Multishell Structures of Virus Coat Proteins

Peter Prinsen,*,†,¶ Paul van der Schoot,‡ William M. Gelbart,† and Charles M. Knobler†

Department of Chemistry and Biochemistry, University of California, Los Angeles, 607 Charles E. Young Drive East, Los Angeles, California 90095-1569, and Group Theory of Polymers & Soft Matter, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands

Received: November 20, 2009; Revised Manuscript Received: March 12, 2010

Under conditions of low ionic strength and a pH ranging between about 3.7 and 5.0, solutions of purified coat proteins of cowpea chlorotic mottle virus (CCMV) form spherical multishell structures in the absence of viral RNA. The outer surfaces of the shells in these structures are negatively charged, whereas the inner surfaces are positively charged due to a disordered cationic N-terminal domain of the capsid protein, the arginine-rich RNA-binding motif that protrudes into the interior. We show that the main forces stabilizing these multishells are counterion release combined with a lower charge density in the RNA-binding motif region of the outer shells due to their larger radii of curvature, arguing that these compensate for the outer shells not being able to adopt the smaller, optimal, radius of curvature of the inner shell. This explains why the structures are only stable at low ionic strengths at pHs for which the outer surface is negatively charged and why the larger outer shells are not observed separately in solution. We show how to calculate the free energy of shells of nonoptimal radius of curvature from the elastic properties of the native shell. The spacing between shells is determined mainly by the entropic elasticity of the RNA-binding motifs. Although we focus on CCMV multishells, we also predict the solution conditions under which multishells formed by CCMV coat protein mutants with a lower RNA-binding motif charge are stable, and we examine other viruses as well. We conclude that at a given surface charge density, the boundaries separating regions of stable multishells with different numbers of shells shift to lower ionic strengths upon either increasing the length of the RNA-binding motif, increasing the stiffness of the shells, or decreasing the charge per RNA-binding motif.

1. Introduction

The simplest viruses consist of a single-stranded RNA genome that is enclosed in a container, the capsid, which consists of many copies of a single protein. Some, like tobacco mosaic virus (TMV), are cylindrical, while others, like cowpea chlorotic mottle virus (CCMV), are icosahedral. In 1955, Fraenkel-Conrat and Williams[1] showed that infectious particles of TMV can spontaneously self-assemble when the purified RNA and protein are mixed in a buffer solution at an appropriate pH and ionic strength. Twelve years later, Bancroft and Hiebert[2] demonstrated that as in TMV, CCMV RNA and protein self-assemble into infectious viruses, indistinguishable from the wild type. Further, Bancroft et al.[3] demonstrated that in the absence of RNA, CCMV capsid proteins self-assemble into a variety of structures, including capsids, tubes, and sheets. They sketched an ionic strength/pH phase diagram that showed the conditions of stability of each polymorph; a closely similar diagram for CCMV capsid protein was later determined by Adolph and Butler.[4]

The CCMV protein phase diagram has recently been re-examined by Lavelle et al.,[5] who also showed that the protein assemblies are equilibrium structures, that is, reversible. As in the earlier studies, empty capsids are observed at pH < 3.7 and at high ionic strengths; but when the pH is increased at low ionic strength, multishell structures assemble. They consist of an empty capsid equal in size to the native capsid, surrounded by one or more larger shells that each consist of a single layer of protein. Multishells are not observed at pH > 5.0; tubular structures are found instead.

Lavelle et al.[5] suggested that the appearance of the multishells may be understood in terms of a balance between electrostatic interactions and the curvature energy of the protein shells. The 26 N-terminal amino acids of the CCMV capsid protein are positively charged because of the presence of basic residues. This arginine-rich motif (ARM) extends into the interior of the capsid where it can interact with the negatively charged RNA. The p*K*s of the ARM residues are high; therefore, they remain essentially fully charged until pH > 8. On the other hand, the charge on the capsid exterior varies with pH. It becomes negative when the pH exceeds the isoelectric point, 3.7 (a value that is independent of the charge on any cargo in the capsid, such as RNA), and then increases in magnitude and levels off at a pH of about 5.[6] Lavelle et al. argued that the formation of additional shells can be attributed to the electrostatic interaction between the negative shell exterior and the positive ARM protein tails on the interior of the shell. CCMV capsids have a radius of 14 nm associated with the spontaneous curvature of the protein. Successive shells have smaller curvatures (larger radii) and a corresponding cost in curvature energy. The number of shells that form must then be limited by the balance between these two effects, the stabilizing effect of electrostatic interaction between successive shells and the energy cost associated with reduced curvature.

While these arguments seemed plausible, they are purely qualitative. We have therefore undertaken a more rigorous theoretical study to investigate their validity. We focus on the

* To whom correspondence should be addressed. E-mail: prinsen@lorentz.leidenuniv.nl.
† University of California, Los Angeles.
‡ Eindhoven University of Technology.
¶ Current address: Instituut-Lorentz for Theoretical Physics, Leiden University, P.O. Box 9506, 2300 RA Leiden, The Netherlands.

10.1021/jp911040z
Published on Web 04/06/2010

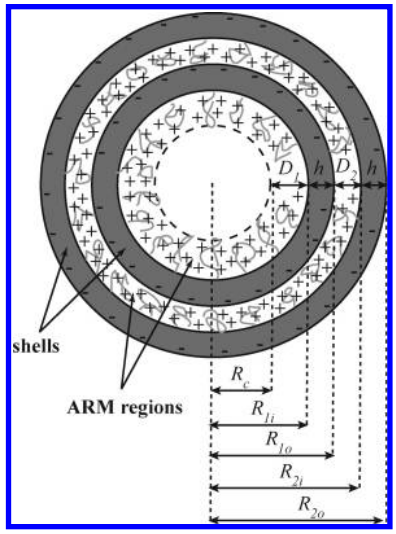

Figure 1. Model of a multishell structure, in this case consisting of two shells of thickness h. The outside surfaces of shells are negatively charged, and the arginine-rich binding motifs (ARMs), of width D_1 and D_2, are positively charged and protrude on the inside of the shells. $R_{1i(o)}$ is the inner (outer) radius of shell 1, and similarly, $R_{2i(o)}$ is that of shell 2; $R_c = R_{1i} - D_1$.

relative stabilities of single shells and multishells, and we develop a model to calculate the free energy per protein in each of those structures. For a given structure, we optimize the spacing between shells to find the lowest free energy per protein for that particular structure, which we subsequently use to determine the most stable one. We assume the shells are spherical and of constant thickness and made from a material of low dielectric constant. The charges on the shell exteriors are taken to be uniformly distributed, and the ARMs are assumed to form a layer of uniform charge density and to fill the entire region between shells (see Figure 1). Because this charge density is high, we solve the Poisson−Boltzmann equation to find the electrostatic potential from which we can then calculate the electrostatic free energy. We employ a simple entropic spring model to describe the stretching or compression of the layer of ARMs, and we use continuum mechanics to determine the free-energy penalty associated with the deformations of the shells away from their favored radius of curvature.

The width of the region between two shells is determined by minimization of the free energy and is generally close to the equilibrium width of the entropic spring free energy, typically a few nm. Increasing the width of this region, by increasing the radius of the outer shell, decreases the charge density of the ARM region and thus the electrostatic free energy, whereas it increases the deformation free energy of the outer shell (since the radius of curvature of that shell increases). Decreasing the width has the reverse effect. The electrostatic free energy has two main components, namely, the entropy of the free ions and the interaction of the surface charges with the electric field. The lower the ionic strength, the larger the entropy difference between ions between two shells and ions in the solution. This means that at low ionic strength the electrostatic free energy is the main component of the free energy, and the ARMs are stretched; at high ionic strength, the width of the ARM region is close to the equilibrium value. In other words, the deformation free energy of the shells favors an ARM region as narrow as possible because it favors a radius of curvature of a shell as close as possible to that of the native capsid, whereas electrostatics favor a region as wide as possible. The entropic elasticity of the ARMs prevents the region from becoming too narrow or too wide; it essentially determines its width.

Shells with a large radius of curvature have a large deformation free energy. Stable multishell structures can form because two effects cause a decrease in the electrostatic free energy. First, if the width of the ARM region is held constant, the charge density in that region decreases as the radius of curvature of the shell increases, leading to a decrease in the electrostatic free energy. Second, the free energy includes a contribution from the interaction between the positively charged ARMs and the charged outside surface of a shell. If the outside surface is negatively charged, the overlap of its electric double layer with the positively charged ARM region leads to release of counterions, which decreases the electrostatic free energy. The decrease in electrostatic free energy turns out to be large enough to offset the deformation free energy as long as the ionic strength is sufficiently low.

We find that for a certain range of surface charge densities under any solution condition (i.e., pH and ionic strength) only one type of multishell is dominant and that sharp boundaries exist between regions in which multishells with different numbers of shells are stable. Increasing the stiffness of the shells shifts the boundaries between single- and multishell structures to lower ionic strengths at fixed surface charge density because it leads to an increase in the elastic deformation energy which has to be compensated for by a larger difference in electrostatic free energy between the two types of structures at the boundary. This can be achieved by lowering the ionic strength. Increasing the length of an ARM has the same effect because it increases the equilibrium separation between shells, which effectively increases their radius of curvature (except for the central capsid), leading to a larger deformation energy of the shell and also to a smaller charge density between shells. Increasing the charge per ARM has the opposite effect. It increases the charge density between shells and thus the electrostatic free-energy difference between two structures at a boundary. An increase in ionic strength compensates for this; therefore, boundaries are shifted to higher ionic strengths.

