The remainder of this paper is structured as follows. In Section 2 the tetrahedral patch–anti-patch model is defined. Section 3 gives a brief outline of the different Monte Carlo methods that we used to compute the free energy and phase equilibria of the liquid and the solid phases. In this section, we report the phase diagrams of the patch-patch p-p and the patch–anti-patch p-ap tetrahedral model and discuss the implications of our findings.

2 Model

To explore the effect of the patch specificity on the equilibrium phase behaviour, we consider the Kern–Frenkel model with four patches ($\alpha = 1$ to 4) placed in a tetrahedral arrangement on the surface of the particle as shown in Fig. 1. In addition, every patch p has a label α such that p^α only interacts with p^α. We examine three realizations of this tetrahedral patchy-sphere model:

(a) a 'patch-patch' model (p-p) where $p^\alpha = 1$ for all α and all patches on particle i interact with all patches of particle j (Fig. 1a).

(b) a 'patch–anti-patch' model (p-ap$_1$) where every particle has four identifiable patches labelled $p^\alpha = \alpha$; therefore a patch interacts only with a single ($N_a = 1$) complementary anti-patch on other particles (Fig. 1b).

(c) an intermediate model (p-ap$_2$) where every particle has two patches labelled 1 and two labelled 2 ($p^1 = p^2 = 1$ and $p^3 = p^4 = 2$); hence each patch has $N_a = 2$ complementary anti-patches (Fig. 1c).

In these models, two patchy particles i and j, located at $\mathbf{r}_i$ and $\mathbf{r}_j$ respectively, feel an attraction given by
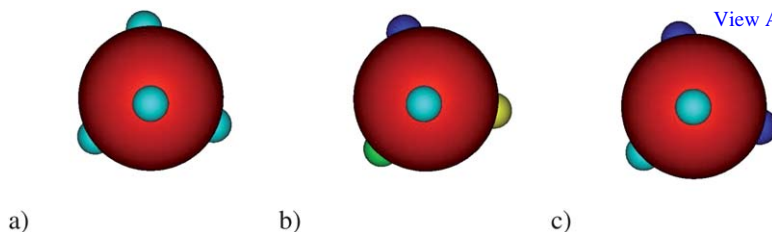
a)   b)   c)

Fig. 1 Schematical view of the three different patchy models studied. Note that the shape of the patches is not precisely the shape of the Kern-Frenkel patches, and is only meant to identify the patch location. A patch on a particle interacts only with a complementary patch of the same color on another particle. (a) usual 'patch-patch' model (p-p) where all 4 patches on particle i interact with all patches on particle j. (b) 'patch–anti-patch' model (p-ap$_1$) where every particle has four identifiable patches ($p^\alpha = \{1, 2, 3, 4\}$) (c) intermediate 'patch–anti-patch' model (p-ap$_2$) where every particle has two distinct patch (anti-patch) types ($p^\alpha = \{1, 1, 2, 2\}$).

$$u_{\text{patch}}(i,j) = u_{sw}(\boldsymbol{r}_{ij}) \sum_{\alpha,\beta=1}^{4} \delta_{p^\alpha,p^\beta} \Phi\left(\boldsymbol{r}_{ij}, \hat{\boldsymbol{p}}_i^\alpha\right) \Phi\left(\boldsymbol{r}_{ji}, \hat{\boldsymbol{p}}_j^\beta\right) \qquad (1)$$

where $\boldsymbol{r}_{ij} = \boldsymbol{r}_j - \boldsymbol{r}_i$, $\hat{\boldsymbol{p}}_i^\alpha$ is the normalized vector pointing from the center of particle i towards patch α on the same particle, $\delta_{p^\alpha,p^\beta}$ is the Kronecker delta function which is equal to one if patch α and β are complementary patches ($p^\alpha = p^\beta$), and u_{sw} is a square-well potential of hard diameter σ, range δ and depth ε. The reduced temperature T is expressed in units ε/k_B, where k_B is Boltzmann's constant. The reduced pressure P and the number density ρ will be expressed in units ε/σ^3 and $1/\sigma^3$ respectively. The function $\Phi(\boldsymbol{r}_{kl}, \boldsymbol{p}^\alpha)$ is defined as

$$\Phi(\boldsymbol{r}, \boldsymbol{p}) = \begin{cases} -1 & \text{if } \hat{\boldsymbol{r}} \cdot \hat{\boldsymbol{p}} < \cos(\theta_m) \\ 0 & \text{otherwise} \end{cases} \qquad (2)$$

with $\hat{\boldsymbol{k}}$ a normalized vector in the direction of $\boldsymbol{k}$.

