Physical Regulation of the Self-Assembly of Tobacco Mosaic Virus Coat Protein

Willem K. Kegel* and Paul van der Schoot†
*Van't Hoff Laboratory for Physical and Colloid Chemistry, Debye Research Institute, Utrecht University, Utrecht, The Netherlands; and †Eindhoven Polymer Laboratories, Technische Universiteit Eindhoven, Eindhoven, The Netherlands

ABSTRACT We present a statistical mechanical model based on the principle of mass action that explains the main features of the in vitro aggregation behavior of the coat protein of tobacco mosaic virus (TMV). By comparing our model to experimentally obtained stability diagrams, titration experiments, and calorimetric data, we pin down three competing factors that regulate the transitions between the different kinds of aggregated state of the coat protein. These are hydrophobic interactions, electrostatic interactions, and the formation of so-called "Caspar" carboxylate pairs. We suggest that these factors could be universal and relevant to a large class of virus coat proteins.

INTRODUCTION

The spontaneous formation of virus-like particles in aqueous solutions of the coat protein (CP) of tobacco mosaic virus (TMV) is generally seen as the paradigm for self-assembly in biology (1–3). Indeed, as was shown half a century ago, infective virus particles of helical symmetry spontaneously form upon mixing aqueous solutions of the coat protein and the RNA of the virus (4). The coat protein alone in fact exhibits, in solution, various aggregated states: mono- and oligomers, disk-like assemblies, and extended helices (5,6). The various aggregated states interconvert reversibly upon variation of the temperature, pH, and ionic strength. It appears that the propensity to form virus-like particles is an intrinsic property of the CP.

Simple mass-action models have proven quite successful in describing isolated experiments (7–11), but a theory that predicts the transitions between the various equilibrium aggregation states of TMV coat protein as a function of the external conditions is still lacking. In this work, we identify three factors of physical origin involved in the stability of the virus-like particles. These are 1), hydrophobic interactions; 2), electrostatic interactions; and 3), intersubunit carboxylate or Caspar pair interactions (7). Incorporated into a minimal statistical mechanical (mass-action) model, they explain the main features of the in vitro self-assembly behavior of the tobacco mosaic virus coat protein.

Our conclusions are based on a comparison with experimental findings, summarized in Fig. 1. Fig. 1 *A* gives the ranges over which the various aggregation states of the CP subunits are thought to be stable as a function of the ionic strength and pH, but at fixed temperature and concentration. It indicates that electrostatic interactions must play a role in the stability of the assemblies. The indicated stability boundaries do not demarcate true phase boundaries; they show where larger self-assembled species become detectable, yet do not imply that the smaller species actually disappear. The diagram includes both stable (reversibly formed) species and what presumably are metastable species (2,3).

The two-layered disks and the single helices form reversibly: they appear/disappear upon increasing/decreasing the proton concentration. The ''lock-washer'' species, on the other hand, slowly grows into larger helices, whereas the ''stacked disk'' structure is thought to represent an irreversible, partly proteolyzed aggregated state (see Klug (2) and Butler (3), and references cited therein). For simplicity, we shall ignore, in our model description, the appearance of cylindrical species consisting of more than two layers but discuss the implications of this idealization. Thus, in the following, only equilibria between ''monomers'', disks, and helices will be considered, where the ''monomers'' are also thought to include oligomeric species that we do not need to explicitly include in the model.

Fig. 1 *B*, taken from Sturtevant et al. (11), shows the excess heat capacity associated with the reversible assembly of TMV CP as a function of the temperature, measured at three different pH values. At low temperatures, the free ''monomers'' are the preferred species, whereas disk and helix aggregates form upon increasing the temperature (11). A more detailed discussion of this process is presented in the Results and Discussion section. It is important to note that the excess heat capacities are larger than zero, implying that the aggregation must be endothermic and is in all likelihood driven by hydrophobic interactions. What is not shown in the figure is that at temperatures $>\sim 35°C$, irreversible denaturation of the CPs takes place, a process accompanied by a large excess heat capacity.

Finally, in Fig. 1 *C*, typical acid-base titration curves of the TMV CP for a number of different temperatures are shown (reprinted from Butler et al. (10)). Helices form upon lowering the pH, whereby a total of approximately two protons per protein subunit are absorbed. Proton absorption seems to

Submitted August 12, 2005, and accepted for publication May 9, 2006.

Address reprint requests to Willem K. Kegel, E-mail: w.k.kegel@chem.uu.nl.