In the following sections, we present our theoretical treatment of the ARMs, of the relevant electrostatic effects, and of shell elasticity contributions. In section 3, we calculate phase diagrams, that is, regions of pH and ionic strength where these competing structures are successively dominant. Our results for CCMV capsid protein are compared directly with experimental results, and predictions are made for CCMV mutants and for other viral coat proteins.

2. Theory

Here, we develop a model to calculate the free energy per protein in an assembly consisting of an arbitrary number of shells. This free energy is subsequently used to determine the most stable assembly, that is, the one with the lowest free energy per protein, and the spacing between shells in that assembly. Figure 1 shows a sketch of our model of a multishell, which in this case consists of two shells. Although CCMV and most other viruses have icosahedral symmetry, for simplicity, we assume that each shell is spherical and of constant thickness h. We assume, as well, that a shell has a uniform dielectric constant $\varepsilon_{\text{protein}}$ that is much smaller than that of the surrounding medium and that no charges are present within the shell itself. The charge on the capsid exterior is taken to be uniform; that on the interior surface is small compared to that on the positively charged ARMs and can be neglected. We assume that the ARMs fill the entire region between shells and that the charge density due to them is uniform.

The total free energy F_{tot} of a multishell consists of three parts

$$F_{tot} = F_s + F_{elec} + F_{elas} \quad (1)$$

F_s is the entropic elastic free energy associated with stretching or compression of the layer of ARMs. We use a simple ideal chain elasticity to model this (see section 2.1). F_{elec} is the electrostatic free energy associated with the charges on the ARMs, the counterions, and the charges on the outside surface of a shell. Due to the high charge density between shells, we need to solve the Poisson−Boltzmann equation to determine the electrostatic potential, from which we subsequently calculate the free energy (for a similar case, see ref 7). Instead of solving the Poisson−Boltzmann equation numerically over the whole region, we calculate a numerical solution inside of the inner shell and then derive an approximate analytical solution in the rest of the region (described in section 2.2). The third contribution is F_{elas}, which is associated with the local elastic deformations of a shell due to the fact that proteins in the shells surrounding a capsid in a multishell structure experience a larger local radius of curvature than those in the capsid. We assume that the shell can be modeled as a continuous elastic medium, and we then show how to calculate this elastic free energy from the Young's modulus and Poisson ratio of the native shell (see section 2.3). Finally, we discuss the total free energy (see section 2.4). Because the multishells contain different numbers of proteins, we must, in principle, use the law of mass action to calculate the relative abundances of the structures. As shown in section S1 of the Supporting Information, however, taking into account the law of mass action changes the results only slightly.

2.1. Entropic Elasticity of ARMs. We model the ARMs as freely rotating chains consisting of three segments per amino acid, each of length $a = 0.144$ nm[8] and with a bond angle of 116°[8] between segments; their typical length is about 10 to a few tens of nm, and their collective properties are described by polyelectrolyte brush theory.[9] Specifically, we describe the ARM region as an osmotic brush.[10] In other words, the equilibrium size of the layer is determined by the osmotic pressure of the counterions and the entropic elasticity of the ARMs. Since the contribution of the counterion osmotic pressure to the free energy will be included in the electrostatic free energy, we need examine only the entropic elasticity of the ARMs. We use the familiar free energy for an ideal chain[11]

$$f = \frac{3R_e^2}{2R^2} + \frac{3R^2}{2R_e^2} \quad (2)$$

where R is the end-to-end distance, R_e^2 is the average square end-to-end distance of the isolated chain in equilibrium, that is, with free ends, and where we have omitted an irrelevant constant term. For a freely rotating chain consisting of n segments of length a with a bond angle of 116° between segments,[12] $R_e^2 \approx 2.56na^2$. We set $R = D_i$ and $R_e = D_0$. Then, the total free energy due to stretching of the ARMs in the ith shell is

$$F_{s,i} = \frac{3}{2}N_i\left(\frac{D_0^2}{D_i^2} + \frac{D_i^2}{D_0^2}\right) \quad (3)$$

where N_i is the number of coat proteins in the shell and D_i is the width of the ARM region of the ith shell. D_i is different, in principle, for different shells. We come back to this assumption when we discuss the electrostatic part of the free energy in the next section.

As we will see in section 3, the amino acid density in the ARM region in the case of CCMV is about 14% of that in the protein shell itself, which is close to the maximum ("close-packed") value. This means that excluded volume effects might play a role. It turns out, however, that at high ionic strength, when the electrostatic free energy is relatively small, the width of the ARM region is close to the equilibrium width. Changing the shape of the free energy but keeping the equilibrium width the same will then have little influence on the results. At low ionic strength, the ARMs are generally stretched, especially if they are highly charged. Changing the shape of the free energy changes the amount of stretching, but it does not change the results qualitatively; it slightly shifts the boundaries between regions of stability of different structures. The same is true for relatively small changes in the equilibrium spacing D_0. The ARMs are relatively short, and the amino acid density is quite high; therefore, as in the case of concentrated polymer solutions and melts, the excluded volume interactions are not likely to be important. Taking into account these considerations, the simple expression eq 3 is sufficiently accurate for our purposes.

2.2. Electrostatics. In our model, the shells are taken as spherical, and all charge distributions are assumed to be spherically symmetric, which implies that the electrostatic potential Φ is spherically symmetric. In the mean-field (Poisson−Boltzmann) approximation, in which the ions are assumed to be point charges, the electrostatic free energy consists of two contributions. The first is the energy of the electric field or, equivalently, the energy of the charges in the electric field,[13] and the second contribution comes from the translational entropy of the free ions. Because the calculation of the electrostatic free energy is relatively straightforward, we give only an outline of the calculation in this section, with emphasis on some important issues, and we refer to section S2 of the Supporting Information for details. The total electrostatic free energy is given by eq S.6 in the Supporting Information.

In order to calculate the electrostatic free energy, we need to know the electrostatic potential in the whole region. Although the negative surface charge on the outside of a shell is small (with a typical surface charge density smaller than 0.1 e nm^{-2}), the electrostatic potential is still large due to the charged ARMs inside of the native capsid and between the shells (with a typical volume charge density of about 0.3 e nm^{-3}). A treatment of the electrostatics at the Debye−Hückel level is therefore not justified; to find the electrostatic potential, we need to solve the nonlinear Poisson−Boltzmann equation. We follow ref 14 and assume that the charge density in the ARM region due to the ARMs is constant, which simplifies the calculations considerably. We can assume that the ARMs fill the whole region between shells because for a given width D of the ARM region of a shell (other than the capsid), the lowest free energy is obtained when the separation between a shell and the surrounding ARM region is zero. If this separation were to be increased, the radius of curvature of the outer shell would have to increase, increasing the elastic deformation energy of that shell (see section 2.3). At the same time, the overlap between the electric double layer of the negative exterior of the inner surface and the ARM region would decrease, increasing the electrostatic free energy. Note that an increase in the radius of curvature causes a slight decrease of the charge density in the ARM region, leading to a slight decrease in the electrostatic free-energy density. However, the slight increase in volume of the ARM region increases the free energy per protein. (We have

checked this numerically.) The net effect would thus be an increase in free energy, showing that the ARMs indeed fill the entire region between shells. In general, the ARM region within the central capsid (the inner shell) will not extend to the center.

We assume that the mobile salt ions do not penetrate the protein shells themselves and that no charged groups are present in the protein shell itself. The electrostatic potential in the protein shell then obeys the Laplace equation. Salt can move freely from one side of a shell to the other because shells contain pores that are large enough for the salt ions to diffuse through. We assume that these pores do not influence the electrostatic energy of the assembly significantly, and we do not model them explicitly. At the boundary between a protein shell and the surrounding liquid, the potential is continuous but generally not differentiable. This discontinuity in the derivative is determined by the surface charge density and the dielectric permittivities of the two materials and can be calculated with Gauss's law. For the dielectric permittivity of the protein shells, we use $\varepsilon_{\text{protein}} = 5$. Note that the dielectric permittivity of proteins is a subject of much discussion.[15] In fact, the value of the dielectric constant depends on the model in which it is used; the more the dielectric response is modeled explicitly (e.g., conformational changes due to ionization), the lower the dielectric constant that one needs.[15,16] We find that in our case, the exact value influences the results very little as long as it is small compared to the dielectric permittivity of water $\varepsilon_{\text{water}} = 80$.

There is no analytical solution for the electrostatic potential in our model, but instead of constructing a numerical solution over the whole region, we construct a numerical solution inside of the innermost shell, where the potential changes considerably, and an approximate analytical solution elsewhere. An advantage of this approach is that it is computationally less expensive. This is important here since we will minimize the total free energy numerically with respect to the separations D_i between shells, which means the electrostatic potential has to be calculated very many times.

Figure 2 shows a typical numerical solution of the dimensionless potential $\phi \equiv e\Phi/k_BT$, in this case for a multishell consisting of two shells. As one can see, the potential between shells is quite high, but it is more or less constant. In fact, numerical solution of the problem shows that the dimensionless potential between shells $i - 1$ and i is approximately equal to $\operatorname{arcsinh}(\rho_i/2c_0)$, where ρ_i is the charge density (in units of e) between the shells and c_0 is the salt concentration (see, for example, Figure 2). This result can also be derived by assuming that the total charge between the shells (including the surface charge) is zero. Because the surface charge is much smaller than the charge in the ARM region, we can neglect it and find $\overline{2c_0 \sinh \phi} \approx \rho_i$, where the bar denotes an average over the whole ARM region. If ϕ is approximately constant, this leads to $\phi \approx \operatorname{arcsinh}(\rho_i/2c_0)$.