We study the patchy models for $\cos(\theta_m) = 0.94$ and $\delta = 0.24\sigma$, parameters that ensure that no two patches on one particle can bind to the same patch on another particle. The equilibrium phase diagram and nucleation barriers of the usual 'patch-patch' realization with this set of parameters have been reported by Saika-Voivod et al.[30] It is worth noting that the p-ap$_1$ model can qualified as 'chiral' since there are two different ways to choose the label of the last two patches once the label of the first two have been fixed. Here, we study one realization of the p-ap$_1$ model, but it would be interesting to investigate if solutions with different mixing ratio of the two p-ap$_1$ 'enantiomeric' forms have different phase behaviors.

3 Methods and results

In this section we examine the effect of the patch specificity on the equation of state of the liquid, the liquid–gas coexistence, and the equilibrium phase diagram.

3.1 Equation of state of the liquid phase

The equations of state for the liquid were calculated using isobaric NPT Monte Carlo simulations[31] of $N = 512$ particles, and using the second order virial expansion at very low density. The reduced second virial coefficient of the KF model reads:[18]

$$\frac{B_2}{B_2^{\text{HS}}} = 1 - \chi^2 \left[(1 + \delta/\sigma)^3 - 1 \right] \left(e^{\varepsilon/k_B T} - 1 \right) \qquad (3)$$

with $B_2^{HS} = (2/3)\pi\sigma^3$ the second virial coefficient of hard spheres and χ the fraction of View Article Online the particle surface covered by the N_p attractive patches and is given by

$$\chi^{p-p} = \sqrt{N_p^2(1 - \cos(\theta_m)/2)}. \qquad (4)$$

The B_2 of the two p-ap models have the same functional form but with covering ratios

$$\chi^{p-ap_{Na}} = \sqrt{N_p N_a}(1 - \cos(\theta_m)/2), \qquad (5)$$

that account for the $N_p * N_a$ ($N_a = 1,2$) possible patch–anti-patch pairs that can form between two particles, instead of N_p^2s for the fully patchy case. Equations of state for the p-p and the p-ap_1 models are shown in Fig. 2 together with their compressibility factor $Z = \beta P/\rho$ and B_2 expansion at low density.

3.2 Liquid–gas coexistence

The liquid–gas phase coexistences were calculated using a Grand Canonical Wang-Landau based scheme[32] proposed by Ganzenmuller and Camp[33] and are presented in Fig. 3. Given the reduction in the effective coverage of the particles when considering patch–anti-patch interactions, the critical point of the liquid–gas coexistence curve gets shifted towards lower temperatures as expected, while the critical density is almost unaffected. Interestingly, the liquid–gas phase boundaries of the p-p and the p-ap models collapse on top of each other when plotted as a function of the reduced second virial coefficient defined by eqn (3) and (5). This observation suggests that there may be a generalised law of corresponding states[34,35] for patchy particles that share the same distribution of patches (here tetrahedral) but with a different number of corresponding anti-patchs to a given patch. This observation is potentially useful because it suggests that knowledge of the liquid–gas phase coexistence of the fully patchy model will be sufficient to predict the coexistence for other patch–anti-patch realizations. One could envisage inverting this argument as a method to classify "similarity" in protein–protein interactions.
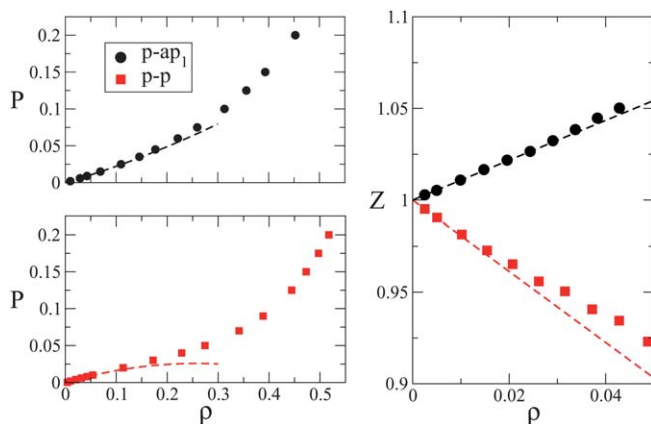

