Simple models for charge and salt effects in protein crystallization

2002 J. Phys.: Condens. Matter 14 7617

(http://iopscience.iop.org/0953-8984/14/33/305)

INSTITUTE OF PHYSICS PUBLISHING JOURNAL OF PHYSICS: CONDENSED MATTER

J. Phys.: Condens. Matter **14** (2002) 7617–7629 PII: S0953-8984(02)38379-6

Simple models for charge and salt effects in protein crystallization

Patrick B Warren

Unilever Research and Development Port Sunlight, Bebington, Wirral, CH63 3JW, UK

Received 19 June 2002
Published 9 August 2002
Online at stacks.iop.org/JPhysCM/14/7617

Abstract

A simple extension of existing models for protein crystallization is described, in which salt ions and charge neutrality are explicitly incorporated. This provides a straightforward explanation for the shape of protein crystallization boundaries and the associated scaling properties seen for lysozyme, and can also explain much of the salt dependence of the second virial coefficient. The analysis has wider implications for the use of pair potentials in understanding protein crystallization.

1. Introduction

Protein crystallization is of great practical importance, both as a purification method and since high-quality crystals are needed for x-ray diffraction work [1, 2]. Under typical conditions where crystals are obtained, protein–protein interactions appear to be characterized by hard-core repulsions with short-ranged attractions [3, 4]. For instance, Rosenbaum *et al* [5] successfully collapsed the crystallization boundaries for a number of proteins onto the adhesive hard-sphere (AHS) model [6–8] by matching the second virial coefficient B_2. Whilst the AHS model is frequently used to understand protein crystallization, it is also possible that crystallization is an energetic ordering transition driven by highly directional interactions. A model of hard spheres (HS) with sticky patches was introduced by Sear [9] to capture this important possibility.

In this paper I examine the properties of a simple model for protein crystallization which takes into account the effects of charge in a very elementary way. Such a model suggests that some generic aspects of protein crystallization can be explained as a straightforward consequence of charge neutrality, at least for lysozyme which readily undergoes crystallization. These various aspects are:

(i) the overall shape of the crystallization boundary;
(ii) a scaling collapse of the crystallization boundary noticed by Poon *et al* [10] when plotted as a function of c_s/Q^2, where c_s is the NaCl concentration and Q the protein charge;

(iii) a similar scaling of the lysozyme solubility data in Guo *et al* [11]; and
(iv) a similar scaling for B_2 noticed by Egelhaaf and Poon [12].

The appearance of c_s/Q^2 in all of these is very reminiscent of the Donnan membrane equilibrium [13], where one finds a contribution to B_2 equal to $Q^2/4c_s$ provided one is in the high-salt limit $c_s \gg Qc_p$ where c_p is the protein concentration. The effect is well known in polyelectrolyte theory [14, 15]. This result is usually derived by supposing ideal solution behaviour and imposing electrical neutrality on the two sides of a semi-permeable membrane. In the present case, I argue there is an analogous Donnan equilibrium between protein crystals and a dilute protein solution, with the crystal–solution interface playing the role of the membrane. The appearance of c_s/Q^2 as a scaling variable reflects the fact that the solution (not necessarily the crystal) is in the high-salt limit, as will be explained below. Donnan's theory appears to have little to do with the more conventional McMillan–Mayer approach [16, 17] in which effective potentials between proteins are invoked to explain crystallization and the effects of added salt, often in the context of DLVO potentials [10] where Debye–Hückel theory is used to account for the charge interactions. However, a careful study of the relationship between the two approaches made by Hill in the 1950s indicates that the two approaches should converge to the same results in the high-salt limit [18].

To apply these ideas to the protein crystallization problem, one constructs a free energy which enforces the charge neutrality constraint. Small ions can be treated as ideal solution species, but one must at least incorporate non-ideality of the protein solution, to allow it to form the ordered phase which represents the protein crystals. The first thing to do is to set up this general theory.

2. General theory

I suppose the proteins under consideration have a molecular volume V_p; for example lysozyme is a charged globular protein of approximate size $V_p = (\pi/6) \times 4.5 \times 3.0 \times 3.0$ nm^3 = 21.2 nm^3 [10, 19], or a molar volume $N_A V_p = 12.8$ M^{-1} where N_A is Avogadro's number. If the protein concentration is c_p, the effective volume fraction is $\phi = V_p c_p$. In the absence of charge effects, a baseline model which incorporates a freezing transition to correspond to protein crystallization will be fully specified by a dimensionless free energy density $f^{(0)} = V_p \beta F^{(0)}/V$, where $F^{(0)}$ is the free energy of the protein solution or crystal, V is the system volume, and $\beta = 1/kT$. There may be different branches of $f^{(0)}$ to correspond to the fluid and ordered (crystal) phases. Below I shall consider two possibilities for this baseline model: Sear's model of HS with sticky patches, and the AHS model. The full free energy is obtained from the baseline model by adding in contributions from the coions and counterions (which I suppose to be univalent) and imposing charge neutrality. If the added salt concentration is c_s, there will be coions at a concentration c_s and counterions at a concentration $Qc_p + c_s$. Denoting the dimensionless free energy density for the full model by f, I write

$$f(\phi, \phi_s) = f^{(0)}(\phi) + \phi_s[\log \phi_s - 1] + (Q\phi + \phi_s)[\log(Q\phi + \phi_s) - 1] \qquad (1)$$

where the last two terms are the ideal mixing terms from the coions and counterions respectively, written using $\phi_s \equiv V_p c_s$ for notational simplicity. Unimportant constants and terms linear in ϕ or ϕ_s have been dropped. Non-ideality of the small ions is neglected in the present treatment, although this is certainly a refinement which should be considered for more quantitative work.