The way that we calculate the approximate analytical solution is as follows. Outside of the outer surface of the outermost shell (region VI in Figure 2), the potential is small, and we solve the Debye−Hückel equation. In the protein shells (regions III_1 and III_2), we use the exact solution of the Laplace equation; between shells (region II_2), we set $\phi = \operatorname{arcsinh}(\rho_i/2c_0) + \phi_1$, where ϕ_1 is a small correction that obeys the Debye−Hückel equation with a renormalized Debye length, which we solve. Inside of the inner surface of the inner shell (regions I and II_1), we solve the Poisson−Boltzmann equation numerically with boundary condition $\phi = \operatorname{arcsinh}(\rho_1/2c_0)$ on the inside surface of the inner shell, where ρ_1 is the charge density in the ARM region inside of the inner shell. We then calculate the small correction to this solution analytically in the outer half of the ARM region of the inner shell (region II_1) again by solving the Debye−Hückel equation with renormalized Debye length. (We assume that the correction is negligible elsewhere inside the inner surface of the inner shell.) The appropriate boundary conditions are derived from the original ones, and the total electrostatic free energy is calculated partly numerically and partly analytically from the electrostatic potential (see section S2 of the Supporting Information for details).

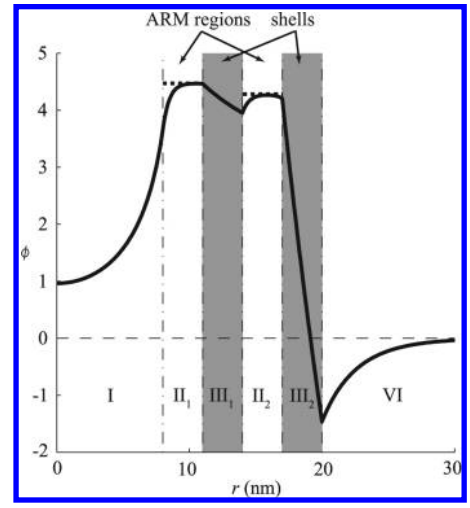

Figure 2. Example of the dimensionless electrostatic potential ϕ as a function of distance r from the center of the multishell structure for a structure consisting of two shells. The ionic strength is $I = 0.01$ M, the inner shell consists of $N_0 = 180$ proteins, each with charge (in units of e) $q_0 = 1$, the charge per ARM (in units of e) is $Q = 10$, and the equilibrium thickness of the ARM region is $D_0 = 2.0$ nm. The shell thickness is $h = 3$ nm, and the thicknesses of the ARM regions are $D_1 = D_2 = 3$ nm. The solid line denotes the electrostatic potential, and the dotted lines denote the approximate solution $\phi = \operatorname{arcsinh}(\rho_i/2c_0)$, where ρ_i is the charge density in the ARM region due to the ARMs and c_0 is the salt concentration.

Generally, increasing the charge per ARM increases the charge density and thus the free energy per protein. Increasing the width of the ARM region decreases the charge density, and even though the volume of the ARM region increases, it still decreases the free energy per protein in that shell (the electrostatic free-energy density scales roughly as ρ_i^2; therefore, the electrostatic free energy per protein in a shell is roughly proportional to ρ_i). Note that this differs from the effect mentioned in the second paragraph of this section. Even though the free energy per protein in a shell decreases, it is possible that the free energy per protein in all of the shells combined increases, which is the case there. Finally, increasing the absolute value of the surface charge, assuming that it is opposite in sign to that of the ARM, decreases the electrostatic free energy.

There are two separate mechanisms that cause additional shells to have a lower electrostatic free energy per protein than the central shell, at least when the width of the ARM region is the same. The first is counterion release. Due to overlap of the electric double layer of the negatively charged outer surface of a shell with the oppositely charged ARM region of the surrounding shell, positive- and negative-charged counterions can escape to the bulk, increasing their entropy and lowering the free energy. This can be seen, for example, in Figure 2. The concentration of positive counterions is $c_0 \exp(-\phi)$, and it is much higher at the negatively charged outside surface of the outer shell, where ϕ is negative, than close to the negatively charged outside surface of the inner shell, where ϕ is positive. The concentration of negative counterions ($c_0 \exp \phi$) between

the surfaces is quite large but still smaller than the charge density due to the positive ARMs (ϕ is smaller than $\text{arcsinh}(\rho_i/2c_0)$), especially close to the negatively charged surface. The reason that the concentration of negatively charged counterions is still large is that the total fixed positive charge (on the ARMs) is much larger than the total fixed negative charge (on the surface). The second reason why additional shells have a lower electrostatic free energy is that additional shells have a larger radius of curvature than the inner shells, which means that the charge density due to the ARMs is smaller at the same ARM region width. As mentioned before, even though the volume is larger, the net effect is a decrease in electrostatic free energy.

If we let the width of the ARM region go to zero, we can compare our calculation of the free energy of a single shell with that of Šiber and Podgornik,[17] who calculated the electrostatic free energy of a shell with surface charge on both the inside and the outside. Our results agree with theirs.

2.3. Elastostatics. An important experimental observation is that in the case of CCMV, the only single shells that are stable in the pH range that we study are native capsids. Large shells are found around smaller shells under certain solution conditions but never alone. This indicates that large shells are less stable (have a higher free energy per protein) than native capsids. We assume that this is mainly due to an energetic penalty for elastic deformation of the shells because they have a radius of curvature that is larger than the preferred radius of curvature, which we assume is equal to the native capsid radius of curvature. We also assume that the shell can be viewed as a continuous elastic medium with Young's modulus E and Poisson ratio ν. The material is undeformed in the native capsid, and the deformation energy for a particular shell can be calculated by comparing the deformation with respect to the capsid. As explained in the Appendix, there is no stretching or compression of the elastic material in the neutral plane of a shell, which is the spherical surface that divides a shell into two shells of equal thickness, even if a shell does not have the preferred radius of curvature. The density of proteins in this neutral plane is assumed to be independent of the size of the shell. Note that this is an assumption since shells with larger T numbers would have a higher ratio of hexamers to pentamers and therefore a different density of proteins. (The T number refers to the Caspar−Klug[18] classification of the ways in which a 2D hexagonal lattice can close on itself with increasing distances between its 12 icosahedrally distributed five-fold disclinations. The number of protein subunits in each of the resulting structures is $60T$, with $T = m^2 + n^2 + mn$ and $m, n =$ positive integers.) Furthermore, even though the outer shells of multishells probably consist of hexamers and pentamers, we do not know how they are arranged, and it seems unlikely that the larger shells correspond to T number structures.[5]

The number of proteins in shell j is then proportional to the area of the midplane

$$N_j = N_1\left(\frac{R_{ji} + R_{jo}}{R_{1i} + R_{1o}}\right)^2 \tag{4}$$

where R_{ji} (R_{jo}) is the inner (outer) radius of the jth shell and the elastic free energy of the jth shell equals (see eq A.17)

$$F_{\text{elas},j} = \frac{\pi E h^3}{3(1-\nu)k_B T}\left(\frac{R_{ji} + R_{jo}}{R_{1i} + R_{1o}} - 1\right)^2 = \frac{1}{2}k_{\text{elas}}N_j\left(1 - \frac{R_{1i} + R_{1o}}{R_{ji} + R_{jo}}\right)^2 \tag{5}$$

The elastic constant is given by

$$k_{\text{elas}} \equiv \frac{2\pi E h^3}{3(1-\nu)N_1 k_B T} \tag{6}$$

with h, as before, the shell thickness, presumed constant. Note that the elastic constant is proportional to $(1-\nu)^{-1}$ and hence is different from the flexural rigidity or cylindrical rigidity,[19] which is proportional to $(1-\nu^2)^{-1}$. However, we are dealing with the deformation of a surface where the deformed surface does not contain the same amount of material as the (smaller) undeformed surface. See the Appendix for details.

2.4. Total Free Energy. The total free energy per protein in the single-shell capsid consists of a contribution from the electrostatics and from the stretching of the ARM region inside of the capsid. It is calculated using eqs 3, S.11, and S.16 (Supporting Information)

$$F_{\text{tot},1} = \frac{F_{\text{elec}} + F_s}{N_1} = \frac{F_{\text{elec},0} + F_{\text{elec},1} + F_{s,1}}{N_1} \tag{7}$$

where the thickness of the ARM region of the first shell D_1 minimizes this free energy. For a structure consisting of two shells, there is also a contribution from the elastic deformations in the outer shell. The free energy per protein is determined from eqs 3, 5, S.11, and S.24 (Supporting Information)

$$F_{\text{tot},2} = \frac{F_{\text{elec}} + F_s + F_{\text{elas}}}{N_1 + N_2} = \frac{F_{\text{elec},0} + F_{\text{elec},2} + F_{s,1} + F_{s,2} + F_{\text{elas},2}}{N_1 + N_2} \tag{8}$$

where, in principle, we should calculate the thickness of the two ARM regions D_1 and D_2 by minimizing this free energy. In practice, we can use the value of D_1 that minimizes eq 7 at the same pH and ionic strength so that we need only minimize with respect to the single quantity D_2. Similar expressions can be derived straightforwardly for the free energy of multishells consisting of more than two shells. In the case of M shells, we have to minimize the total free energy per protein with respect to D_i, $i = 2, ..., M$.