Fig. 2 (left) Fluid equations of state for the patch-patch p-p (squares) and the patch–anti-patch p-ap_1 (dots) models at $T = 0.2$. The Monte Carlo results (symbols) are shown together with the virial expansion up to B_2 (dashed curves). (right) Corresponding compressibility factor $Z = \frac{\beta P}{\rho}$ at low density as obtained from MC simulations and B_2 expansions for the two models.
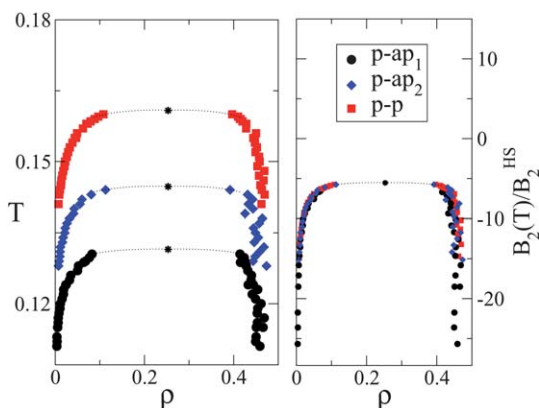
Fig. 3 (left) Liquid–gas coexistence curves for the patch-patch (squares) and the patch–anti-patch models p-ap_1 (dots) and p-ap_2 (diamonds) in the ρ–T plane. Small stars indicate the location of the critical points. The critical temperatures is reduced due to a lower effective coverage of the patchy particles when increasing the specificity of the patches from p-p to p-ap_1; the critical densities are almost unaffected. (right) Liquid–gas coexistence boundary of the three models collapse on top of each other when mapped onto their reduced second virial coefficients $B_2(T)/B_s^{HS}$.

3.3 Phase diagrams

To calculate the phase diagrams for the three models, we first calculate the Helmholtz free energies of the potential phases, and then use common tangent constructions to determine the coexistence regions.

3.3.1 Gas free energy. The Helmholtz free energy of the gas is given by[31]

$$\frac{\beta F(\rho)}{N} = \frac{\beta F_{id}(\rho)}{N} + \beta \int_0^\rho d\rho' \frac{P(\rho') - \rho'/\beta}{\rho^2} \tag{6}$$

where F_{id} is the ideal gas free energy. To determine the equation of state, $P(\rho)$, we used NPT Monte Carlo simulations with 512 particles.

3.3.2 Fluid free energy. The free energy of the fluid at density ρ is calculated by thermodynamic integration using the hard sphere fluid as a reference system; hence the potential used for the thermodynamic integration is

$$U_{liq}(\lambda) = U_{HS} + \lambda U_{patch}. \tag{7}$$

where λ is the coupling parameter, U_{HS} is the hard sphere potential and U_{patch} is the potential associated with the patches given by

$$U_{patch} = \frac{1}{2} \sum_{i,j=1}^N u_{patch}(i,j). \tag{8}$$

Thus, for $\lambda = 0$ the system reduces to a hard sphere fluid and for $\lambda = 1$ we have the full potential energy. The free energy of the fluid is then given by[31]

$$F_{liq} = F_{HS} + \int_0^1 d\lambda \left\langle \frac{\partial U_{liq}(\lambda)}{\partial \lambda} \right\rangle_\lambda \tag{9}$$

where F_{HS} is the free energy of the hard sphere fluid, which has been determined previously by Speedy,[36] and z_λ denotes an ensemble average for a system with

potential energy $U(\lambda)$. The integrand was measured using NVT Monte Carlo simulations with $N = 512$ particles, and the integral was evaluated using a 20 point Gauss–Legendre quadrature.

Using the free energy determined at a reference density $F(\rho_0)$, the free energy as a function of density $F(\rho)$ can be determined using the equation of state. In this case, the free energy is given by[31]

$$\frac{\beta F(\rho)}{N} = \frac{\beta F(\rho_0)}{N} + \beta \int_{\rho_0}^{\rho} d\rho' \frac{P(\rho')}{(\rho')^2} \qquad (10)$$

where $\dfrac{\beta F(\rho_0)}{N}$ is the free energy at a reference density ρ_0. The equation of state $P(\rho)$ was determined using Monte Carlo NPT simulations with $N = 512$ particles.

3.3.3 Solid free energies.

Previous work on this model had identified three possible solid phases: face-centered-cubic, body-centered-cubic, and diamond. All three crystal structures are fully bonded, although the bonding in the FCC system does not have patches pointing directly towards each other. Snapshots of the three crystal phases, and their bonding is shown in Fig. 4.