This free energy is a function of two density variables: ϕ and ϕ_s. A phase equilibrium such as between protein crystals and protein solution corresponds to equality of osmotic pressure and chemical potentials for *both* components. As a consequence there will in general be *different*

values of ϕ_s in coexisting phases, a salt repartitioning effect which is not usually taken into account. To solve for phase equilibria, it is useful to transform f into a semi-grand potential $h(\phi, \phi_s^{(\mathrm{R})})$ where $\phi_s^{(\mathrm{R})} \equiv V_p c_s^{(\mathrm{R})}$ is a dimensionless salt *reservoir* concentration [20]. To make this transformation note that the salt chemical potential is

$$\beta\mu_s = \frac{\partial f}{\partial \phi_s} = \log\phi_s + \log(Q\phi + \phi_s). \tag{2}$$

Now $\phi_s \to \phi_s^{(\mathrm{R})}$ in the limit $\phi \to 0$; therefore $\beta\mu_s = 2\log\phi_s^{(\mathrm{R})}$. This gives

$$(\phi_s^{(\mathrm{R})})^2 = \phi_s(Q\phi + \phi_s) \tag{3}$$

which is the usual Donnan equilibrium result that the product of the coion and counterion concentrations takes a constant value in all phases including the reservoir. Solving this gives

$$\phi_s = [(Q^2\phi^2 + 4(\phi_s^{(\mathrm{R})})^2)^{1/2} - Q\phi]/2. \tag{4}$$

The semi-grand potential $h = f - \beta\mu_s\phi_s$ and the first two derivatives with respect to ϕ at constant $\phi_s^{(\mathrm{R})}$ are, after a few lines of calculus,

$$h = f^{(0)} + Q\phi\log(Q\phi + \phi_s) - (Q\phi + 2\phi_s), \tag{5}$$

$$\partial h/\partial\phi = \partial f^{(0)}/\partial\phi + Q\log(Q\phi + \phi_s), \tag{6}$$

$$\partial^2 h/\partial\phi^2 = \partial^2 f^{(0)}/\partial\phi^2 + Q^2/(Q\phi + 2\phi_s). \tag{7}$$

The advantage of this transformation is that h is effectively a one-component free energy and can be treated accordingly. To use these results, one should remember to substitute for ϕ_s from equation (4). For example, the osmotic pressure follows from $V_p\beta\Pi = \phi(\partial h/\partial\phi) - h$; i.e.,

$$\Pi = \Pi^{(0)} + kT(Qc_p + 2c_s). \tag{8}$$

This shows explicitly the small ions behaving ideally, at a total concentration $Qc_p + 2c_s$.

What is obvious from these results is that there is a crossover in behaviour at $Qc_p \sim c_s$ or salt concentrations of the order $Q\phi/V_p$. The *difference* in protein volume fraction between the solution and crystal is often $\Delta\phi \sim 0.5$. Putting $Q \sim 10$, and $V_p \sim 10\ \mathrm{M}^{-1}$, this corresponds to a crossover salt concentration ~0.5 M. If the salt concentration is much less than this, there will be a large osmotic pressure difference between the crystal and the solution due to the counterions, having the effect of narrowing the coexistence gap. This is the basic reason that the coexistence boundary occurs at salt concentrations of this order of magnitude in this model. It is a much larger crossover salt concentration than intuition might have suggested based on experience with colloidal systems (e.g. $Q \sim 10^3$, $V_p \sim 10^6\ \mathrm{nm}^3$ gives $c_s \sim 10^{-3}$ M), but then globular proteins are much smaller than colloids.

On the other hand, typical protein solutions on the crystallization boundary have $\phi \sim 0.05$, corresponding to a crossover salt concentration ~50 mM. For salt concentrations much larger than this, the effects of salt and charge are subsumed into a scaling variable $\phi_s^{(\mathrm{R})}/Q^2$. Since $\phi_s^{(\mathrm{R})}$ is more or less the salt concentration in the solution if the protein concentration is small, this is a natural explanation for the scaling properties described in the introduction. In the following sections, I place these arguments on a firm footing, starting with discussions of the second virial coefficient and the high-salt scaling of solubility, since these are not dependent on any particular baseline model.

3. Second virial coefficient

Any theory for the free energy of a protein solution contains a prediction for the second virial coefficient. For the present theory, B_2 can be obtained from the osmotic pressure result

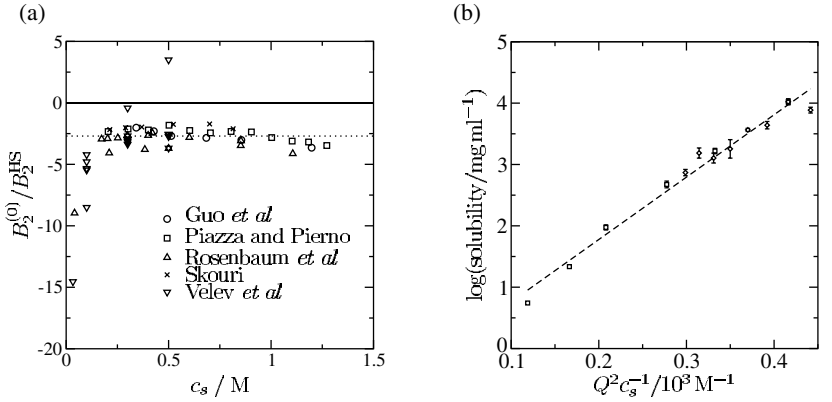

Figure 1. (a) $B_2^{(0)}/B_2^{HS}$ for lysozyme, where $B_2^{(0)} = B_2 - Q^2/4c_s$, as a function of salt concentration c_s [12]. The dotted line is the average, equation (9), for all data with $c_s > 0.25$ M. (b) Solubility data for lysozyme from Guo *et al* [11]. The dashed line is the least-squares fit to the data, equation (12) in the text.

equation (8) (recalling that equation (4) should be used for ϕ_s), or directly by inspection from equation (7). Either way, as stated in the introduction, $B_2 = B_2^{(0)} + Q^2/4c_s$ where $B_2^{(0)}$ is the contribution from the baseline model. One can prove the same result holds in light scattering determination since charge density fluctuations are suppressed in the long-wavelength limit. We can hope that the baseline model is insensitive to the net charge and salt concentration, so that $B_2^{(0)}$ is independent of Q and c_s, but is there any evidence for this? For lysozyme, Egelhaaf and Poon [12] have collected experimental data on B_2 from the available literature[1]. These data are used to construct figure 1(a) which shows $B_2 - Q^2/4c_s$ normalized by the value expected if lysozyme proteins behave as HS, namely $B_2^{HS} = 4V_p \approx 85$ nm^3 (this is close to $B_2^{HS} = 82.3$ nm^3 used in [22]). Inspecting figure 1(a), it appears that there is indeed a data collapse to an approximate plateau for $c_s \gtrsim 0.25$ M, although a downwards trend of 40% or so can be detected at high salt levels and there may also be a problem if the protein charge is too small (the 'rogue' point with $B_2^{(0)} > 0$ is for $Q = 3.6$). There is considerable variation in the data, even for measurements at supposedly identical state points, which reflects the experimental difficulties in obtaining B_2. Averaging over all results in the plateau region in figure 1(a), one arrives at

$$B_2^{(0)} \approx (-2.7 \pm 0.2) B_2^{HS}. \tag{9}$$

This result can be used as a constraint for any particular baseline model for lysozyme. Furthermore, the approximate constancy of $B_2^{(0)}$ as Q and c_s are varied suggests that it is at least reasonable to make the baseline model independent of Q and c_s, as is assumed in the remainder of this paper.