There are seven independent quantities in the model, the preferred size of the capsid R_0, the thickness of the capsid h, the number of proteins in the capsid N_1, the elastic constant of the capsid material k_{elas}, the number of amino acids in an ARM n (or the equilibrium spacing of the ARM region D_0), the total charge of an ARM Q, and the surface charge per protein q_0. However, because the law of mass action does not influence the results significantly, as shown in section S1 of the Supporting Information, we can replace the three quantities N_1, Q, and q_0 by N_1Q, and N_1q_0.

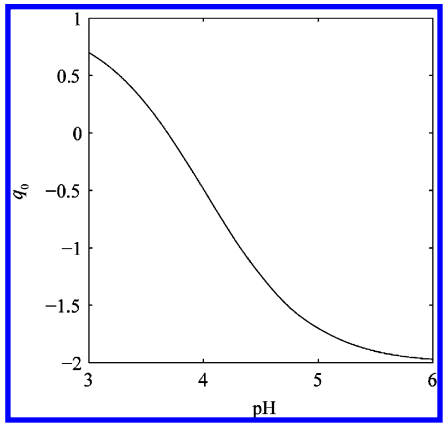

Figure 3. The surface charge per protein q_0 (in units of e) as a function of pH determined by estimating the total charge of the amino acids close to the outside surface using the Henderson−Hasselbalch equation with the bare pK values of the amino acids and the sequence from the protein data bank[26] (www.pdb.org).

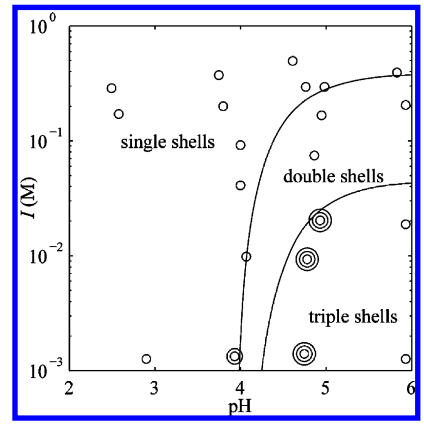

Figure 4. Diagram showing the majority type of structure formed by CCMV coat protein as a function of pH and ionic strength I (NaCl). The symbols denote experimental results from ref 5. Single circles: wild-type empty capsids (single shells); double circles: double shells; triple circles: triple shells. The curves are the results of our calculations. The top curve denotes the boundary at which the concentrations of single shells and double shells are equal. The lower curve denotes the boundary at which the concentrations of double shells and triple shells are equal. In both cases, we used the estimated values $k_{elas} = 31$ and $D_0 = 2.0$ nm, and the surface charge was calculated from the pK values of the amino acids close to the surface (see Figure 3).

3. Results and Comparison to Experiments

3.1. CCMV. We first look at multishells consisting of CCMV coat proteins. A native CCMV capsid consists of 180 identical proteins and has icosahedral symmetry, which means it is not exactly spherical. We therefore use average values for its inner radius, $R_{1i} = 10.4$ nm, and outer radius, $R_{1o} = 13.2$ nm, as determined in ref 20. The average width of a shell is then $h = 2.8$ nm. To model the elastic properties of the shells, we use the same Poisson ratio $\nu = 0.4$ as that in ref 20, and we use the Young's modulus $E = 280$ MPa determined in that paper from nanoindentation experiments.[21] Using these values in eq 6, we find $k_{elas} = 31$. If we use the values $E = 140$ MPa and $h = 3.8$ nm of ref 21, we find $k_{elas} = 39$. The nanoindentation experiments were performed at room temperature, whereas the assembly experiments were performed at 4 °C, but we assume that the Young's modulus is independent of temperature over this narrow range.

In principle, we can determine the surface charge of a shell from the electrophoretic mobility measurements on full and empty capsids performed by Johnson et al.[6] by using, for example, the first-order equation derived by Henry.[22] However, the measurements were performed at an ionic strength of 0.22 M, and the mobility at lower ionic strength (and the charge) can be as much as 3−4 times higher (see, e.g., ref 23).

An alternative method for estimating the surface charge is by using the Henderson−Hasselbalch equation with tabulated pK values of isolated amino acids to calculate the total charge of the amino acids at the capsid exterior. In the absence of electrophoretic mobility measurements made at low ionic strengths, we will employ this method. There are four amino acids that are potentially charged and close to the surface, two glutamic acids (residues 63 and 166, pK = 4.15), one lysine (residue 65, pK = 10.67), and one aspartic acid (residue 168, pK = 3.71).[24] These residues were identified using the crystal structure of CCMV (PDB ID: 1CWP[25]) from the Protein Data Bank[26] (www.pdb.org). Figure 3 shows the calculated surface charge as a function of pH.

The ARM consists of 26 amino acids,[27] corresponding to $n = 78$ and $D_0 = R_e(= (2.56na^2)^{1/2}) = 2.0$ nm, and the stretched backbone of an ARM is 9.5 nm long. The only amino acids in the ARM with charged side groups are six arginines and three lysines, which are all positively charged in the relevant pH range ($3.0 < \text{pH} < 6.0$) so that the total charge is +10 e (including the N-terminus).[27] These charges are distributed fairly homogeneously along the ARM.[27] The shells are about 3 nm thick on average, which is comparable to the average width of the ARM region. Because there are 26 amino acid residues in the ARM and 190 in the whole coat protein, the amino acid density in the ARM region is about 14% of that in the protein shell itself.

The concentration of protein used in the experiments is 0.1−0.5 mg/mL.[5] Since the molecular weight of the CCMV capsid protein is 19.8 kDa,[25] this corresponds to a concentration of $0.5-2.5 \times 10^{-5}$ M. We are interested in several boundaries in the "phase diagram". The first is the curve on which the concentration of double shells equals the concentration of single-shell capsids. If the spacing between the shells is equal to the equilibrium spacing D_0, the number of proteins in the outer shell is about 360, assuming that the density of proteins is the same as that in the capsid.[28] The total number of proteins in the double shell is then about 540. If we assume that a significant fraction of the proteins is in single or double shells, then a concentration of capsids and double shells of $10^{-9}-10^{-8}$ M seems reasonable (most of the other proteins are actually present as dimers instead of monomers[4] in the case of CCMV, but this does not change the results significantly; therefore, for simplicity, we ignore this fact). Using eq S.3 in the Supporting Information to calculate the free-energy contribution from mass action, that is, due to mixing, we then conclude that the free energy per protein in the double shell is about 0.08−0.10 k_BT lower than that in the single-shell capsid when their concentrations are equal. The third shell in a triple shell contains about 590 proteins,[29] assuming $D_2 = D_3 = D_0$, and the triple shell contains about 1130 proteins in total. At the line of equal concentrations of double shells and triple shells, the free-energy difference per protein in the two structures is about 0.020−0.023 k_BT. These numbers are indeed small, as remarked earlier, and we can therefore neglect the effects of mass action.

Figure 4 shows the calculated boundaries together with the experimental data points. The experiments were performed at 4 °C using several different buffers, and the ionic strength was adjusted using NaCl (see ref 5 for details). The boundaries calculated using the charge determined from the pK values agree reasonably well with the experimental results. The boundary between the double-shell and triple-shell regions is in agreement

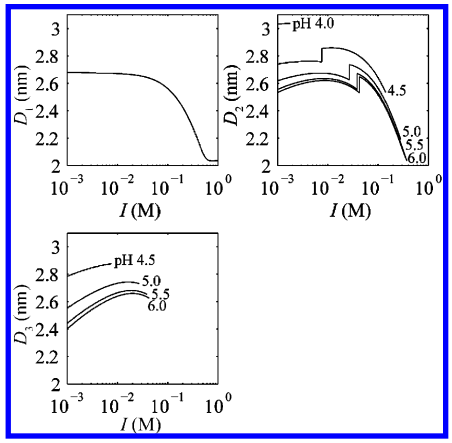

Figure 5. Shell spacings as a function of ionic strength I for several values of the pH. Only spacings for the dominant assemblies are depicted (see Figure 4 for the dominant assemblies). That is why there are jumps in the lines for D_2 at the boundary of the double-shell and triple-shell regions and why some lines end at a different pH than others. Since we assume that D_1 is not influenced by the charge on the outside of the shell, it is independent of pH.

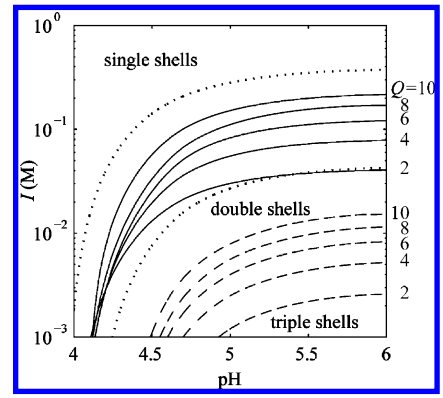

Figure 6. Same as Figure 4 but now for salt-stable mutants. For reference, the two solid lines in Figure 4 are reproduced here as dotted lines. The solid lines denote the boundary between the region with capsids and the region with double shells for $Q = 10, 8, 6, 4, 2$ (from top to bottom), and the dashed lines denote the boundary between the regions of double shells and triple shells, again for $Q = 10, 8, 6, 4, 2$ (from top to bottom). We used the values of $k_{elas} = 40$ and $D_0 = 2.0$ nm and the surface charge calculated from the pK values of the amino acids (see Figure 3).

with experiment, whereas the boundary between the capsid and double-shell region seems to lie at slightly too high of an ionic strength. We stress that there are no adjustable parameters in our model.