We are specifically interested in the stability of the low density crystal phase, *i.e.* the diamond phase, and we will only examine the stability of the diamond and the associated bcc phases in the remainder of this paper. It should be noted that the FCC crystal will be stable for high densities (and hence, pressures). and the stable region will be strongly limited by the fact that the particles cannot rotate significantly while remaining bonded.
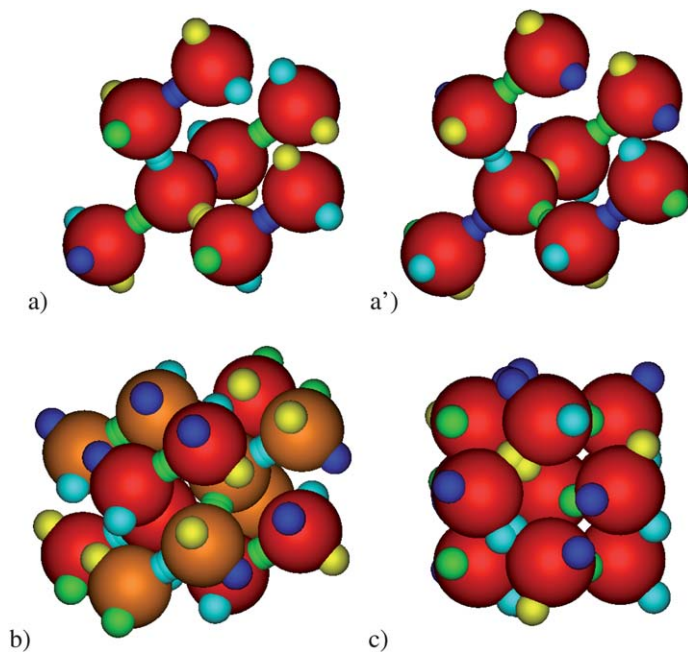

Fig. 4 Three possible fully bonded solid phases for the thetrahedral patch–anti-patch model $p\text{-}ap_1$: (a) cubic diamond (DC), (b) body-centered-cubic (BCC) and face-centered-cubic (FCC). Different patch realizations of DC exist (*e.g.* a and a') while BCC is the result of two inter-penetrating DC lattices (darker and lighter red respectively in b). Note also in c) that patches cannot exactly aligned in the fully bonded FCC and therefore the angular stability range of this crystal is reduced. The particles and patches are not drawn to size.

In the case of the patch-patch model, there is trivially a single realization of the diamond, BCC and FCC phases. However, in the patch–anti-patch case, it is possible for the particles to be slocated on lattice sites, but with a different bonding structure. As shown in Fig. 4, in the case of diamond, there are at least two ways to bond this system. Further examinations of the system show that there are indeed a large number of different ways to bond the system, however, as soon as the orientations of the particles in a single plane are determined, there are no remaining degrees of freedom for the other particle in the system. Hence, this contribution to the free energy per particle will be of the form $\beta F_{\text{bond}}/N = 1/N \log (CN^{2/3})$, where C is a constant; this "surface" term vanishes for large N and hence will be set to zero for the remainder of this paper. This is also the case for the BCC and FCC phases.

In the p-ap_2 system, however, the situation is significantly different: the possibility of locally rearranging the bond identities adds extensively to the total entropy of the crystal phases. As each bond arrangement has the same energy, all possible bonding configurations are equally likely, and the entropy associated with rearranging the bonds can be calculated by estimating the number of possible bond configurations in the system:

$$\beta \delta F_{\text{bond}} = -\log N_{\text{conf}}. \tag{11}$$

The set of configurations in which the system is fully bonded can be sampled by starting from a 'blank' diamond crystal lattice, where the patches are not assigned an identity yet, and generating a bonded configuration according to the following scheme:

1. Select the first blank patch on the first particle in the system that still has blank patches.
2. Randomly assign one of the two possible identities to the selected patch.
3. If any blank patch is bonded with a non-blank patch, assign it the same identity as the patch it is bonded to.
4. If any particle already has two patches with the same identity, give all remaining patches the other identity.
5. Repeat steps 3 and 4 until the system no longer changes.
6. If there are any blank patches remaining, and no particles with more than two patches of the same kind, start again from step 1.