4. High-salt scaling behaviour

In the high-salt limit, a simple result can be obtained for the crystallization boundary (equivalently the protein solubility) by assuming that the baseline model is independent of salt and charge, the density in the crystal remains constant, and the protein concentration in

[1] Data for B_2 is collected from Guo *et al* [11], Piazza and Pierno [21], Rosenbaum *et al* [22], Skouri cited in [21], and Velev *et al* [23].

the solution phase becomes very small with the result that the solution can be treated as ideal. These features are certainly seen for the two specific baseline models discussed below. The simplifications amount to treating $\partial f^{(0)}/\partial\phi$ in equation (6) as a constant independent of Q and c_s in the protein crystal, and taking the dilute limit $\partial f^{(0)}/\partial\phi \to \log\phi_{\rm F}$ in the solution where $\phi_{\rm F}$ is the protein concentration in the solution phase. Equating protein chemical potentials in the two phases results in

$$\log\phi_{\rm F} + Q\log(Q\phi_{\rm F} + \phi_{s,{\rm F}}) = \text{constant} + Q\log(Q\phi_{\rm X} + \phi_{s,{\rm X}}) \tag{10}$$

where $\phi_{s,{\rm F}}$ and $\phi_{s,{\rm X}}$ are the salt concentrations in the solution and crystal phases respectively, and $\phi_{\rm X}$ is the protein volume fraction in the crystal. The terms in this can be expanded assuming that $\phi_{\rm F} \ll 1$, $\phi_{s,{\rm F}} \approx \phi_s^{({\rm R})}$, $\phi_{s,{\rm X}} \gg Q\phi_{\rm X}$, and making use of equation (4) to get

$$\log\phi_{\rm F} \approx \text{constant} + \frac{Q^2\phi_{\rm X}}{2\phi_{s,{\rm F}}}. \tag{11}$$

The prediction is that the logarithm of the protein solubility should be proportional to the square of the protein charge, and inversely proportional to the salt concentration in the solution. Figure 1(b) shows that such a law is indeed satisfied for the solubility data for lysozyme from Guo *et al* [11]. The best-fit line in figure 1(b) is

$$\log\left(\frac{\text{solubility}}{\text{mg ml}^{-1}}\right) = (-0.25 \pm 0.15) + (0.010 \pm 0.001)\frac{Q^2}{c_s/{\rm M}}. \tag{12}$$

The predicted slope from equation (11) is $\phi_{\rm X}/2V_p$. If we take $\phi_{\rm X} \approx 0.5$ as a reasonable estimate of the protein volume fraction in the crystal, then $\phi_{\rm X}/2V_p \approx 0.02$ M which around twice the measured slope. However, there are a number of omitted effects in the present theory which could account for the discrepancy. Nevertheless, the observation that the logarithm of the protein solubility is inversely proportional to the inverse salt concentration is apparently quite common[2]. Also note that solubility $\sim {\rm e}^{Q^2}$ indicates a very strong dependence on the protein charge. This is a natural explanation for the dramatic decrease in solubility around the isoelectric point described in [1].

5. Two specific baseline models

For lower salt concentrations, the protein solubility is not small and the above treatment needs refinement. I now consider the minimal extension to the basic model that takes into account protein non-ideality by specifying a particular baseline model. In this way, the advantage of an analytic transformation to the semi-grand potential h in equations (5)–(7) is retained. The baseline model should encompass both the fluid and crystal phases, and have a second virial coefficient consistent with the analysis above. One such suitable model has been devised by Sear.

5.1. Sear's model

Sear's model comprises HS with 'sticky patches' [9]. The HS diameter is σ, chosen such that $\pi\sigma^3/6 = V_p$ (i.e. $\sigma \approx 3.4$ nm for lysozyme). There are n_s sticky patches per sphere which associate in pairs and are characterized by a range $r_c > \sigma$, an angular width θ_c, and a depth ϵ. The free energy of the fluid (F) phase (I reproduce only the essential details of the model here) is

$$f_{\rm F}^{(0)}(\phi) = f_{\rm HS}(\phi) + \phi n_s[\log p + (1-p)/2]. \tag{13}$$

[2] Reference [1] suggests alternatively that the logarithm of the protein solubility should be proportional to ionic strength, but this is a poorer fit to the solubility data of Guo *et al* [11] than equation (12).

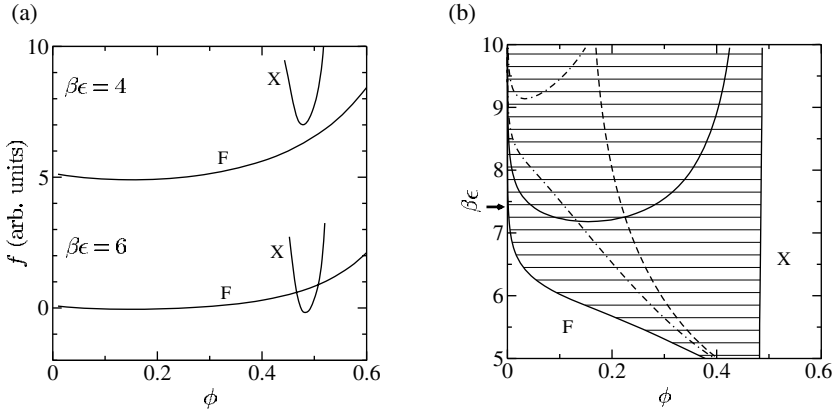

Figure 2. (a) Free energies for the baseline model illustrating metastability of the crystal branch (X) with respect to the fluid branch (F) for $\beta\epsilon \lesssim 4.8$. (b) Phase behaviour in the $(\phi, \beta\epsilon)$ plane for the full model: solid curves are for the baseline model ($Q = c_s^{(\mathrm{R})} = 0$), the dashed curve is for $Q = 10$ and $c_s^{(\mathrm{R})} = 0$, and the chain curves are for $Q = 10$ and $c_s^{(\mathrm{R})} = 0.2$ M. Other parameters are $n_s = 6$, $r_c = 1.05\sigma$, $\theta_c = 0.45$.