We have also calculated the boundaries at which the ratio between the two types of shells (capsids and double shells or double shells and triple shells) is 0.01 and 100. These lie very close to the boundaries at which the concentrations are equal; therefore, we do not show them here. This implies that the transitions between regions are fairly sharp.

At high ionic strengths, the equilibrium spacing between shells is close to the equilibrium end-to-end distance of the ARMs of 2.0 nm, while it increases to about 2.5−3 nm at lower ionic strength (see Figure 5). A value of about 3 nm at low ionic strengths is in agreement with the experimental observations.[5]

There are no multishell structures at pH > 5.0. This is probably caused by the pentamers being relatively less stable than hexamers at high pH, leading to structures with a relatively larger number of hexamers, like tubes.[5,25,30]

3.2. CCMV Mutants. Mutants of the CCMV capsid protein have been made in which some of the charged amino acids in the ARM have been replaced by uncharged amino acids, which means the ARM still has the same length but a lower charge.[31] In this section, we predict what effect this has on the stability of the multishells.

Besides the substitutions of the amino acids in the ARM, there is also a single substitution of lysine (residue 42) to arginine. Coat proteins with this substitution are called salt-stable mutants because virions formed from this mutant do not dissociate at pH 7 and high ionic strength ($I > 1$ M),[32] in contrast to the native virion. Because this substitution is not in the ARM but in the part of the coat protein that forms the shell, it leads to a change in the Young's modulus of the capsid. Unfortunately, nanoindentation experiments have only been performed on salt-stable mutant capsids containing ssRNA and not on empty capsids.[21] The Young's modulus of these filled capsids is indeed larger than that of the full wild-type capsids.[21] However, measurements are available for empty capsids of another mutant, SubE, which has the same substitution as the salt-stable mutant but whose arginines and lysines in the ARM are replaced by glutamic acid, which effectively gives the ARMs a charge of $Q = -8$ instead of $+10$.[33] We assume that this change in the charge of the ARMs has little influence on the Young's modulus, and we use the value $E = 360$ MPa[20] of the SubE mutant for the salt-stable mutant. Then, the elastic constant is $k_{elas} = 40$ (see eq 6), which is about 30% higher than the value for native CCMV.

Since the size of the CCMV mutant capsids is equal to the size of the native capsid, and since the ARMs are equally long, the other parameters are the same as for native CCMV. We assume the charge of the outside surface is also the same and we use the charge calculated using the Henderson−Hasselbalch equation and the pK values of the amino acids close to the outside surface (see Figure 3).

Figure 6 shows our predictions of the boundaries between the different regions for the salt-stable mutant for several values of the charge Q of the ARM. For comparison, the boundaries for native CCMV are also shown. If we compare the boundaries of native CCMV and the CCMV mutant with $Q = 10$, we see that increasing the stiffness of the shells decreases the ionic strength at which a boundary between two regions is found. The reason for this is that in the latter case the gain in electrostatic free energy due to the formation of a shell has to offset a larger elastic penalty than in the former, which can be achieved by lowering the salt concentration. Decreasing the total charge of the ARM has a similar effect. The figure also shows that one can still expect multishell structures for $I > 0.001$ M for most values of Q, but they will show up at lower ionic strengths than in the case of the native coat protein. For the larger values of Q, the spacing between shells is somewhat larger than the equilibrium spacing D_0, whereas for smaller values of Q ($Q \leq 6$), it is essentially equal to D_0.

3.3. Other Viruses. In this section, we predict where the boundaries between stability regions of different types of multishells are for capsids of several sizes and how these boundaries change when we change the stiffness of the shells or the size or total charge of the ARMs. We should point out that with a few notable exceptions, multishells have so far not been widely observed in viruses other than CCMV. There is some evidence of double shells for BMV,[34] but in these experiments, trypsin was added to the protein, and this removes 63 amino acid residues at the N-terminus. It implies that after this treatment, the coat proteins no longer have an ARM and

that, therefore, the shells must be stabilized through a mechanism not covered by our theory. Double shells were observed in alfalfa mosaic virus (AMV),[35] except that "empty" capsids and double shells only assemble in the presence of a small amount of RNA. As we discuss in the next section, AMV might, despite the presence of RNA, still be a good candidate to which our theory applies. Finally, bacteriophage fr, a $T = 3$ ssRNA phage,[36] does form multishells, despite it not having an ARM, albeit only at high concentrations. The same reservation holds here as that for the quoted experiments on BMV.[34] Although we know of no unambiguous other examples than CCMV, this does not mean that they do not exist, and we include this section for completeness and as a guide for possible experiments. We note that ssRNA viruses almost always have an ARM region and that, therefore, we expect that our theory must quite generally be valid.

For simplicity, we assume that the shells have a fixed width of $h = 3$ nm. Then, there are five physical quantities that need to be specified, the preferred size of the capsid R_0, the number of proteins in a capsid N_1, the elastic constant k_{elas}, the number of amino acids in an ARM n, and the total charge of an ARM Q (although we could eliminate N_1 because mass action is not very important here; see section 2.4). We will only consider $T = 3$ and 4 capsids, and since most $T = 3$ capsids have an outer diameter of about 28 nm and most $T = 4$ capsids have an outer diameter of about 40 nm, we consider two cases, namely, $N_1 = 180$ and $R_0 = 12.5$ nm and $N_1 = 240$ and $R_0 = 18.5$ nm. We make no assumptions about the surface charge density but plot the ionic strength at a boundary as a function of the surface charge per protein q_0 instead of pH.

We start with a few qualitative remarks about how boundaries shift when parameters are changed. Increasing the elastic constant k_{elas} shifts boundaries to lower ionic strengths at fixed surface charge density because it leads to an increase in the elastic deformation energy. This has to be compensated by a larger difference in electrostatic free energy between the two types of structures at the boundary, which is achieved by lowering the ionic strength. Increasing the number of amino acids in an ARM has the same effect because it increases the equilibrium extension D_0, which effectively increases the radius of curvature of shells (except the capsid), leading to a larger deformation energy of the shell and also to a smaller charge density between shells. Increasing the charge Q per ARM has the opposite effect. It increases the charge density between shells and thus the electrostatic free-energy difference between two structures at a boundary. An increase in ionic strength compensates for this; therefore, boundaries are shifted to higher ionic strengths.

For the $T = 3$ ssRNA viruses listed in Table 1 of ref 37, the charge Q per ARM is typically between 10 and 20, and the number of amino acids n in the ARM is between 25 and 75. (See the Protein Data Bank[26] (www.pdb.org) for information on what part of the protein is disordered in the crystal structure. We interpret this part as the ARM. Reference 38 also lists values of Q and n for a number of $T = 3$ ssRNA viruses.) We use the average values $Q = 15$ and $n = 50$ ($D_0 = 2.8$ nm), and we assume that the elastic constant is about equal to the one for CCMV and set $k_{elas} = 30$. In Figure 7, we plot the regions of stability. From the discussion in the previous paragraph, one can make a qualitative estimate of what happens to the boundaries when the values of the physical properties are changed. Notice that we plot the results only for $-q_0 \geq 1$ because for $-q_0 < 1$, there is less than one charge per protein. This means that the outsides of the proteins are uncharged for part of the time and charge fluctuations might then play a role and the stretching of the tails might become inhomogeneous. Because we do not take this into account in our model and are thus not sure how accurate our predictions are for very low surface charge densities, we do not consider this specific regime.

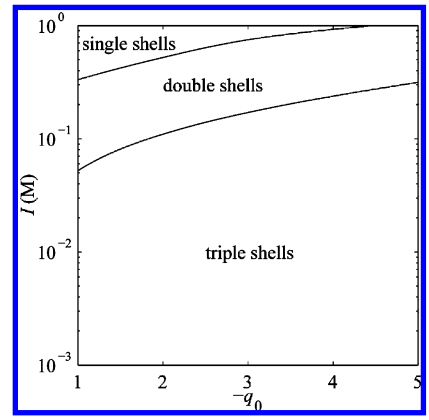

Figure 7. Prediction of the majority type of structure formed by a typical $T = 3$ capsid protein as a function of surface charge per protein q_0 and ionic strength I. The top line denotes the boundary at which the concentrations of single shells and double shells are equal. The lower line denotes the boundary at which the concentrations of double shells and triple shells are equal. We use the values $R_0 = 12.5$ nm, $N_1 = 180$, $k_{elas} = 30$, $Q = 15$, and $n = 50$ ($D_0 = 2.8$ nm).

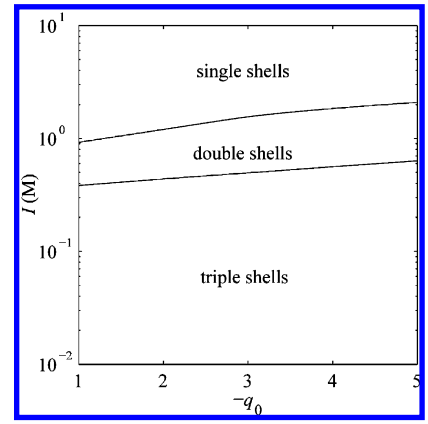

Figure 8. Same as Figure 7 but now for a typical $T = 4$ capsid protein. The solid lines are the boundaries for $R_0 = 18.5$ nm, $N_1 = 240$, $k_{elas} = 20$, $Q = 30$, and $n = 100$ ($D_0 = 4.0$ nm). Note the change in range of the vertical axis compared to the previous plots.