This strategy results in either a fully-bonded configuration, or a configuration with invalid particles (*i.e.* the wrong number of patches of each type). Since at each iteration of the algorithm there are only two choices, and each set of choices will result in a different configuration, this algorithm can be seen as a random walk in a binary tree, ending up at a specific configuration c after taking $d(c)$ decisions. The probability $P(c)$ of finding a specific configuration c is then given by:

$$P(c) = 2^{-d(c)}. \tag{12}$$

As a result, the total number of configurations can be calculated by:

$$N_{\text{conf}} = \sum_c f(c) = \sum_c P(c) 2^{d(c)} f(c) = \langle 2^{d(c)} f(c) \rangle, \tag{13}$$

where $f(c)$ is a function that equals 1 if a configuration is fully bonded, and 0 if it is invalid. The sum is taken over all possible configurations, and the angular brackets denote averaging over a set of configurations resulting from random walks through the binary tree, as described above.

Thus, the number of fully bonded configuration for a fixed system size can be calculated by sampling a large number of randomly generated configurations. Of course, the number of valid configurations is highly dependent on the number of particles. To estimate the contribution of the bonding entropy in the thermodynamic limit, we calculated the bonding free energy for a range of system sizes and

extrapolated the result to an infinitely large system, starting from either a completely blank crystal, or one where only particles in a cube-shaped region were blanked. For each system size, we sampled at least 10^5 configurations. The resulting extrapolation is shown in Fig. 5, and yields a bonding entropy of $\beta F_{bond}/N = -0.34 k_B T$ per particle. While accurately sampling the number of configurations requires a large number of runs, we find that changes of around $0.05 k_B T$ in the bonding entropy do not significantly affect the phase diagram for the p-ap_2 system. Apart from local rearrangements, we expect the strongest contribution to the entropy for large systems to be either plane defects (for the fully free crystal), or surface effects (for the crystal with constrained boundaries). For both of these, the entropic contribution to the total free energy should be proportional to $N^{2/3}$. Thus, the leading finite-size correction term for the free energy per particle will be on the order of $N^{-1/3}$.

To calculate the complete free energies of the diamond and bcc solid phases we again used thermodynamic integration. In this case, we choose an Einstein crystal with a fixed center of mass and constrained orientation of the particles as the reference state:

$$U(\lambda) = U_{HS} + (1 - \lambda)U_{patch} + \lambda U_{Ein} + \lambda U_{rot} \qquad (14)$$

where U_{HS} is the hard sphere potential, U_{Ein} is the typical Einstein term which attaches particles to their "ideal" lattice sites denoted r_i^0 and is given by

$$U_{Ein} = \sum_{i=0}^{N}(r_i - r_i^0)^2,$$

and the fourth term ties the orientation of the particles to their "ideal" orientation; U_{rot} is given by

$$\beta U_{rot}^{p-p} = \sum_{i=1}^{N} \min_{k \neq l} \left\{ 2 - \text{sign}(c_{a,k})c_{a,k}^2 \, \delta_{p^a,p^k} - \text{sign}(c_{b,l})c_{b,l}^2 \delta_{p^b,p^l} \right\} \qquad (15)$$

where $c_{a,k} = \hat{p}_i^k \cdot \hat{p}_i^{a,0}$, $\hat{p}_i^{a,0}$ is the "ideal" direction for patch a on particle i, and a and b correspond to two, nonidentical patches. In the tetrahedral case, any two patches are sufficient since all patches are not colinear. In a more general case, such as for
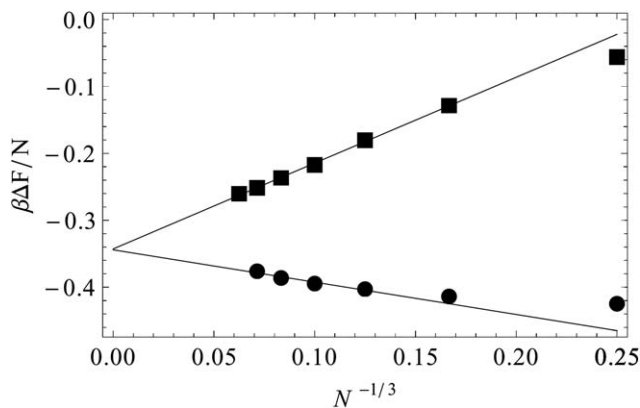

Fig. 5 Bonding entropy per particle as a function of the system size, both for a fully free system (circles), and a system where the bond identity for the particles at the edge is fixed (squares). The lines are linear fits, leading to an estimate of $\beta\delta F_{bond}/N = -0.34$ in the thermodynamic limit.