0006-3495/06/08/1501/12 $2.00

doi: 10.1529/biophysj.105.072603

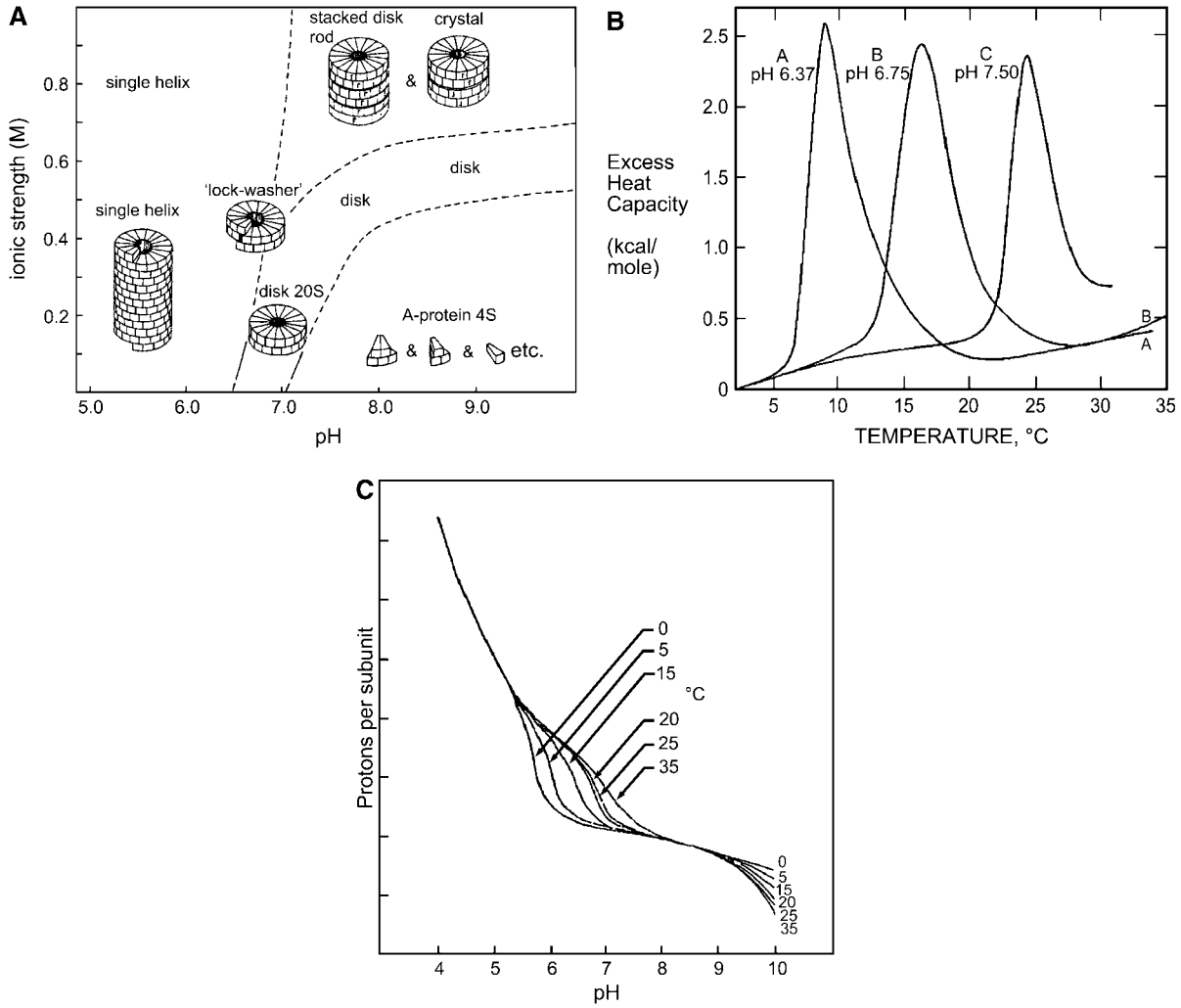

FIGURE 1 (*A*) Diagram of states of TMV protein subunit aggregates in aqueous solution as a function of the acidity and ionic strength (in the absence of RNA). The terms 4S and 20S refer to the sedimentation rate of the clusters. (Data taken from Klug (2).) The 20S disk is identified as a two-layer disk consisting of 34 subunits. It has a central hole with a diameter of ~2 nm and its outer diameter is ~18 nm. The molecular weight of a coat-protein subunit is 17,500 g/mol. Indicated are observations for a total subunit concentration of 0.3 mM. As explained in the main text, the boundaries are not true phase boundaries but rather crossovers, and not all species of aggregate form reversibly. (*B*) Results of excess heat capacity measurements on solutions of TMV coat protein at three different pH values (reprinted from Sturtevant et al. (11)). The ionic strength in the experiments was 0.1 M, and the total protein concentrations are 4.48 g/l at pH 6.37, 9.41 g/l at pH 6.75, and 5.45 g/l at pH 7.50. These concentrations are comparable to that of *A*. (*C*) Experimental acid-base titration curves of aqueous solutions of TMV coat protein at the temperatures indicated (reprinted from Butler et al. (10)). The measurements were made at a protein concentration of 1.7 g/l, and an ionic strength of 0.1 M.

take place in two steps, as indicated by the large difference in slope. At low pH values, there is a steep variation of the absorbed number of protons with pH, whereas at higher pH values this variation is significantly less pronounced. We shall argue that this points to the existence of two types of proton-binding process, and that the absorption of the protons not only decreases the overall charge on the subunits but in fact also involves the formation of Caspar pairs that strongly stabilize the helical state.

The remainder of this article is organized as follows. In the following section, we introduce our idealized model based on the principle of mass action. The potential of mean force between the protein subunits that enters our model consists of both attractive and repulsive contributions, and extends earlier work on hepatitis B virus capsid assembly (12). This potential of mean force, or binding strength, is a function of the temperature, the pH, and the ionic strength. An important experimental observation that we explicitly take into account is the absorption of protons by the CP subunits upon the formation of the helices. Although it is widely recognized that so-called ''Caspar'' carboxylate pairs are responsible for this, it has to our knowledge not yet been put into a predictive model.