Table 1 of ref 37 only lists details of a number of $T = 4$ ssRNA viruses. For those viruses, we find average values of about $Q = 30$ and $n = 100$ ($D_0 = 4.0$ nm). We assume that the Young's modulus is similar to the one for CCMV. Because $N_1 = 240$ for $T = 4$, the elastic constant is about 25% smaller; therefore, we choose $k_{elas} = 20$. (Our particular choice of k_{elas} makes it proportional to N_1^{-1}; see eqs 5 and 6.) Figure 8 shows the predictions for these values. Notice the high ionic strength at the boundary between single and double shells for large values of $|q_0|$. It is doubtful whether our theory is still valid at these high ionic strengths because the Debye length then drops below the size of an ion.

4. Discussion

In this paper, we have shown that multishell structures of CCMV coat protein are stable at sufficiently high pH, when the outside surface of a shell has a sufficiently high negative charge, and also at sufficiently low ionic strength. Competing contributions stemming from elasto- and electrostatics determine

the stability of these structures. Shells larger than the native capsid have an elastic free-energy penalty due to a radius of curvature that is larger than the preferred one. We have related this elastic free-energy penalty to the elastic properties of the native capsid. At the same time, there is a decrease in electrostatic free energy due to two separate effects. The first is that, due to the larger radius of curvature, every ARM has a slightly larger volume at its disposal, which decreases the charge density and thus the electrostatic free energy. The second is the interaction between the positively charged ARMs and the negatively charged outside surface of a shell, which leads to counterion release and a decrease in the electrostatic free energy.

The increase in elastic free energy is independent of the ionic strength, whereas the decrease in electrostatic free energy increases in magnitude with decreasing ionic strength, resulting in multishells becoming stable at sufficiently low ionic strengths. As the ionic strength decreases, the number of shells in the stable multishell structure increases. A third contribution to the free energy, the entropic free energy associated with stretching or compressing the layer of ARMs, determines the spacing between shells. Our model predicts what type of multishell structure is stable as a function of ionic strength and surface charge for a typical $T = 3$ and 4 capsid. These results can be used to determine the stable structures if one keeps in mind that increasing the bending stiffness of the shell decreases the ionic strength at which boundaries between regions of different stable multishells are found. Increasing the length of the ARMs has the same effect, whereas increasing the charge per ARM has the opposite effect.

We remark that our predictions are consistent with results for brome mosaic virus (BMV), a virus from the same family as CCMV. The BMV coat protein has only 70% homology with that of CCMV, but the ARM is very similar[39] and has about the same charge. No multishells are observed in a pH range of about 1−8,[40] which is what we would expect because in that range, the charge of the outside surface of a capsid has the same sign as the charge on the ARMs.[41] There are a few other cases of double-shell formation in viral protein self-assembly. Lebeurier et al.[35] studied the self-assembly of the capsid protein from alfalfa mosaic virus (AMV) and observed double-shelled structures in a buffer at pH 7.4 and 8.0 containing 0.1 M NaCl. The presence of a small amount of RNA seems necessary; attempts to form capsids using protein alone were unsuccessful.[42] The AMV capsid protein can assemble both into ($T = 1$) icosahedra and bacilliform particles, and double shells were observed for both polymorphs.

AMV is insoluble below pH 6; therefore, it is not possible to determine its isoelectric point directly, but extrapolation from measurements at several pH's suggests that it is around 4.6.[43] Thus, it is negatively charged under all conditions where it is water-soluble. The virus protein has a basic N-terminus consisting of 26 amino acids, of which 6 are lysines and 2 are arginines;[44] therefore, $Q = 9$ (including the N-terminus). Nothing is known about the Young's modulus of the AMV capsid, but if we assume that it is the same as that for CCMV, then our model predicts that double shells of which the inner shell is the $T = 1$ icosahedron are stable for $I \lesssim 0.01$ M if the outer capsid surface has a net charge of a few negative charges per protein. If the Young's modulus is half of that of CCMV, this is the case for $I \lesssim 0.1$ M, in agreement with observations.

If we consider all of the charged amino acids close to the surface, then AMV should be almost neutral at a pH between 7.4 and 8.0, in contradiction to electrophoretic mobility measurements. However, AMV is one of the few plant viruses that is sensitive to pancreatic RNase,[45] indicating that the surface contains relatively large holes. This is confirmed by the crystal structure,[46] which shows holes of radius larger than 1.7 nm in the pentamers. It is likely that some of the RNA, which is needed for the assembly, sticks out through these holes, giving the capsid a negative charge and allowing the formation of double shells.

As we have noted, bacteriophage fr forms multishells. They are not very stable, however, appearing only at high protein concentrations >0.5 mg/mL, and there is no long, highly charged, ARM region sticking out into the interior of the shell.[47] This implies that our theory does not apply to this case and that the multishells might be stabilized by some other mechanism.

We end this discussion with a simple calculation that shows the effect of counterion release and that explains the qualitative features of the phase diagram. We assume that the entropy of the ARMs determines to first order the width of the ARM region and that the balance between free-energy gain due to counterion release and the free-energy penalty of a larger radius of curvature determines the position of the boundaries in the phase diagram.

If we assume that each released ion gains about 1 k_BT in free energy, then the free energy associated with the counterion release is $F_{elec} \approx -2N_1|q_{eff}|$ (in units of k_BT), where N_1 is the number of proteins in the inner shell. The factor 2 comes from the fact that counterions are released in pairs, one positive counterion from the negatively charged outer surface of the shell and one negatively charged counterion from the positively charged ARMs. Finally, $|q_{eff}|$ is the number of counterions that is released per protein, which depends on the overlap of the Debye layer of the outside surface of the shell with the ARM layer. Provided the excess number of counterions in the Debye layer $N_1|q_0|$ is lower than the number of counterions of the ARMs in the overlapping region, $N_2Q\kappa^{-1}/D_2$, then all of the excess counterions can be released, and $q_{eff} = q_0$. Here, N_2 is the number of proteins in the outer shell. If the excess number of counterions is larger than the number of counterions of the ARMs in the overlapping region, then the number of released counterions is equal to the latter number, $N_1|q_{eff}| \approx N_2Q\kappa^{-1}/D_2$. For simplicity, we use the interpolation $q_{eff}/q_0 \approx (1 + \kappa D_2N_1|q_0|/N_2Q)^{-1}$.

The elastic penalty associated with the deformation of the shell is $F_{elas} \approx (1/2)k_{elas}N_2[1 - (N_1/N_2)^{1/2}]^2$ (see eq 5). At the boundary between the single-shell region and the double-shell region, F_{elec} and F_{elas} balance each other. In other words $|q_{eff}| \approx (1/4)k_{elas}N_2/N_1[1 - (N_1/N_2)^{1/2}]^2 \equiv \gamma$. We now insert the expression for q_{eff} and for the inverse Debye length $\kappa \approx 3.3I^{1/2}$ (with I in M and κ in nm^{-1}; see also section S2 of the Supporting Information), and we find that at the boundary between the regions of single and double shells

$$I \approx 0.09\left[\left(\frac{1}{\gamma} - \frac{1}{|q_0|}\right)\frac{N_2}{N_1}\frac{Q}{D_2}\right]^2 \tag{9}$$

From this we conclude that $I \downarrow 0$ and $\mathrm{d}\log I/\mathrm{d}|q_0| \rightarrow \infty$ as $|q_0| \downarrow \gamma$, which is what we see in our phase diagrams; see, for example, Figure 4. This happens upon approach of the isoelectric point of the outer surface of the shell. We also see that I is a monotonically increasing function of $|q_0|$ (and thus of pH in our case since $|q_0|$ increases with pH beyond the isoelectric point) and that $\mathrm{d}\log I/\mathrm{d}|q_0|$ is a monotonically decreasing function of $|q_0|$ (and pH) for $|q_0| \geq \gamma$. For $|q_0| \gg \gamma$, at high pH, I levels off, also in agreement with our phase diagram.

In the case of native CCMV capsid proteins, we have $N_2/N_1 \approx 2.0$ and $k_{elas} = 30$; therefore, $\gamma \approx 1.3$. This overestimates the value of $|q_0|$ at the pH at which $I \downarrow 0$ by about a factor of 2 (see Figures 3 and 4), which is not surprising considering the simplicity of our scaling estimate. The value of I at the boundary at higher pH is overestimated by a little less than an order of magnitude by eq 9. The equation also shows that decreasing Q does not really influence the phase diagram close to $|q_0| = \gamma$, at least not if we use a logarithmic scale for the ionic strength axis, whereas it shifts down the boundaries at higher values of $|q_0|$. This is exactly what we see in Figure 6. Hence, we conclude that our interpretation of counterion release as the stabilizing force for multishell formation is correct.

The existence of multishell structures and of other empty capsid polymorphs in several viruses makes evident that the concept of spontaneous curvature and the balance between it and electrostatic interactions has a biological relevance in that they must play a role in the selection of viral capsid size.

Acknowledgment. We thank Drs. Jaime Ruiz-Garcia, William Klug, Alex Levine, Laurence Lavelle, and Robijn Bruinsma for insightful discussions. The work described in this paper was partially supported by the U.S. National Science Foundation (Grant CHE 0714411 to W.M.G. and C.M.K.). P.P. acknowledges support form a Rubicon grant from The Netherlands Organisation for Scientific Research.