Table 1. Fluid–fluid critical points for Sear's model showing critical parameter values $\phi^{(\mathrm{C})}$, $k^{(\mathrm{C})}$, and $B_2^{(\mathrm{C})}$ for several values of n_s.

n_s	$\phi^{(\mathrm{C})}$	$k^{(\mathrm{C})}$	$B_2^{(\mathrm{C})}/B_2^{\mathrm{HS}}$
4	0.09	16.8	−7.38
5	0.12	7.02	−3.39
6	0.15	3.90	−1.93
8	0.21	1.77	−0.77

In this, f_{HS} is the HS fluid free energy and p is the proportion of non-bonded sites, solving $(1-p)/p^2 = k\phi g_{\mathrm{HS}}$ where

$$k = 6(r_c/\sigma - 1)(1 - \cos\theta_c)^2 \mathrm{e}^{\beta\epsilon} \tag{14}$$

is a dimensionless bond association constant. The association equilibrium includes an enhancement factor, g_{HS}, for the HS pair correlation function at contact. The second virial coefficient in this model is given by

$$B_2^{(0)} = B_2^{\mathrm{HS}} - (n_s/2)kV_p. \tag{15}$$

It is apparent that the fluid phase properties are completely determined by k and n_s. The model predicts fluid–fluid phase separation for sufficiently large values of k, and table 1 gives the critical point for $n_s = 4$–8. Note that the second virial coefficient at the critical point in this model provides an marked counterexample to the observation of Vliegenthart and Lekkerkerker [24] that $B_2^{(\mathrm{C})}/B_2^{\mathrm{HS}} \approx -1.5$ for a wide variety of other models. Whether fluid–fluid phase separation is metastable in this model depends on the actual values of r_c, θ_c, and $\beta\epsilon$, and examples of both are given by Sear.

Sear provides a cell model for the protein chemical potential in an ordered cubic phase, arguing that osmotic pressure is not important. My approach here is slightly different. I use the same cell model to specify the free energy and take into account the osmotic pressure. The results are not essentially different from Sear's results, although the analysis does highlight an

important property of Sear's model (figure 2(a)). Following Sear, the free energy per protein in the crystal (X) is

$$\frac{f_{\mathrm{X}}^{(0)}(\phi)}{\phi} = C - 3\log\left(\frac{a}{\sigma} - 1\right) - \log\left(\frac{\theta_c^3}{\pi^2}\right) - \frac{n_s}{2}\beta\epsilon\, w\left[\frac{(\phi - \phi_{min})}{\delta\phi}\right] \qquad (16)$$

where $a/\sigma = (6\phi/\pi)^{-1/3}$ is the unit-cell size relative to the HS diameter. The constant in this is $C = \log(V_p/\sigma^3)$, provided that $f_{\mathrm{HS}} \sim \phi(\log\phi - 1)$ in the fluid phase as $\phi \to 0$. This free energy is appropriate for $\phi \gtrsim \phi_{min}$ where $\phi_{min} = \pi(\sigma/r_c)^3/6$ is the volume fraction around which the bonds in the crystal become dissociated. As ϕ decreases past ϕ_{min}, the last term in equation (16) vanishes rapidly. To implement this I have introduced an *ad hoc* cut-off function $w[(\phi - \phi_{min})/\delta\phi]$ in the last term, where $\delta\phi$ sets the rate at which the cut-off operates. For the present calculations I take $w[x] = 1/(1+\mathrm{e}^{-x})$ and $\delta\phi = 0.01$. This cut-off function represents the way in which the short-range attraction potential falls off with distance in the model. The actual details may shift the phase boundaries but are not important for the broad picture.

Although several values of the parameters in the model were examined, I only report in detail here on calculations for $n_s = 6$, $r_c = 1.05\sigma$, and $\theta_c = 0.45$ (so the range of the attraction is $r_c - \sigma \approx 2$ Å and the angular width about 26°). These values were chosen to give quite good agreement with the crystallization boundaries for lysozyme in the present model. Interestingly, in a separate analysis Curtis *et al* [25] also conclude that $n_s = 6$–8 is appropriate, and Oki *et al* [26] identify three 'macrobonds' between lysozyme molecules from crystallographic data, again corresponding to $n_s = 6$ contacts per protein.

Figure 2(a) shows the fluid and crystal free energies for $n_s = 6$, $r_c = 1.05\sigma$, and $\theta_c = 0.45$, for two values of $\beta\epsilon$. It is clear that there is a certain minimum value $\beta\epsilon \approx 4.8$ below which the crystal is *metastable* with respect the fluid. This is an important contrast with the AHS model (see the next subsection) in which the ordered phase is present at $\beta\epsilon \to 0$. Crystallization in Sear's model is essentially an *energetic* transition to an ordered phase dominated by the directional interactions (compare the 'energetic fluid' concept introduced by Louis [27]) and not a continuation of the *entropic* HS freezing transition. The position of the sticky patches controls the crystal structure, and, as suggested by Sear, this may explain the relative ease or difficulty of crystallizing various proteins.

Figure 2(b) shows the phase behaviour for the model for the chosen parameter set. The solid lines in figure 2(b) are for the baseline model and include representative tie lines. As $\beta\epsilon$ increases, the fluid–crystal phase transition widens (the re-entrant fluid phase expected at larger ϕ is not shown). There is a metastable fluid–fluid phase separation for $\beta\epsilon \gtrsim 7.18$.