In fact, we will make plausible that two types of Caspar pair regulate the transitions between the several aggregation states of the coat protein, i.e., between carboxyl groups located on

the same CP and between those on different CP subunits. The formation of the first presumably depends on pH only. That of the second type couples to the helix formation and depends sensitively on the ionic strength, the temperature, and the overall subunit concentration of the solution.

In the third section, we demonstrate that our model is indeed able to consistently describe the main features of the equilibrium aggregation behavior of the TMV coat protein. In particular, our model quite naturally explains the experimental fact that at low pH and increasing temperature, a transition from monomers to helices occurs, whereas at high pH, monomers are transformed into disk-like aggregates if the temperature is increased beyond some critical temperature. Our theory in fact makes a testable prediction, which is that the fraction of helices goes through a maximum as a function of temperature in a small range of intermediate pH.

Finally, in the fourth section, we compare our findings with results obtained for the hepatitis B virus capsid, which is spherical, not rodlike. The contact energies between the coat proteins that we find for these two very different types of virus are remarkably similar, suggesting that these might be universal.

TWO-STATE MODEL FOR DISKS AND HELICES

As may be inferred from the schematic drawing of a disk and a helix fragment in Fig. 2 *A*, disks must have a larger contact area per monomer than the other aggregation states, irrespective of the details of the molecular structure of the subunits and how they are arranged in the assemblies. Hence, if the bare attractive energy per unit area were the same in both aggregation states, monomers in disks would benefit from stronger attractive interactions than those in helices and should therefore be more stable. As can be seen in Fig. 1 *A*, at low pH this is not the case. In the model that we put forward, we attribute the stability of helices at low pH to an increase of the net binding strength resulting from proton binding by Caspar carboxylate pairs.

Two types of carboxylate pair will be considered. The groups of one type of pair are located on the same subunit, and have an anomalous dissociation constant (10). Protonation of this pair reduces the net charge on the CPs, thereby lowering the electrostatic repulsion between them. The other pair forms between groups on neighboring subunits, but only if they are part of a helix, because only then does the juxtaposition of the CPs allow for hydrogen bonds to form. The proton binding by these intersubunit Caspar pairs, apart from lowering the electrostatic repulsion between the CPs, contributes to the binding free energy between them.

Although hypothesized by Caspar (7), it should be stressed that the issue of whether carboxyl pairs between groups located on the same subunit actually form is not quite settled (see, e.g., the discussion in Lu et al. (13) and Wang et al. (14). It is important to note, in this context, that similar carboxylate pairs are known to play a significant role in structure formation in proteins and protein assemblies (see, e.g., Wohlfahrt et al. (15)).

As for the Caspar pairs between groups on neighboring subunits, there are very strong indications that the adsorption of protons forces a twist in the relative orientation of CP subunits (1) (see also the discussions in Caspar (7) and Butler et al. (10)). Adsorption of a proton by a carboxylate pair involving two carboxyl groups on different CP subunits is schematically illustrated in Fig. 3. In principle, a second proton can be taken up by the carboxylate pair, leaving a charge-neutral complex, but we do not consider this here as it only occurs at very low pH values where other carboxyl groups also get protonated.

If the amino acids putatively involved in the intersubunit carboxylate pair, Glu-50 and Asp-77, are removed and replaced by Gln and Asn, the rate of depolymerization of RNA-containing helices at high pH is reduced significantly

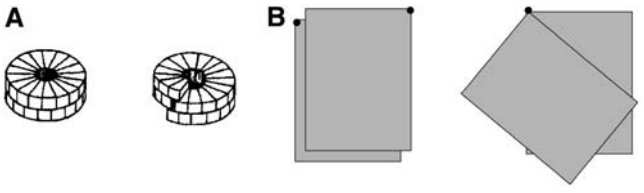

FIGURE 2 (*A*) Schematic drawing of a disk (*left*) and a helix (*right*). Note that the number of contacts (and hence the total contact area) between protein subunits is larger for disks than for helices. (*B*) Schematic of our model for subunits in a disk (*left*) and helix (*right*). Subunit bonding sites are modeled as rectangles, where the black dots symbolize carboxylate groups that may form ''Caspar'' pairs. Subunits in the disk configuration (*left*) are thought to maximize the interaction area. Helices (*right*) are in a way ''twisted,'' i.e., rotated relative to each other, so their interaction area is reduced. By twisting (or reorienting), they allow a ''Caspar'' carboxylate pair to form, which liberates a proton binding energy and reduces the electrostatic repulsion between the proteins.