patches arranged in an octahedral symmetry, two non-colinear patches would need to be chosen. In the patch-antipatch case (p-ap_1), U_{rot} simplifies to

$$\beta U_{rot}^{p-ap} = \sum_{i=1}^{N} \left(2 - \text{sign}(c_{a,a})c_{a,a}^2 - \text{sign}(c_{b,b})c_{b,b}^2 \right). \qquad (16)$$

The free energy of the non-interacting system consists of the free energy of the Einstein crystal (F_{ein}) for symmetric particles with a fixed center of mass plus the non-interacting rotational free energy (F_{rot}). The expression for an Einstein crystal with a fixed center of mass is given by[37]

$$\frac{\beta F_{ein}^{CM}}{N} = -\frac{3}{2}\ln\frac{2\pi}{\lambda} - \frac{3}{2N}\ln\frac{\lambda}{2\pi} + \frac{\ln\rho}{N} - 2\frac{\ln N}{N} \qquad (17)$$

while the rotational free energy is[19]

$$\frac{\beta F_{rot}}{N} = -\ln\left[\frac{\int d\Omega \, \exp(-\beta\lambda U_{rot})}{\int d\Omega} \right]. \qquad (18)$$

Note that F_{Ein}^{CM} does not depend on the choice of model, $i.e.$ patch-patch or patch–anti-patch. However, F_{rot} does depend on the model.

The full free energy of the system is given by

$$\frac{\beta F(\rho)}{N} = \frac{\beta F_{ein}^{CM}}{N} + \frac{\beta F_{rot}}{N} + \frac{\beta F_{int}}{N} + \frac{\beta F_{bond}}{N} \qquad (19)$$

where

$$\frac{\beta F_{int}}{N} = -\frac{\beta}{N}\int \langle U_{Ein} + U_{rot} - U_{patch}\rangle_\lambda d\lambda. \qquad (20)$$

Recall that $\beta F_{bond}/N$ is non-zero only in the p-ap_2 case. As in ref. 37, we calculate the integral in eqn (20) using 20 point Gauss–Legendre quadrature in combination with NVT MC simulations.

3.3.4 Phase coexistence. The phase diagrams of the three models presented in Fig. 6 summarize the effect of the interaction specificity on the equilibrium phase diagram of a colloidal model that include both a liquid–gas (L-G) and a fluid-diamond coexistence. The main result of our investigation is that, while the L-G boundaries shift towards lower temperatures when increasing the specificity of the patches from p-p to p-ap_1, the fluid-DC coexistence is less affected and does not follow the L-G critical point. The result is a large metastability gap that opens up between the liquid–gas critical temperature and the fluid-DC melting temperature at the critical density when increasing the specificity ot the interaction. The metastability gap M that measures the degree of metastability is usually the driving force for crystal nucleation at the critical point:

$$M = \frac{T_x - T_c}{T_c} \qquad (21)$$

with T_x the freezing temperature of the solid (here DC) at the critical density and T_c the metastable L-G critical temperature.

In the case of isotropic potentials, T_c decreases faster than T_x upon reducing the interaction range and hence M increases as the range λ gets narrower, making crystallization just above or just below the critical point possible. However, Romano et al.[27] observed that for the the tetrahedral patch-patch model, the thermodynamic driving force for crystallisation at the critical point increases only slightly with decreasing λ. As a consequence, low-valence p-p particles have to be cooled well
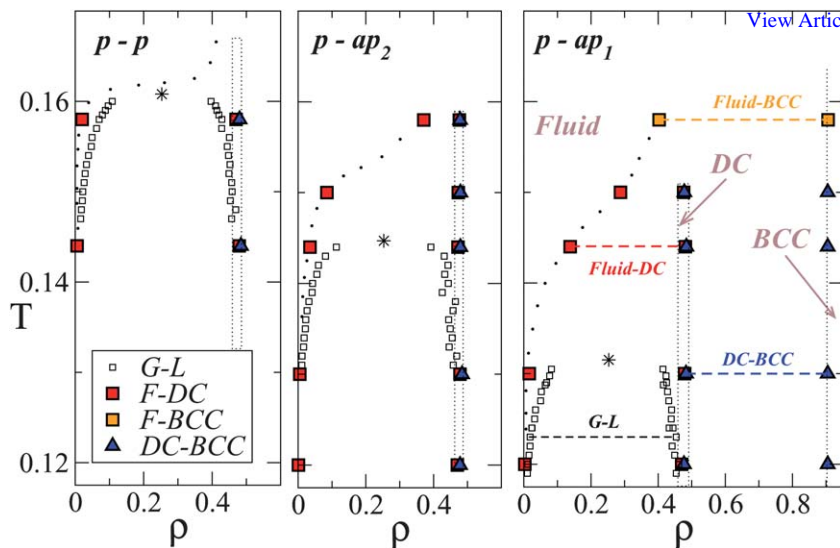
Fig. 6 Phase diagrams of the patch-patch and patch–anti-patch models. The fluid, diamond-cubic (DC) and body-centered-cubic BCC phases were considered. Increasing the specificity of the interaction from p-p to p-ap_1 enlarges the metastability gap M between the liquid–gas (G-L) critical temperature and the Fluid-DC melting temperature at the critical density. Dotted lines are guide to the eye for the boundary of the different phases and dashed lines account for phase coexistence.