The effect of charges and added salt is obtained by inserting the baseline model into the general formalism in section 2. The dashed curve in figure 2(b) shows the fluid–crystal phase boundary at $Q = 10$ in the absence of salt. The transition has been markedly narrowed and the fluid–fluid transition moves to such a high value of $\beta\epsilon$ that it is no longer on the diagram. Repeating the calculation for $c_s^{(\mathrm{R})} = 0.2$ M obtains the chain lines in figure 2(b). The fluid–crystal phase boundary is intermediate between the zero-salt limit and the baseline model, and the metastable fluid–fluid transition has reappeared in the diagram for $\beta\epsilon \gtrsim 9.14$. Finally as $c_s^{(\mathrm{R})} \to \infty$, the phase boundaries all move back to coincide with the baseline model. Thus the effect of charge in the model is to strongly suppress existing phase transitions in the absence of added salt. Adding salt weakens and eventually destroys the effect. The salt concentration required to do this is $c_s \sim Qc_p$, as discussed already in section 2.

I now set $\beta\epsilon = 7.4$, marked by an arrow in figure 2(b), so that $B_2^{(0)}/B_2^{\mathrm{HS}} = -2.7$ reproduces the value calculated in section 3 above[3]. Figure 3(a) shows a typical phase diagram

[3] Oki *et al* [26] suggest 'macrobond' energies of the order 50 kcal mol^{-1} $\approx 80\,kT \gg \epsilon$ but this estimate is for the bare macrobond energy and does not include re-solvation of the protein surface.

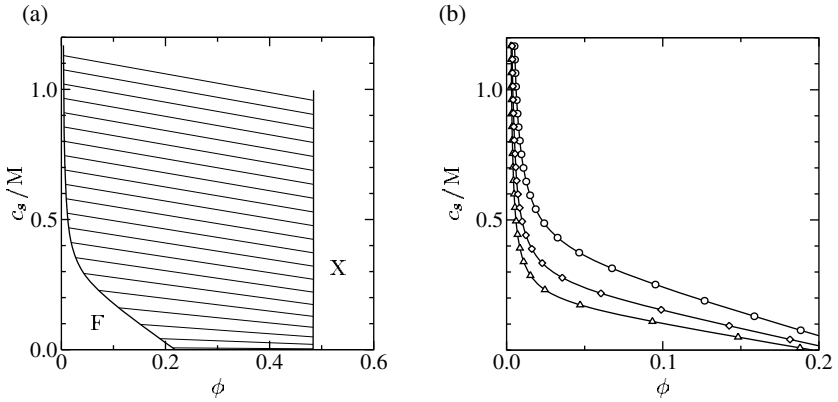

Figure 3. (a) Phase behaviour in the (ϕ, c_s) plane for Sear's model for $Q = 10$. (b) Fluid–crystal binodals for $Q = 8.0$ (triangles), 9.4 (diamonds), and 11.4 (circles). Other parameters are $n_s = 6$, $r_c = 1.05\sigma$, $\theta_c = 0.45$, $\beta\epsilon = 7.4$.

for the full model at this value of $\beta\epsilon$, as a function of added salt. Salt reservoir concentrations have been converted back to actual salt concentrations. The sloping tie lines indicate a salt repartitioning effect in which the coion concentration is enhanced in the more dilute phases. In this simple model I have not taken into account excluded volume in the dense phases, so in reality the salt repartitioning would be more marked (see below for a further discussion on this). In this representation we can clearly see how adding salt broadens the fluid–crystal coexistence. Metastable fluid–fluid phase separation does not appear until salt concentrations $\gtrsim$2.3 M for this value of $\beta\epsilon$. Note that the shape of the fluid–crystal binodal (which I interpret as being the same thing as the crystallization boundary) is in agreement with common experience. This is a natural but non-trivial consequence of charge neutrality in the present model.

This calculation is now repeated for $Q = 8.0$, 9.4, and 11.4, which are the three values of lysozyme charge examined experimentally by Poon *et al*. Figure 3(b) shows the crystallization boundary as a function of added salt, for the three values of Q. Clearly, the higher the charge, the more the phase transition is suppressed. Finally, these same phase boundaries are replotted in figure 4(a) as a function of the scaling variable c_s/Q^2. In this representation, the curves all collapse to lie on approximately the same quasi-universal crystallization boundary. The scaling collapse is robust: if the calculations are repeated for different parameter values, the quasi-universal crystallization curve moves up or down but a similar scaling collapse is always obtained.

The mean crystallization boundary from figure 4(a) is shown in 4(b), which now includes the experimental data from Poon *et al*. There is reasonable agreement on the location of the quasi-universal crystallization boundary, although it appears there is always some discrepancy between the shape of the theoretical and experimental boundaries at low ϕ which recalls the discrepancy in the slope of the high-salt law in figure 1(b). It is worth re-emphasizing that the present model is constrained to have the correct B_2, at least for $c_s \gtrsim 0.25$ M. The absence of a metastable fluid–fluid phase separation agrees with Poon *et al* who observe that no such phase transition occurs at the temperature of the experiments (22.5 °C). However, a temperature decrease of just a few per cent in the model, so that $\beta\epsilon = 7.7$ for example, is sufficient to bring fluid–fluid phase separation to accessible salt concentrations in the range 0.5–1.0 M. This is in accordance with the observations of Muschol and Rosenberger [28] (but additional caution is required since it is unlikely that the only effect of temperature is through the value of $\beta\epsilon$). In the scaling representation of figure 4(a), such metastable fluid–fluid binodals also collapse

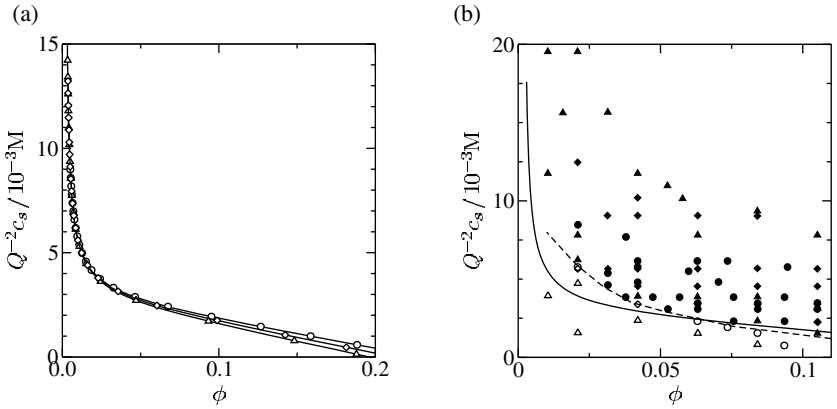

Figure 4. (a) The data of figure 3(b) replotted in the $(\phi, Q^{-2}c_s)$ plane. (b) Experimental data of Poon *et al* [10], for $Q = 8.0$ (triangles), 9.4 (diamonds), and 11.4 (circles). Filled (open) symbols indicate the occurrence (non-occurrence) of crystals. The dashed curve is the approximate experimental crystallization boundary. The solid curve is the mean crystallization boundary from (a).

to a quasi-universal curve—this is a prediction of the theory that would be interesting to test experimentally.