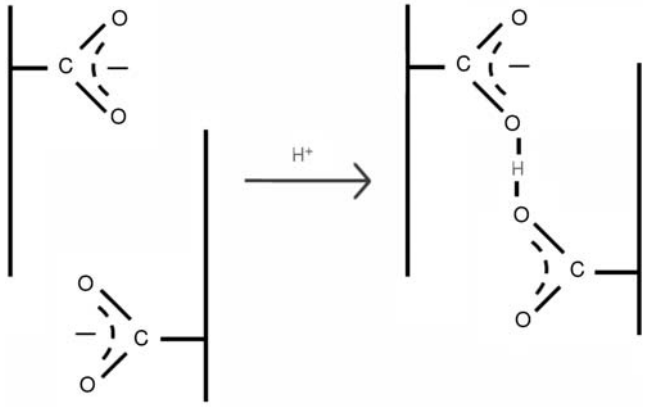

FIGURE 3 Schematic illustration of the protonation of a ''Caspar carboxylate pair,'' where the two carboxyl groups that form the pair are located on different subunits. In principle, a second proton can be taken up by the complex to neutralize the complex.

(16). Although this does not prove that helices become thermodynamically stable (as opposed to kinetically stable) at higher pH after replacing these amino acids, it does, in our view, suggest such a process. More convincing evidence would of course be a shift of the stability regions of the helices and disks as a function of pH and ionic strength, accompanied by a change in the acid-base titration curves of the subunits. We have not been able to find any studies in the literature that resolve the issue, so it remains somewhat contentious.

As helix formation is accompanied by proton adsorption (7,10), it seems sensible to model this by explicitly taking it into account in the chemical equilibrium between monomers and helices, which indeed is what we do. Our minimal model for disks and helices is given schematically in Fig. 2 *B*, where the dots represent the carboxyl groups involved in the stabilization of the helical configuration. The coaligned squares represent pairs of proteins in disk configuration (Fig. 2, *A* and *B*, *left*), and the ones rotated (or ''twisted'') relative to each other so as to bring into contact the carboxyl group involved in a Caspar pair represent proteins in helical aggregated states (Fig. 2, *A* and *B*, *right*). This does not imply that we presume the helix to be a twisted disk. It is not, because the disk is known to be bipolar, whereas the helix is polar (see Caspar and Namba (1)). In other words, the transition from disk to helix requires disassembly of the disk and does not form by simply twisting the CPs.

To describe the interconversion of the monomers, disks, and helices as a function of the ambient conditions, we assume, for simplicity, that the disks and helices consist of an equal number of monomers, $q = 34$, which is equal to the number of monomers in a two-layer disk. In reality, there is a broad distribution of helix sizes present in the solution under conditions in which they form (7). It so happens, however, that in helical aggregation, only the smallest stable size of linear equilibrium assemblies is relevant for establishing stability limits if a nucleation step is involved in the polymerization (17). Here, the nucleation step is provided by the relatively large aggregation number of the disk and lock-washer configurations. We (numerically) verified this by including all helix sizes, and found no significant impact on our results. Likewise, whereas in reality cylindrical aggregates of four and more layers may present themselves in the region(s) where we find two-layer disks, this does not alter the essential physics.

In the model, monomers, B, and q-mers in the form of disks, B_q^D, are thought to be in thermal equilibrium with each other. This may be expressed as

$$qB \leftrightarrow B_q^D. \tag{1}$$

For monomers to form a helix, B_q^H, we maintain that p protons, H^+, are absorbed, so

$$qB + pH^+ \leftrightarrow B_q^H. \tag{2}$$

Note that disks and helices may form via many intermediates and pathways. Within a statistical thermodynamic theory, however, only the properties of the initial and final states matter, so we need not consider the intermediate states here explicitly.

The reason CPs form assemblies is that this produces a free-energy gain. Let the sticking energy V of a pair of monomers be nil if they are free, $V = 0$, and negative if they are bound, $V < 0$. Suppose further that the total number of bonds in a disk is $n_D(q)$ and the number in a helix $n_H(q)$. This is equivalent to assuming that the effective interaction areas in disks and helices differ by a factor $n_D(q)/n_H(q)$, as schematically shown in Fig. 2.

We stress that in our coarse-grained description, the question of whether the 34 CP disk is made up of two 17 CP disks in a head-to-head or head-to-tail arrangement is irrelevant. For a discussion of this issue see, e.g., Caspar and Namba (1). We simply assume here that the total binding energy (excluding the contribution of the carboxylate pairs) is proportional to the contact area of the CPs in the different aggregation states, and watch where that simplifying assumption takes us.

Applying multichemical equilibrium statistical thermodynamics, we find for the mass fractions f_D of disks

$$f_D = qx_D/x = \frac{qx_1^q y^{n_D(q)}}{x}, \tag{3}$$

and for mass fractions of helices, f_H,

$$f_H = qx_H/x = \frac{qa_{H^+}^{p(q)} x_1^q y^{n_H(q)}}{x}. \tag{4}$$

In Eqs. 3 and 4, x_1, x_D, and x_H denote the mol fractions of monomers, disks, and helices, respectively. x is the overall mole fraction of CP subunits in the system. The quantity $y = e^{-V/kT}$ represents the Boltzmann weight of the CP sticking interaction, V, with k as Boltzmann's constant and T the absolute temperature. Essentially, $n_D(q)V$ and $n_H(q)V$ represent the chemical potentials of a disk and a helix of size q relative to those of q free monomers. The larger number of CP contacts in disks is reflected by $n_D(q) > n_H(q)$ for $q = 34$. Finally, for the Caspar pair we have ''activity'' $a_{H^+} \approx c_{H^+} e^{-\Delta\mu^0_{H^+}} = e^{-\Delta\mu^0_{H^+}} 10^{-pH}$, with c_{H^+} the dimensionless proton concentration (normalized to 1 M) and $kT\Delta\mu^0_{H^+}$ the chemical potential of a proton adsorbed by a carboxylate pair relative to that of a free proton in a 1-M solution.