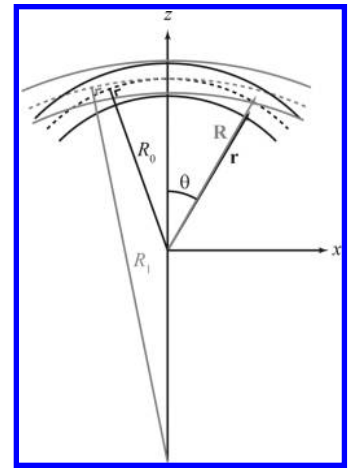

Figure 9. Cross section (at $\phi = 0$) of a small part (in principle, infinitesimally small but exaggerated here for clarity) of a capsid before (black lines) and after (gray lines) deformation. The dotted lines denote the neutral surfaces, and the two parts are positioned such that their neutral surfaces coincide at $\theta = 0$. The radius of curvature of the neutral surface of the undeformed section is R_0, while the radius of curvature of the neutral surface of the deformed one close to $\theta = 0$ is R_1 in this cross section and R_2 in a perpendicular cross section through the z axis; **r** is a point before deformation, and **R** is the same point after deformation.

Appendix

Elastostatic Free Energies. We assume that we can model the protein shell as an elastic shell of thickness h, Young's modulus E, and Poisson ratio ν. We define the radius of curvature of a shell as the radius of curvature of the neutral surface of the shell, which is the spherical surface that dissects the shell into two shells of equal thickness. We also assume that the capsid radius R_0 is the preferred radius of curvature. In other words, we assume that in a native capsid, there is no stretching or compression anywhere in the shell.

Shells are only one protein layer thick; therefore, even though larger shells contain more elastic material (proteins) than smaller ones, there will be some stretching and compression because the ratio between the outer radius and the inner radius decreases as the radius of curvature of the shell increases, whereas the ratio of material at the outer surface to material at the inner surface is constant (the thickness of the shell stays more or less constant, as we shall see further on). Since the shells are spherically symmetric, the amount of stretching or compression will depend only on the radial coordinate; therefore, we only need to study an infinitesimal section of the shell to calculate the deformations and the corresponding elastic free energy. If we increase the radius of curvature of an infinitesimal section, we expect compression parallel to the surface on the outside, stretching parallel to the surface on the inside, and no stretching or compression in the neutral surface; this is the reason why it has this name. (In fact, due to the asymmetry between the outside of a shell and the inside, we do not expect the neutral surface to be exactly in the middle, but corrections to its position will be small, on the order of h^2/R_0.) This may be counterintuitive because one might expect that there is always stretching or compression if one changes the radius of curvature of a piece of an infinitely thin shell (i.e., the neutral surface). However, as we will show soon, this is not the case if the piece is infinitesimally small. This leads to the important conclusion that even though we have to add material to a small shell to make a bigger one, we do not have to add material when we deform an infinitesimally small section of a shell to make a corresponding infinitesimally small section of a bigger shell.

In order to calculate the free energy, we need to calculate the strains and stresses accompanying the deformation.[19] We start with a Cartesian coordinate system with its origin coinciding with the center of the capsid. A point **r** inside, the capsid shell is given by

$$\mathbf{r} = r\begin{pmatrix}\sin\theta\cos\phi\\ \sin\theta\sin\phi\\ \cos\theta\end{pmatrix} = r\begin{pmatrix}\theta\cos\phi\\ \theta\sin\phi\\ 1\end{pmatrix} + O(\theta^2) \quad \text{(A.1)}$$

where $R_0 - h/2 \le r \le R_0 + h/2$, $0 \le \phi < 2\pi$, and where we are only interested in a small section $0 \le \theta \ll 1$.

It will be insightful to study first the more general case where the two principal radii of curvature R_1 and R_2 of the deformed neutral surface are not necessarily equal. We assume that the points on a normal to the neutral surface of the undeformed section are all on the same normal after deformation. It is clear that this has to be the case for $R_1 = R_2$ because of the spherical symmetry of the shell of which the small section is a part. We expect some stretching in the radial direction, which is accounted for by a factor of $(1 + \varepsilon)$, where $\varepsilon = O(h/R_0)$ and where ε could, in principle, depend on r, θ, and ϕ. We will determine ε later by demanding that the stresses in the radial direction be small compared to the other stresses in the shell (consistent with the thinness of the shell, i.e., $h/R_0 \ll 1$). If we demand that the point $(r, \theta) = (R_0, 0)$ be in the same place after the deformation, that the normal to the neutral surface at this point still point in the same direction, and that the principal direction of R_1 be along the x-axis, then the new position **R** of point **r** after deformation is (see also Figure 9)

$$\mathbf{R} = R_0\begin{pmatrix}\theta\cos\phi\\ \theta\sin\phi\\ 1\end{pmatrix} + (r - R_0)(1+\varepsilon)\begin{pmatrix}\frac{R_0}{R_1}\theta\cos\phi\\ \frac{R_0}{R_2}\theta\sin\phi\\ 1\end{pmatrix} + O(\theta^2) \tag{A.2}$$

Note that even though the neutral surface of the deformed section has different principle radii of curvature from the undeformed section, it is undeformed to first order in θ. The direction of the surface normal does change slightly however.

The displacement vector is defined as

$$\mathbf{u} \equiv \mathbf{R} - \mathbf{r} = -(r - R_0)\begin{pmatrix}\left(1 - \frac{R_0}{R_1}\right)\theta\cos\phi\\ \left(1 - \frac{R_0}{R_2}\right)\theta\sin\phi\\ -\varepsilon\end{pmatrix} \times \left(1 + O\left(\frac{h}{R_0}\right)\right) + O(\theta^2) \tag{A.3}$$

and its components in a spherical coordinate system are

$$u_r \equiv \mathbf{u}\cdot\mathbf{e}_r = (r - R_0)\varepsilon + O(\theta^2) \tag{A.4}$$

$$u_\theta \equiv \mathbf{u}\cdot\mathbf{e}_\theta = -(r - R_0)\theta\left(1 - \frac{R_0}{R_1}\cos^2\phi - \frac{R_0}{R_2}\sin^2\phi\right) \times \left(1 + O\left(\frac{h}{R_0}\right)\right) + O(\theta^3) \tag{A.5}$$

and

$$u_\phi \equiv \mathbf{u}\cdot\mathbf{e}_\phi = -(r - R_0)\theta\left(\frac{R_0}{R_1} - \frac{R_0}{R_2}\right) \times \sin\phi\cos\phi\left(1 + O\left(\frac{h}{R_0}\right)\right) + O(\theta^3) \tag{A.6}$$

where the Cartesian components of the unit vectors are given by

$$\mathbf{e}_r = \begin{pmatrix}\theta\cos\phi\\ \theta\sin\phi\\ 1\end{pmatrix}(1 + O(\theta^2)) \qquad \mathbf{e}_\theta = \begin{pmatrix}\cos\phi\\ \sin\phi\\ -\theta\end{pmatrix}(1 + O(\theta^2)) \qquad \mathbf{e}_\phi = \begin{pmatrix}-\sin\phi\\ \cos\phi\\ 0\end{pmatrix} \tag{A.7}$$

The surface of the deformed shell must be stress-free. In other words, we should have $\sigma_{ik}n_k = 0$ at the surface, where σ_{ik} is the stress tensor and $\mathbf{n} = \mathbf{e}_r(1 + O(\theta))$ is the surface normal. This implies that to leading order in θ, the stresses σ_{rr}, $\sigma_{\theta r}$, and $\sigma_{\phi r}$ vanish at the surface. Since the shell is thin ($h \ll R_0$) we expect these stresses to be small compared to the other stresses throughout the shell. Equating σ_{rr} to zero leads to the following equation

$$(1 - \nu)u_{rr} = -\nu(u_{\theta\theta} + u_{\phi\phi}) \tag{A.8}$$

ε can be determined from this equation by demanding that it is satisfied to leading order in θ. We first compute the relevant components of the strain tensor[19]

$$u_{rr} = \varepsilon + (r - R_0)\varepsilon' + O(\theta^2) \tag{A.9}$$

$$u_{\theta\theta} = -\frac{r - R_0}{R_0}\left(1 - \frac{R_0}{R_1}\cos^2\phi - \frac{R_0}{R_2}\sin^2\phi\right) \times \left(1 + O\left(\frac{h}{R_0}\right)\right) + O(\theta^2) \tag{A.10}$$

$$u_{\phi\phi} = -\frac{r - R_0}{R_0}\left(1 - \frac{R_0}{R_1}\sin^2\phi - \frac{R_0}{R_2}\cos^2\phi\right) \times \left(1 + O\left(\frac{h}{R_0}\right)\right) + O(\theta^2) \tag{A.11}$$

and we then find

$$\varepsilon = \frac{\nu}{2(1 - \nu)}\frac{r - R_0}{R_0}\left(2 - \frac{R_0}{R_1} - \frac{R_0}{R_2}\right) \times \left(1 + O\left(\frac{h}{R_0}\right)\right) + O(\theta^2) \tag{A.12}$$

and

$$u_{rr} = \frac{\nu}{(1 - \nu)}\frac{r - R_0}{R_0}\left(2 - \frac{R_0}{R_1} - \frac{R_0}{R_2}\right) \times \left(1 + O\left(\frac{h}{R_0}\right)\right) + O(\theta^2) \tag{A.13}$$

The remaining components of the strain tensor are

$$u_{\theta\phi} = u_{\phi\theta} = -\frac{r - R_0}{R_0}\left(\frac{R_0}{R_1} - \frac{R_0}{R_2}\right) \times \sin\phi\cos\phi\left(1 + O\left(\frac{h}{R_0}\right)\right) + O(\theta^2) \tag{A.14}$$

$u_{r\theta} = u_{\theta r} = O(\theta)$ and $u_{r\phi} = u_{\phi r} = O(\theta)$. These last two equations are consistent with the fact that $\sigma_{\theta r} \approx u_{\theta r}$ and $\sigma_{\phi r} \approx u_{\phi r}$ should be small.