The scaling collapse in figure 4 occurs because the effects of charge and salt in fluid free energy are effectively combined into a single scaling variable c_s/Q^2, at salt concentrations in the vicinity of the crystallization boundary. To show that this is not just a feature of Sear's model, I now turn briefly to the more commonly studied AHS model.

5.2. Adhesive hard-sphere model

The AHS model is also a suitable baseline model for the general theory in section 2. In this model, HS interact with a short-range isotropic attractive potential. As noted by Rosenbaum *et al* [5], the phase behaviour is largely insensitive to the details of the potential provided that the second virial coefficient is used as the effective temperature axis. My approach to the baseline model here is closely based on that of Noro *et al* [29] who investigated the effects of long-range forces on the AHS model.

For the fluid phase, I use Barboy's treatment of Baxter's analytic theory [30]. In Baxter's theory [31], the attractive potential is characterized by a 'stickiness' parameter τ which is related to the second virial coefficient by[4]

$$B_2^{(0)}/B_2^{\mathrm{HS}} = 1 - (4\tau)^{-1}. \tag{17}$$

The crystal phase is expected to be FCC since this has the greatest density of intersphere contacts. As the stickiness is switched off ($\tau \to \infty$ or $\beta\epsilon \to 0$), the fluid–crystal phase transition goes over into the usual HS freezing transition, which is a contrast to Sear's model. The AHS ordered phase has always proved rather more difficult to treat analytically than the fluid phase, and approaches have ranged through density functional calculations by Marr and Gast [6], and Tejero and Baus [7], to detailed simulation studies [8]. Here I use the cell model of Daanoun *et al* [32] for the crystal phase free energy, assuming that the attractive part of the AHS potential is $-\epsilon(r/\sigma)^{-n}$ with $n \gg 1$. The cell model free energy is almost identical to that already used in equation (16) with $n_s = 12$ for the FCC structure. The difference is that there

[4] Barboy's definition of τ differs from the present use and Baxter's original work by a factor 6.

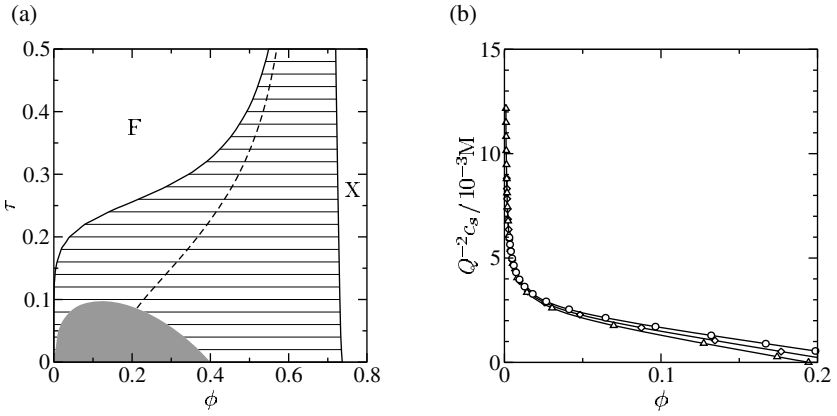

Figure 5. (a) Phase behaviour in the (ϕ, τ) plane for the AHS model with $n = 30$: solid curves are for the baseline model with $Q = c_s^{(\mathrm{R})} = 0$; the dashed curve is for $Q = 10$ and $c_s^{(\mathrm{R})} = 0$. The shaded region is the 'danger zone' equation (19). (b) Fluid–crystal binodals in the $(\phi, Q^{-2}c_s)$ plane for $Q = 8.0$ (triangles), 9.4 (diamonds), and 11.4 (circles); other parameters are $n = 30$, $\tau = 0.1$.

is no restriction on orientation so the term $-\log(\theta_c^3/\pi^2)$ is absent, and the role played by the *ad hoc* cut-off function is taken over by the functional form of the actual short-range potential, i.e. $w[x] \to (a/\sigma)^{-n}$. Like the choices for the cut-off function in the previous section, the actual value of n may shift the phase boundaries but does not change the broad picture. For the purposes of the present calculation I set $n = 30$.

A link between the cell model for the crystal phase and the Baxter–Barboy treatment of the fluid phase can be made by matching B_2 of equation (17) with the exact expression

$$B_2^{(0)}/B_2^{\mathrm{HS}} = 1 + 3\int_1^\infty \mathrm{d}x\, x^2[1 - \exp(-\beta\epsilon x^{-n})]. \tag{18}$$

Figure 5(a) shows the phase behaviour for the model. The solid curve in this figure is the fluid–crystal phase boundary for the baseline model, and the dashed curve is the corresponding boundary for $Q = 10$ with no salt. Similarly to the case for figure 2(b), the effect of charge is to narrow the fluid–crystal transition. As salt is added (not shown here), the boundary moves back towards that of the baseline model.

There is a complication that arises for Baxter's solution for the AHS fluid phase free energy. Baxter's theory involves the solution of a quadratic equation. If

$$\tau < \frac{(12\phi + 6\phi^2)^{1/2} - 6\phi}{6(1-\phi)}, \tag{19}$$

then this quadratic equation has complex roots and Baxter's theory becomes inadmissible. Normally this 'danger zone', shown shaded in figure 5(a), is happily hidden within the fluid–crystal two-phase region. In the charged version though, it emerges into the single-phase fluid region. For this reason one cannot choose $\tau < (2 - \sqrt{2})/6 \approx 0.098$, the maximum of equation (19), and unfortunately this excludes $\tau = 0.067$ which would match $B_2^{(0)}/B_2^{\mathrm{HS}} = -2.7$ from section 3.