As for the equilibrium fraction of monomers in the system, this follows from mass conservation, i.e., from the condition that

$$f_1 + f_D + f_H = 1, \tag{5}$$

with $f_1 = x_1/x$ the monomer fraction. Solving Eq. 5 with Eqs. 3 and 4 numerically for x_1 leads to the equilibrium fraction of all species, provided that the binding energy, V, is known. This, of course, is where the essential physics comes in.

We put forward the following ansatz for the potential of mean force between the CPs:

$$V = 0\,(\text{not bound}), \tag{6a}$$

and

$$V = V_{\text{attr}} + V_{\text{el}} \approx -\gamma A_{\text{p}} + akTz^2\kappa^{-1}\,(\text{bound}), \tag{6b}$$

based on the presumption of competing hydrophobic and screened Coulomb interactions between the coat proteins, as advertised in the Introduction. In the following, we discuss in more detail the origin of these two contributions to the binding strength and explain the symbols in the equations.

It is well established that the aggregation of CPs into disks and helices is promoted by increasing the temperature, at least up to their denaturation temperature, and that the aggregation process is endothermic (5,11). Only hydrophobic attraction between subunit monomers can plausibly explain the experimental observations. See also Kegel and van der Schoot (12). The attractive part, V_{attr}, of the binding potential V should then be equal to the gain in interfacial free energy of the hydrophobic patches on each protein subunit that are removed from contact with water in an assembly. Hence, $V_{\text{attr}} \approx -\gamma A_{\text{p}}$, with $\gamma \geq 0$ the surface tension of these hydrophobic patches in contact with the aqueous solvent, and A_{p} their total surface area.

As we demonstrated in Kegel and van der Schoot (12) in the context of hepatitis B virus capsid assembly, the temperature dependence of the binding strength is overwhelmingly dominated by the contribution from the hydrophobic interactions, at least for concentrations of inert salt not below a few tens of mM. For future reference, and following Kegel and van der Schoot (12), we therefore write

$$\begin{aligned} V_{\text{attr}}(T) &= V_{\text{attr}}(T_0) + \left(\frac{\partial V_{\text{attr}}}{\partial T}\right)_{\text{T=T}_0}(T - T_0) \\ &= A_{\text{p}}(-\gamma(T_0) + s_{\text{E}}(T - T_0)), \end{aligned} \tag{7}$$

with T_0 a reference temperature that we need not specify at this point. Here, the temperature coefficient is equal to the product of the contact area, A_{p}, and surface excess entropy, s_{E}. For macroscopic hydrophobic surfaces in contact with water $s_{\text{E}} < 0$ (18), explaining why the binding strength becomes stronger (more negative) with increasing temperature.

The repulsive contribution to the binding potential, $V_{\text{el}} \approx akTz^2\kappa^{-1}$, accounts for the electrostatic interactions between the net charge on the protein subunits, which are screened by the presence of inert salt. Here, a denotes a reciprocal length, specified below, and $\lambda_{\text{B}} = e^2/4\pi\varepsilon kT$ the Bjerrum length, with e the elementary charge and ε the solvent permittivity. The latter quantity is equal to ~0.7 nm in aqueous media and only weakly dependent upon the temperature (12). Finally, z denotes the number of charges on the surface of the protein subunits, and $\kappa^{-1} = (8\pi\lambda_{\text{B}}\rho)^{-1/2}$the Debye screening length, with ρ the number density of monovalent electrolyte. The Debye screening length is related to the concentration of 1:1 electrolytes, c_{s}, according to $\kappa^{-1} \approx 0.3/\sqrt{c_{\text{s}}}$ in units of nanometers if c_{s} is given in units of molarity.

The reciprocal length a is defined as $a = 4\pi GA^{-1}r\lambda_{\text{B}}$,with G a geometrical factor of order unity, the precise value of which depends on the size, shape, and relative orientations of the subunit patches involved in the Coulomb interactions. For instance, for infinitely large, parallel plates we have $G = 2$ (18). A is (to a first approximation) the surface area of the protein subunits in an assembly exposed to contact with water, and r the fraction of unit charges on a subunit patch involved in one single subunit-subunit contact. If the charge is homogeneously distributed over the subunits, then $r \approx A_{\text{p}}/A$. The fraction r will be absorbed into the quantities G and A_{p}. A natural, dimensionless quantity to compare between different virus CPs is $\lambda_{\text{B}}a$, which has a clear physical meaning: it quantifies the size of the charged surface on the scale of the Bjerrum length. Note that with the above ''continuum'' description, we ignore the impact of specific ionic interactions.

The strength of the electrostatic repulsion between the protein subunits depends on the ionic strength via the Debye length κ^{-1}, and on the acidity or pH via the net charge per subunit, z. The net charge per subunit, z, is the difference between the numbers of negatively and positively charged groups on its surface. For (native) TMV at around neutral pH, the primary structure of the CP subunits suggests that there must be 15 negatively and 12 positively charged groups. The α-amino terminus is methylated and therefore assumed to be uncharged (10). This makes $z = 3$ at pH values larger than, say, 7.5, but smaller than the pK_{a} values of the amino groups, which are ~8.5.