The free-energy density in the small section is equal to[19]

$$f = \frac{E}{2(1 + \nu)}\left(\frac{r - R_0}{R_0}\right)^2\left[u_{rr}^2 + u_{\theta\theta}^2 + u_{\phi\phi}^2 + 2u_{\theta\phi}^2 + \frac{\nu}{1 - 2\nu}(u_{rr} + u_{\theta\theta} + u_{\phi\phi})^2\right] \tag{A.15}$$

from which the free energy per unit area of neutral surface can be calculated by integration

$$F = \frac{1}{2\pi R_0^2(1-\cos\theta)}\int_0^{2\pi} d\phi \int_0^{\theta} \sin\tilde{\theta}\, d\tilde{\theta} \int_{R_0-h/2}^{R_0+h/2} r^2\, dr\, f$$
$$= \frac{1}{2}\frac{Eh^3}{12(1-\nu^2)}\left(\frac{2}{R_0}-\frac{1}{R_1}-\frac{1}{R_2}\right)^2 - \frac{Eh^3}{12(1+\nu)}\left(\frac{1}{R_0}-\frac{1}{R_1}\right)\left(\frac{1}{R_0}-\frac{1}{R_2}\right) \quad \text{(A.16)}$$

The elastic constant in the first term, $Eh^3/12(1-\nu^2)$, is called the flexural rigidity or cylindrical rigidity;[19] if one of the radii of curvature of the neutral plane of the deformed section is equal to the radius of curvature in the undeformed section, the second term in the free energy vanishes. If $R_1 = R_2 = R$, we find

$$F = \frac{1}{2}\frac{Eh^3}{24(1-\nu)}\left(\frac{2}{R_0}-\frac{2}{R}\right)^2 \quad \text{(A.17)}$$

Note that even though we integrated over a circular section of the neutral surface in eq A.16, the result is valid for a section of any shape as long as it is small.

Supporting Information Available: Section S1 discusses the influence of the law of mass action on the relative abundances of the different multishells. Section S2 describes the calculation of the electrostatic free energies of multishells in detail. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Fraenkel-Conrat, H.; Williams, R. C. *Proc. Natl. Acad. Sci. U.S.A.* **1955**, *41*, 690.
(2) Bancroft, J. B.; Hiebert, E. *Virology* **1967**, *32*, 354.
(3) Bancroft, J. B.; Wagner, G. W.; Bracker, C. E. *Virology* **1968**, *36*, 146.
(4) Adolph, K. W.; Butler, P. J. G. *J. Mol. Biol.* **1974**, *88*, 327.
(5) Lavelle, L.; Gingery, M.; Phillips, M.; Gelbart, W. M.; Knobler, C. M.; Cadena-Nava, R. D.; Vega-Acosta, J. R.; Pinedo-Torres, L. A.; Ruiz-Garcia, J. *J. Phys. Chem. B* **2009**, *113*, 3813.
(6) Johnson, M. W.; Wagner, G. W.; Bancroft, J. B. *J. Gen. Virol.* **1973**, *19*, 263.
(7) Hagan, M. F. *J. Chem. Phys.* **2009**, *130*, 114902.
(8) Engh, R. A.; Huber, R. *Acta Crystallogr., Sect. A* **1991**, *47*, 392.
(9) Rühe, J.; Ballauff, M.; Biesalski, M.; Dziezok, P.; Gröhn, F.; Johannsmann, D.; Houbenov, N.; Hugenberg, N.; Konradi, R.; Minko, S.; Motornov, M.; Netz, R. R.; Schmidt, M.; Seidel, C.; Stamm, M.; Stephan, T.; Usov, D.; Zhang, H. *Adv. Polym. Sci.* **2004**, *165*, 79.
(10) Pincus, P. *Macromolecules* **1991**, *24*, 2912.
(11) de Gennes, P.-G. *Scaling Concepts in Polymer Physics*; Cornell University Press: Ithaca, NY, 1979.
(12) Flory, P. J. *Statistical Mechanics of Chain Molecules*; Hanser: New York, 1989.
(13) Overbeek, J. Th. G. *Colloids Surf.* **1990**, *51*, 61.
(14) Šiber, A.; Podgornik, R. *Phys. Rev. E* **2008**, *78*, 051915.
(15) Sheinerman, F. B.; Norel, R.; Honig, B. *Curr. Opin. Struct. Biol.* **2000**, *10*, 153.
(16) Schutz, C. N.; Warshel, A. *Proteins* **2001**, *44*, 400.
(17) Šiber, A.; Podgornik, R. *Phys. Rev. E* **2007**, *76*, 061906.
(18) Caspar, D. L. D.; Klug, A. *Cold Spring Harbor Symp. Quant. Biol.* **1962**, *27*, 1.
(19) Landau, L. D.; Lifshitz, E. M. *Theory of Elasticity*, 3rd ed.; Butterworth-Heinemann: Oxford, U.K., 1986.
(20) Gibbons, M. M.; Klug, W. S. *Phys. Rev. E* **2007**, *75*, 031901.
(21) Michel, J. P.; Ivanovska, I. L.; Gibbons, M. M.; Klug, W. S.; Knobler, C. M.; Wuite, G. J. L.; Schmidt, C. F. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 6184.
(22) Henry, D. C. *Proc. R. Soc. London, Ser. A* **1931**, *133*, 106.
(23) Dahlgleish, D. G.; Dickinson, E.; Whyman, R. H. *J. Colloid Interface Sci.* **1985**, *108*, 174.
(24) Properties of Amino Acids. In *CRC Handbook of Chemistry and Physics*, 89th ed. (internet version 2009); Lide, D. R., Ed.; CRC Press/Taylor and Francis: Boca Raton, Florida, 2009.
(25) Speir, J. A.; Munshi, S.; Wang, G.; Baker, T. S.; Johnson, J. E. *Structure* **1995**, *3*, 63.
(26) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000**, *28*, 235.
(27) Lavelle, L.; Michel, J.-P.; Gingery, M. *J. Virol. Methods* **2007**, *146*, 311.
(28) This corresponds to a T number of 6. This is not an allowed T number, but we quote it here for comparison. As mentioned before, it seems unlikely that the larger shells correspond to T number structures.[5]
(29) This corresponds to a T number of about 10.
(30) Bruinsma, R. F.; Gelbart, W. M.; Reguera, D.; Rudnick, J.; Zandi, R. *Phys. Rev. Lett.* **2003**, *90*, 248101.
(31) Brumfield, S.; Willits, D.; Tang, L.; Johnson, J. E.; Douglas, T.; Young, M. *J. Gen. Virol.* **2004**, *85*, 1049.
(32) Fox, J. M.; Zhao, X.; Speir, J. A.; Young, M. J. *Virology* **1996**, *222*, 115.
(33) Douglas, T.; Strable, E.; Willits, D.; Aitouchen, A.; Libera, M.; Young, M. *Adv. Mater.* **2002**, *14*, 415.
(34) Cuillel, M.; Jacrot, B.; Zulauf, M. *Virology* **1981**, *110*, 63.
(35) Lebeurier, G.; Fraenkel-Conrat, H.; Wurtz, M.; Hirth, L. *Virology* **1971**, *43*, 51.
(36) Schubert, D.; Frank, H. *Virology* **1971**, *43*, 41.
(37) Belyi, V. A.; Muthukumar, M. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 17174.
(38) Hu, T.; Zhang, R.; Shklovskii, B. I. *Physica A* **2008**, *387*, 3059.
(39) Dasgupta, R.; Kaesberg, P. *Nucleic Acids Res.* **1982**, *10*, 703.
(40) Pfeiffer, P.; Hirth, L. *Virology* **1974**, *61*, 160.
(41) Bockstahler, L. E.; Kaesberg, P. *Biophys. J.* **1962**, *2*, 1.
(42) Hull, R. *Virology* **1970**, *40*, 34.
(43) Lauffer, M. A.; Ross, A. F. *J. Am. Chem. Soc.* **1940**, *62*, 3296.
(44) Guogas, L. M.; Filman, D. J.; Hogle, J. M.; Gehrke, L. *Science* **2004**, *306*, 2108.
(45) Bol, J. F.; Veldstra, H. *Virology* **1969**, *37*, 74.
(46) Kumar, A.; Reddy, V. S.; Yusibov, V.; Chipman, P. R.; Hata, Y.; Fita, I.; Fukuyama, K.; Rossmann, M. G.; Loesch-Fries, L. S.; Baker, T. S.; Johnson, J. E. *J. Virol.* **1997**, *71*, 7911.
(47) Liljas, L.; Fridborg, K.; Valegård, K.; Bundule, M.; Pumpens, P. *J. Mol. Biol.* **1994**, *244*, 279.

JP911040Z