To demonstrate the scaling collapse for this model, I therefore choose $\tau = 0.1$ (equivalent to $B_2^{(0)}/B_2^{\mathrm{HS}} = -1.5$ [24]). I repeat the calculations of the previous section to obtain the crystallization boundaries (fluid–crystal binodal) as a function of added salt, for $Q = 8.0$, 9.4, and 11.4 as used previously. Figure 5(b) shows these boundaries plotted using the scaling variable c_s/Q^2. Again the curves collapse to a quasi-universal crystallization boundary,

Table 2. Lysozyme salt repartitioning data from Palmer *et al* [33]. The solution is 0.855 MNaCl/0.2 M NaAc buffer (pH = 4.5). Ion concentrations in the third column are calculated by multiplying those in the second column by $1/(1-\phi)$, where $\phi = 0.64$ is the crystal volume fraction. The charge unaccounted for in the crystal, assuming $Q = 11.4$, is 0.05 M or $\lesssim$3% of the total charge.

Concentration/M	Solution	Crystal	Crystal free volume
Lysozyme^{Q+}		0.056	
Ac^-	0.2		
Na^+	1.055	0.19	0.53
Cl^-	0.855	0.78	2.2
$Na^+ \times Cl^-$	0.9		1.2

similarly to those in figure 4(a). This result confirms that the scaling collapse is a common feature of two different baseline models for the phase behaviour.

6. Salt repartitioning and interface structure

These ideas have some additional interesting physical implications for the equilibrium between crystals and solution which are explored in this section. The first implication concerns salt repartitioning. In the simple theory of section 2, the product of coion and counterion concentrations is a constant as in equation (3). Each phase has to be electrically neutral, so a phase enriched in protein is also enriched in counterions, and consequently *depleted* in coions (the Donnan common ion effect). This explains the slope of the tie lines in figure 3(a) for example. We can ask: is there any independent evidence for this phenomenon? Let me preface the answer to this by some cautionary remarks. Deviations from the simple Donnan equilibrium result may arise from three sources. Firstly, excluded-volume effects mean that one should really consider the small ion concentrations in the available free volume. This will be an important consideration in the crystal phase where the volume fraction occupied by the protein is $\sim$50%. Secondly, deviations from ideal solution behaviour can be expected at high salt concentrations—such deviations are usually absorbed into 'activity coefficients'. Thirdly, there may be significant specific ion effects, such as those seen in the Hoffmeister series.

Repartitioning of salt was studied in detail by Vekilov *et al* [34]. They found non-uniform salt concentrations inside protein crystals, but consider that this is likely to be an effect of impurities or growth kinetics. The present theory only addresses the equilibrium repartitioning of salt in perfect lysozyme crystals though. Fortunately, some relevant equilibrium results for salt repartitioning can be found in recent experimental reports by Morozova *et al* [35] on cross-linked lysozyme crystals in contact with salt solutions. For example, table 2 shows concentrations of Na^+ and Cl^- in solution and in the crystal from the early work of Palmer *et al* [33], cited by Morozova *et al*. There is marked repartitioning; for example the coion concentration in the crystal free volume is about half that in the external solution. The product of the two ion concentrations is approximately constant though (final row in table 2), provided that the protein excluded volume is taken into account. Detailed calculations by Morozova *et al* on their more recent data also take into account activity coefficients and give excellent agreement between theory and experimental results: they conclude that '... for small ions capable of penetrating into the crystal channels [the] electrostatic (Donnan) potential controls the equilibrium internal concentration of ions in just the same way as in polyelectrolyte gels' [35]. By way of contrast, the same work also suggests that Br^- has a significant specific interaction with lysozyme.

The second implication concerns the electrical structure of the interface between crystal and solution phases. In a Donnan membrane equilibrium [13–15], an electrostatic potential difference, the *Donnan potential*, develops between the two compartments on either side of the membrane. In the present theory for protein crystallization, it is still true that a difference in the mean electrostatic potential develops between the crystals and the solution phase. Whether this should also be called a Donnan potential, a *Galvani potential*, or perhaps something else, can be debated [36, 37].

At any rate, the potential difference is readily calculated given the salt repartitioning. If c_+ and c_- are the coion and counterion concentrations respectively, then $c_\pm = c_s^{(\mathrm{R})} \exp[\mp\beta e\overline{\varphi}]$, where $\overline{\varphi}$ is the mean electrostatic potential in the phase of interest measured relative to the salt reservoir (this result can also be used to recover equation (3)). The difference in mean potential between fluid (F) and crystal (X) phases, $\Delta\varphi = \overline{\varphi}^{(\mathrm{X})} - \overline{\varphi}^{(\mathrm{F})}$, is $\Delta\varphi = (\beta e)^{-1} \log(c_s^{(\mathrm{F})}/c_s^{(\mathrm{X})})$. For example, for the salt concentrations in table 2, $c_s^{(\mathrm{F})} = 1.055$ M, $c_s^{(\mathrm{X})} = 0.53$ M (in the free volume), and therefore $\Delta\varphi \approx 17$ mV. The potential difference arises because an electrical double layer is formed at the crystal–solution interface and is intimately connected with salt repartitioning. The details of this will be discussed in a forthcoming paper [38]. Another consequence of the Donnan potential is that the pH in the crystal will be $0.434\,\Delta\varphi$ higher than in the solution (where $0.434 = \log_{10} e$). If the solution pH is below the isoelectric point, as is often the case for lysozyme, one might expect this to lower the charge per protein in the crystal although it is unlikely that the charge can be determined with any accuracy from such a naive calculation.

Pair potential theories miss both the effects of salt repartitioning and the potential difference between the two phases, which are essentially many-body phenomena [39, 40]. Whilst this is not a problem at high salt levels where Hill's mapping between the McMillan–Mayer/Debye–Hückel pair potential approach and Donnan's method goes through [18], the effects can be significant at lower salt concentrations. The importance of the effects can be gauged by comparing the magnitude of $\Delta\varphi$ to the thermal energy kT. If $\Delta\varphi \gtrsim kT/e$, then there may be significant errors introduced by using a single effective pair potential (such as a DLVO potential) for both the crystal and the solution phases. On the other hand, if $\Delta\varphi \ll kT/e$, the use of pair potentials cannot be criticized on these grounds. It can easily be shown that $\Delta\varphi \gtrsim kT/e$ corresponds to the by now familiar $c_s \lesssim Qc_p$, in other words salt concentrations smaller than the protein charge density in the crystal, which is typically 0.5 M.