Titration experiments show that TMV coat protein subunits exhibit so-called anomalous behavior (7,10). As discussed in the introduction, at least one pair of carboxyl group residues per protein subunit gets protonated upon lowering the pH below the anomalous pK_{a} of ~7.1. It is believed that carboxyl groups located on the same CP subunits are responsible for this anomalous pK_{a} value (7,10,13), which is much higher than the typical value of ~2.

The issue is further compounded by the observation that protons are released upon depolymerization of the helix (irrespective of the presence or absence of viral RNA), indicating that, as discussed in the Introduction, assembly and proton binding are coupled. Again, this is plausibly due to two carboxyl residues located on two different subunits, the hydrogen bonding between which drives the helical configurational transition. Because of the coupling of the protonation, self-assembly, and the change of the protein binding, these carboxyl groups must also exhibit an anomalous pK_{a}. In Eqs. 2 and 4, this effect enters via the number of charges, z, that we presume to depend on the proton concentration, c_{H^+}, via

$$z = 2 + \alpha - p/q = 2 + \frac{1}{1 + k_a c_{\text{H}^+}} - p/q, \tag{8}$$

where $k_a = 10^{\mathrm{pK_a}}$, and where the degree of deprotonation of a single CP subunit is given by $\alpha = 1/(1+k_a c_{\mathrm{H^+}})$. This equation holds also for disks, except that for these, $p = 0$.

Equation 8 states that the total charge on a subunit depends on an intrinsic $\mathrm{pK_a}$, which would be that of a carboxylate pair located on the same subunit. The value of the intrinsic $\mathrm{pK_a}$ influences the degree of protonation of the subunit, α ($0 \leq \alpha \leq 1$). In addition to the intrinsic Caspar carboxylate pairs, we also have those that form between subunits, as illustrated in Fig. 3. Here, the adsorption of protons diminishes the (net negative) charge on a subunit by an amount equal to p/q. Their protonation is coupled to the thermodynamic state of the entire system.

We emphasize that in Eq. 8 we have assumed that the states of ionization of the other acidic and basic groups on the amino acid residues, as well as those on N- and C-terminal groups, do not depend on proton concentration. Within this approximation, these groups contribute a background charge equal to (minus) 2 in Eq. 8. Hence, we estimate that Eq. 8 should hold in the pH range from ~5 to 8.

In summary, the essential ingredients driving the transitions between monomers, disks, and helices are 1), hydrophobic interactions; 2), electrostatic interactions; and 3), subunit reorientation induced by proton adsorption and subsequent hydrogen-bonding. The strength of the hydrophobic interactions increases with temperature, driving the aggregation of monomers into either disks or helices, depending on the pH. The electrostatic self-repulsion destabilizes the aggregates, an effect that should become more important with increasing temperature. However, since the Coulomb self energy is scaled by the thermal energy kT in the Boltzmann weight, the temperature does not play as important a role in the electrostatics of the problem at hand as in that of the hydrophobic interactions (12). Electrostatic interactions are quite sensitive to the pH of the solution via the subunit charge z in Eq. 8, albeit this sensitivity becomes weaker with increasing ionic strength because of the effects of electrostatic screening by the ions.

The quantity driving the helical configurational ''transition'' of the assemblies is the excess chemical potential $\Delta\mu^0_{\mathrm{H^+}}$. This quantity gauges the affinity of protons for a pair of intersubunit carboxylate residues, and influences to what extent the actual equilibrium constant associated with protonation of an intersubunit carboxylate pair deviates from the bare $\mathrm{pK_a}$ if the pairing did not take place. An effective $\mathrm{pK_a^{(eff)}}$ can be defined to describe this, a quantity that depends on everything involved in the intersubunit interactions, in particular the coupled process of helix formation and protonation of the intersubunit carboxylates.

In view of the electrostatic nature of the proton adsorption, we ignore the temperature dependence of the Boltzmann weight that appears in Eq. 4, i.e., we assume that $a_{\mathrm{H^+}} \approx e^{-\Delta\mu^0_{\mathrm{H^+}}} 10^{-\mathrm{pH}}$does not depend on temperature, just as V_{el}/kT is to a good approximation an invariant of the temperature. A logical definition of $\mathrm{pK_a^{(eff)}}$ is that given by the condition that at the pH equal to $\mathrm{pK_a^{(eff)}}$, half the monomers have assembled into helices. Insisting that $f_1 = f_H = 1/2$, gives

$$2.303pK_{\mathrm{a}}^{(\mathrm{eff})} \approx \frac{1}{p(q)}[\ln q + (q-1)\ln(x/2) - n_{\mathrm{H}}(q)V(T)/kT] - \Delta\mu^0_{\mathrm{H^+}}. \quad (9)$$

This is the equation that we use to estimate a value of the quantity $\Delta\mu^0_{\mathrm{H^+}}$ from experiment.

Finally, we emphasize that we do not assume any mechanism by which the helices form. There is some controversy regarding this issue (see, e.g., Caspar and Namba (1) and Klug (2)). In our approach, we merely assume that cylindrical and helical aggregates form, and investigate under what conditions these aggregates should be stable.