below the critical temperature before there is a pronounced thermodynamic driving force for crystallisation. Yet, the viscosity of the (highly networked) liquid phase of patchy particles will increase very strongly with decreasing temperature and hence, when crystal nucleation becomes thermodynamically allowed it will be kinetically suppressed.

Surprisingly – and potentially relevant for real proteins – crystal nucleation in a system with the more specific p-ap interactions can occur in a regime close to the L-G critical point and is therefore more likely to be kinetically accessible: making the patchy interactions more specific opens up a gap between the freezing curve and the L-G coexistence region, in the same way that decreasing the range of attraction does for isotropic potentials.

There is, however, an important difference between the phases that crystallise in the isotropic and the p-ap case: the patch–anti-patch models that we consider freeze into low-density crystals, whilst purely isotropic interactions crystallize into densely packed phases. This observation seems highly relevant for protein crystallisation. We note that our results are also consistent with the phase diagrams obtained in a recent study on the effect of specific and non-specific interaction for the crystallization of a 2D model model of the SbpA surface-layer protein.[38]

The existence of a large metastability gap between the L-G and the Fluid-DC coexistence curves for p-ap models has important implications for the conditions under which crystal nucleation can take place: in the "symmetric" (patch-patch) case, the DC crystal can only nucleate from a fluid with a density that is close to that of the crystal phase. However, for the p-ap system it should be possible to nucleate the DC crystal from a low density fluid, which is the normal condition for protein crystallisation. Of course, thermodynamic driving force is not the whole story: as a liquid gets more supercooled, it will become more viscous and hence the kinetic pre factor for the nucleation rate would go down with supercooling. However, at the critical point, all patchy systems that we studied have the same

reduced B_2 and this would suggest that the ease with which a bond can be broken in the liquid is the same for the *p-p* and *p-ap* models. If we assume that the time scale for bond breaking determines viscosity, then we expect the different patchy liquids to have very similar viscosities near T_c—however, we did not test this.

Of course, the structure of real proteins is such that they do not crystallize into DC structures.[3] They do, however, crystallise into open crystal structures. We believe that the tetrahedral patch–anti-patch model studied here gives valuable insights into their phase diagram and provides a better insight in the origin of the metastability gap that is typically found in real protein solutions: importantly, the short range of the attraction is not the whole story—anisotropy *and* specificity are important.

Many simulation papers on colloidal self assembly idly invoke the relevance of their results for the design of photonic band-gap materials. Although the authors are reluctant to join this bandwagon, the present study may actually be of real relevance for the strategies to make colloidal crystals with a diamond structure. Our message is: it is not enough to make patchy colloids with tetrahedral symmetry—rather, one should make tetrahedral colloids where some or all 4 patches have a different functionality (something that could, for instance, be achieved with DNA functionalization).