7. Conclusions

There are several conclusions from the present work. The first concerns the lysozyme/NaCl system. The quasi-universal crystallization boundary observed by Poon *et al* [10] can be fitted reasonable well by the present version of Sear's model with ∼6 sticky contacts/molecule, similarly to the conclusions of other workers [25, 26]. The scaling collapse for B_2 noticed by Egelhaaf and Poon [12] is here attributed to a contribution $Q^2/4c_s$ added to a 'bare' $B_2^{(0)}$ which is approximately *independent* of salt concentration above about 0.25 M. In this system therefore, NaCl appears to be acting as an *indifferent electrolyte* in the sense that it does not seem to exhibit specific ion effects. This appears to be confirmed by the experiments of Morozova *et al* [35] discussed in the preceding section. Other salts, for example NaBr, may not behave the same way of course.

The other conclusions are general ones. Firstly, a simple extension of existing models to incorporate salt ions and charge neutrality provides a straightforward explanation for the shape of protein crystallization boundaries and the associated scaling properties seen for lysozyme. Even if the present theory proves inadequate to describe the experiments in quantitative detail, this is surely a robust observation. Secondly, there are a large number of effects which have

been omitted from the present theory, but which could be incorporated with some additional numerical effort, such as excluded-volume effects, non-ideality of the salt ions, and other electrostatic correlation effects. Finally, the twin phenomena of salt repartitioning and the concomitant appearance of a significant Donnan potential difference discussed in the previous section are essentially many-body effects, and are not captured in simple pair potential theories.

Acknowledgments

For discussions and correspondence, I thank S U Egelhaaf, M E Fisher, H N W Lekkerkerker, M G Noro, W C K Poon and R P Sear.

References

[1] McPherson A 1999 *Crystallisation of Biological Macromolecules* (New York: Cold Spring Harbor)
[2] Piazza R 2000 *Curr. Opin. Colloid Interface Sci.* **5** 38
[3] Poon W C K 1997 *Phys. Rev.* E **55** 3762
[4] Lekkerkerker H N W 1997 *Physica* A **244** 227
[5] Rosenbaum D, Zamora P C and Zukoski C F 1996 *Phys. Rev. Lett.* **76** 150
[6] Marr D W and Gast A P 1993 *J. Chem. Phys.* **99** 2024
[7] Tejero C F and Baus M 1993 *Phys. Rev.* E **48** 3793
[8] Hagen M H J and Frenkel D 1994 *J. Chem. Phys.* **101** 4093
[9] Sear R P 1999 *J. Chem. Phys.* **111** 4800
[10] Poon W C K, Egelhaaf S U, Beales P A, Salonen A and Sawyer L 2000 *J. Phys.: Condens. Matter* **12** L569
[11] Guo B, Kao S, McDonald H, Asanov A, Combs L L and Wilson W W 1999 *J. Cryst. Growth* **196** 424
[12] Egelhaaf S U and Poon W C K, private communications
[13] Donnan F G 1911 *Z. Elektrochem.* **17** 572
[14] Rice S A and Nagasawa M 1961 *Polyelectrolyte Solutions* (London: Academic)
[15] Morawetz H 1965 *Macromolecules in Solution* (New York: Wiley–Interscience)
[16] McMillan W G and Mayer J E 1945 *J. Chem. Phys.* **13** 276
[17] Belloni L 2000 *J. Phys.: Condens. Matter* **12** R549
[18] Hill T L 1956 *Discuss. Faraday Soc.* **21** 31
[19] Brodie M L, Tominc T M and Saxowsky M D 1996 *Phys. Rev.* E **53** 6325
[20] Warren P B 1997 *J. Physique* II **7** 343
[21] Piazza R and Pierno M 2000 *J. Phys.: Condens. Matter* **12** A443
[22] Rosenbaum D F, Kulkarni A, Ramakrishnan S and Zukoski C F 1999 *J. Chem. Phys.* **111** 9822
[23] Velev O D, Kaler E W and Lenhoff A M 1998 *Biophys. J.* **75** 2682
[24] Vliegenthart G A and Lekkerkerker H N W 2000 *J. Chem. Phys.* **112** 5364
[25] Curtis R A, Blanch H W and Prausnitz J M 2001 *J. Phys. Chem.* B **105** 2445
[26] Oki H, Matsuura Y, Komatsu H and Chernov A A 1999 *Acta Crystallogr.* D **55** 114
[27] Louis A A 2001 *Phil. Trans. R. Soc.* A **359** 939
[28] Muschol M and Rosenberger F 1997 *J. Chem. Phys.* **107** 1953
[29] Noro M G, Kern N and Frenkel D 1999 *Europhys. Lett.* **48** 332
[30] Barboy B 1974 *J. Chem. Phys.* **61** 3194
[31] Baxter R J 1968 *J. Chem. Phys.* **49** 2770
[32] Daanoun A, Tejero C F and Baus M 1994 *Phys. Rev.* E **50** 2913
[33] Palmer K J, Ballantyne M and Galvin J A 1948 *J. Am. Chem. Soc.* **70** 906
[34] Vekilov P G, Monaco L A, Thomas B R, Stojanoff V and Rosenberger F 1996 *Acta Crystallogr.* D **52** 785
[35] Morozova T Ya, Kachalova G S, Lanina N F, Evtodienko V U, Botin A S, Shlyapnikova E A and Morozov V N 1996 *Biophys. Chem.* **60** 1
[36] Fisher M E, private communications
[37] Sparney M J 1972 *The Electrical Double Layer* (Oxford: Pergamon)
[38] Sear R P and Warren P B 2002 *Preprint* cond-mat/0206067
[39] Langmuir I 1938 *J. Chem. Phys.* **6** 873
[40] Warren P B 2000 *J. Chem. Phys.* **112** 4683