RESULTS AND DISCUSSION

The model parameters

For parameter values, we assume $n_{\mathrm{D}} = 2q + 1$ and $n_{\mathrm{H}} = 2q - 3$, where $q = 34$, as already mentioned. These values follow from a simple geometric (lattice) model of disks and helices, in which each protein subunit has four sides that may contact with their neighbors. For both disks and helices, it can in that case be verified that the given number of contacts is maximal under the applied geometrical constraints. The total interaction area in disks is then a factor $n_{\mathrm{D}}(q)/n_{\mathrm{H}}(q) = 69/65 \approx 1.06$ larger than in helices, as schematically illustrated in Fig. 2 *B* in a somewhat exaggerated fashion.

We model the impact of the Caspar carboxylate pairs on the stability of the helices by presuming that the number of adsorbed protons upon helix formation is given by $p(q) = q - 1$, that is, for every monomer in contact with another in a helix, one proton is adsorbed. This is in accord with experimental observations (10).

The attractive part of the contact potential, $V_{\mathrm{attr}}(T_0)$, the length a, and the excess surface entropy, s_{E}, all influence the enthalpy per monomer, ΔH, which we take from the measurements published in Sturtevant et al. (11). Within our model, this quantity may be expressed as

$$\Delta H = \Delta h\langle n\rangle, \quad (10)$$

where $\langle n\rangle$ denotes the average number of bonds per monomer averaged over all species of aggregate in the solution,

$$\langle n\rangle = (n_{\mathrm{D}}x_{\mathrm{D}} + n_E x_{\mathrm{H}})/x, \quad (11)$$

and Δh is the enthalpy per bond, evaluated at a temperature T_0. The former quantity is given by

$$\Delta h = -kT^2\left(\frac{\partial(V/kT)}{\partial T}\right)_{\mathrm{T=T_0}} = V(T_0) - T_0 s_0, \quad (12)$$

where we introduced $s_0 = A_{\mathrm{p}}s_{\mathrm{E}}$. In this expression, we neglect potential contributions stemming from the protonation of the proteins. This seems reasonable, or at least consistent,

because in reality the buffers that are present in the solution also contribute to the enthalpy, but are difficult to estimate.

In Sturtevant et al. (11), a measured value of $\Delta H \approx +20\ kT_0$ was reported for $T_0 = 293$ K. The ratio of $V_{\rm attr}(T_0)$ and a follows from the shift of the heat capacity peaks with pH, also reported in that work (11) (see also below). The absolute values of these quantities are then fixed from a single-state point in Fig. 1 *A*, being the stability boundary of disks along the pH axis at an ionic strength of 0.1 M. This leads to the estimates $s_0 = -0.06k$, $V_{\rm attr}(T_0) = -8.5\ kT_0$, and $a = 0.45$ nm^{-1}. From the same state point, we obtain for the excess chemical potential of Caspar proton a value of $\Delta\mu^0_{H^+} = -12.5$ by demanding that $pK_a^{(\rm eff)} = 6.6$ in Eq. 9 at an ionic strength of 0.1 M for $T = 293$ K, as indicated by the pH where the stability boundary of the helix is located in Fig. 1 *A*.

The protein subunits are slightly wedge-shaped with the largest dimension ≈8 nm, width ≈3 nm, and thickness ≈2 nm (see Figs. 1 *A* and 2 *A*). As already mentioned, inspired by the structure of the disk in Fig. 2 *A*, we assume a total of four binding sites per CP, two of which are located on the radial sides of the CP, and two on the axial sides of the CP. Hence, we estimate an average interfacial area $A_{\rm p}$ per contact of ≈8 nm^2. With the uncertainty in the value of the geometrical constant G this fixes a within a rather broad range of $\sim 0.1 \le a \le 2.5$ nm^{-1}.

Properties of the coat protein of TMV: comparison with hepatitis B

Let us first compare the thermodynamic properties of the hydrophobic patches on the coat proteins of TMV and of HBV, and next discuss the properties of their charged patches.

The bare binding energy per protein subunit, $\gamma(T_0)A_{\rm mono}$, that we find equals $-(2q-3)V_{\rm attr}/qkT_0 \approx 16$ for helices, and $-(2q+1)V_{\rm attr}/qkT_0 \approx 17$ for disks (in units of thermal energy). Here, we estimated the hydrophobic interaction patch areas $A_{\rm mono} = (2q-3)A_{\rm p}/q$ and $A_{\rm mono} = (2q+1)A_{\rm p}/q$ for the helices and the disks, respectively. These estimates compare well with that for hepatitis B virus (HBV), which forms $T = 4$ icosahedral shells consisting of $q = 120$ dimeric subunits and for which the bare binding energy per subunit was found to equal ~19, again in units of thermal energy (12).