Acknowledgements

References

1 A. McPherson, *Crystallization of Biological Macromolecules*, Cold Spring Harbor Laboratory Press, 1999.
2 R. P. Sear, *J. Chem. Phys.*, 1999, **111**, 4800–4806.
3 S. W. Wukovitz and T. O. Yeates, *Nat. Struct. Biol.*, 1995, **2**, 1062–1067.
4 M. Muschol and F. Rosenberger, *J. Chem. Phys.*, 1997, **107**, 1953.
5 A. Lomakin, N. Asherie and G. B. Benedek, *Proc. Natl. Acad. Sci. U. S. A.*, 1999, **96**, 9465–9468.
6 F. Romano, E. Sanz and F. Sciortino, *J. Chem. Phys.*, 2011, **134**, 174502.
7 N. E. Chayen, *Curr. Opin. Struct. Biol.*, 2004, **14**, 577–583.
8 A. George and W. Wilson, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 1994, **50**, 361, year.
9 Rosenbaum, Zamora and Zukoski, *Phys. Rev. Lett.*, 1996, **76**, 150–153.
10 P. R. ten Wolde and D. Frenkel, *Science*, 1997, **277**, 1975–1978.
11 P. N. Pusey and W. van Megen, *Nature*, 1986, **320**, 340–342.
12 E. Zaccarelli, G. Foffi, K. A. Dawson, S. V. Buldyrev, F. Sciortino and P. Tartaglia, *Phys. Rev. E: Stat. Phys., Plasmas, Fluids, Relat. Interdiscip. Top.*, 2002, **66**, 41402.
13 S. Auer and D. Frenkel, *Nature*, 2001, **409**, 1020–1023.
14 C. Liu, N. Asherie, A. Lomakin, J. Pande, O. Ogun and G. B. Denedek, *Proc. Natl. Acad. Sci. U. S. A.*, 1996, **93**, 377–382.
15 N. Dorsaz, G. M. Thurston, A. Stradner, P. Schurtenberger and G. Foffi, *Soft Matter*, 2011, **7**, 1763–1776.
16 P. R. Banerjee, A. Pande, J. Patrosz, G. M. Thurston and J. Pande, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**, 574–579.
17 C. D. Michele, S. Gabrielli, P. Tartaglia and F. Sciortino, *J. Phys. Chem. B*, 2006, **110**, 8064–8079.
18 N. Kern and D. Frenkel, *J. Chem. Phys.*, 2003, **118**, 9882.
19 E. G. Noya, C. Vega, J. P. K. Doye and A. A. Louis, *J. Chem. Phys.*, 2007, **127**, 054501.
20 J. Chang, A. M. Lenhoff and S. I. Sandler, *J. Chem. Phys.*, 2004, **120**, 3003–3014.
21 E. Bianchi, J. Largo, P. Tartaglia, E. Zaccarelli and F. Sciortino, *Phys. Rev. Lett.*, 2006, **97**, 168301.
22 A. L. M. Hloucha, J. F. M. Lodge and S. Sandler, *J. Cryst. Growth*, 2001, **232**, 195.

23 B. L. Neal, D. Asthagiri and A. M. Lenhoff, *Biophys. J.*, 1998, **75**, 2469–2477. View Article Online

24 D. Asthagiri, B. L. Neal and A. M. Lenhoff, *Biophys. Chem.*, 1999, **78**, 219–231.

25 Y. Yan, A. de Keizer, M. Stuart and N. Besseling, in *Self organized nanostructures of amphiphilic block copolymers II*, Springer Verlag berlin, 2011, ch. From Coordination Polymers to Hierarchical Self-Assembled Structures, pp. 91–115.

26 S. Jones and J. M. Thornton, *Proc. Natl. Acad. Sci. U. S. A.*, 1996, **93**, 13–20.

27 F. Romano, E. Sanz and F. Sciortino, *J. Phys. Chem. B*, 2009, **113**, 15133–15136.

28 F. Romano, E. Sanz and F. Sciortino, *J. Chem. Phys.*, 2010, **132**, 184501.

29 E. G. Noya, C. Vega, J. P. K. Doye and A. A. Louis, *J. Chem. Phys.*, 2010, **132**, 234511.

30 I. Saika-Voivod, F. Romano and F. Sciortino, *J. Chem. Phys.*, 2011, **135**, 124506–124510.

31 D. Frenkel and B. Smit, *Understanding Molecular Simulations: From Algorithms to Applications*, Academic Press, London, UK, 2002.

32 F. Wang and D. P. Landau, *Phys. Rev. E: Stat. Phys., Plasmas, Fluids, Relat. Interdiscip. Top.*, 2001, **64**, 056101.

33 G. Ganzenmller and P. J. Camp, *J. Chem. Phys.*, 2007, **127**, 154504.

34 M. Noro and D. Frenkel, *J. Chem. Phys.*, 2000, **113**, 2941.

35 G. Foffi and F. Sciortino, *J. Phys. Chem. B*, 2007, **111**, 9702–9705.

36 R. J. Speedy, *J. Phys.: Condens. Matter*, 1997, **9**, 8591–8599.

37 J. M. Polson, E. Trizac, S. Pronk and D. Frenkel, *J. Chem. Phys.*, 2000, **112**, 5339.

38 T. K. Haxton and S. Whitelam, *Soft Matter*, 2012, **8**, 3558–3562.