For the excess hydrophobic surface entropy per monomer, $s_{\rm E}A_{\rm mono}$, we find $(2q+1)s_0/q \approx -0.12k$ for disks, and $(2q-3)s_0/q \approx -0.12k$ for helices. These values are to be compared to $\approx -0.11k$ found for Hepatitis B virus capsids (12). From the excess surface free energy and entropy, the surface excess binding enthalpies per monomer, $h_{\rm E}A_{\rm mono}$, can be obtained because $h_{\rm E}(T_0) = \gamma(T_0) + T_0 s_{\rm E}$. We obtain values of ≈−19 for disks and ≈−18 for helices, to be compared with the value of ≈−13 for HBV. We have collected these thermodynamic data in Table 1, together with those for the reciprocal length a that is a measure of the charged area of every coat protein.

Comparing the electrostatic properties of TMV and HBV coat proteins is not so straightforward, not least because of the differences in geometry. To make headway, we first note that Eq. 6b tells us that the strength of the electrostatic repulsion per subunit contact at a given ionic strength must be proportional to az^2, where z is the net number of (interacting) charges on the subunits. The precise value of z varies with pH and with the aggregation state, and is different for TMV and HBV coat proteins. We take, as a measure for the strength of the electrostatic repulsion per monomer and per unit charge, $a' = (2q-3)a/q \approx 0.81$ nm^{-1} for the helices and $a' = (2q+1)a/q \approx 0.91$ nm^{-1} for the disks (see The model parameters). A measure for the strength of the electrostatic repulsion per monomer and independent of ionic strength is $\lambda_B a' z^2$ (see again discussion in The model parameters). Hence, with $2 < z < 3$ for TMV CP, we find a range of $2.2 < \lambda_{\rm B} a' z^2 < 5.7$ for CPs in disks, and $2.2 < \lambda_{\rm B} a' z^2 < 5.1$ for those in helices.

Extracting a comparable quantity for HBV follows from the identification $a' = A_{\rm C}\sigma^2\lambda_{\rm B}/qz^2$, where, according to Kegel and van der Schoot (12) $A_{\rm C}\sigma^2 \approx 1.2 \times 10^{21}$ m^{-2}. Here, $A_{\rm C}$ is the contact surface area per monomer of the (HBV) capsids, and σ is the surface charge density. From the primary structure of its capsid monomers, we deduce that $z = 4$ for HBV dimers at near-neutral pH, so that for HBV, $a' \approx 0.43$ nm^{-1}. Guessing that the net number of charges z on HBV coat protein at lower pH values is reduced to 3, the measure for the strength of the electrostatic repulsion varies between $2.7 < \lambda_{\rm B} a' z^2 < 4.8$. Obviously, this range compares very well with TMV CP.

It is tempting to conclude from Table 1 that, although the amino acid composition, tertiary structure, and geometry of the assemblies are quite different for the CP subunits of TMV and HBV, evolution provided them with comparable (coarse-grained) interaction energies. This suggests some form of universality in virus capsid formation. This would in fact make sense, because weaker interactions would not produce assemblies unless the monomer concentrations are very high. Much stronger interactions, on the other hand, would plausibly produce kinetic dead ends or ''traps'' of ill-formed aggregates.

TABLE 1 Values (dimensionless) of thermodynamic – and model quantities per protein subunit monomer for TMV species and Hepatitis B capsid

	$\gamma A_{\rm mono}/kT_0$	$T_0 s_{\rm E} A_{\rm mono}/kT_0$	$h_{\rm E}A_{\rm mono}/kT_0$	$\lambda_{\rm B}a'z^2$
TMV disk	17	−35	−19	2.2–5.7
TMV helix	16	−34	−18	2.2–5.1
HBV	19	−31	−13	2.7–4.8

The ranges in the last column are based on values of z between 2 and 3 for TMV, and between 3 and 4 for HBV.

Stability boundaries: impact of pH and ionic strength

We now investigate how the prevalence of the various species of assembly depend on the acidity and ionic strength of the solution, if we keep the temperature constant.

The stability line separating monomer- and disk-dominated regimes can be estimated by evaluating under what conditions their fractions are equal and $f_1 = f_D$. This produces the condition

$$\ln q + (q-1)\ln x_1^* - n_D V(T)/kT = 0,$$

with x_1* the ''critical'' monomer mol fraction where $x_1 = qx_D$. (See Eqs. 3 and 6.) Note that x_1* is the equivalent of a critical micelle concentration (18). Presuming that no helices are present, we may substitute $x_1^* \approx x/2$. With $\kappa^{-1} \approx 0.3/\sqrt{c_s}$, and the pH entering via the number of charges on a subunit, z, we get for the isothermal phase boundary in the c_s−pH plane,

$$c_s \approx \left(\frac{0.3az_{p=0}^2}{[\ln q + (q-1)\ln(x/2)]/n_D - V_{attr}/kT}\right)^2 \quad \text{(monomer-disk)}, \tag{13}$$

where $z_{p=0}$ indicates that we put $p = 0$ in the expression for the number of charges per subunit, Eq. 8.

Note that the right-hand side of Eq. 13 represents the ratio of the repulsive and attractive contributions to the binding potential. The attraction part of the sticking energy is renormalized by the loss of translational entropy upon assembly. Because of this renormalization, the stability-bound Eq. 13 depends logarithmically on the concentration of CPs in the solution, implying that its dependence on the total CP concentration is only weak.

Invoking similar arguments, we find the helix-monomer boundary to be given by

$$c_s \approx \left(\frac{0.3az_{p=q-1}^2}{[\ln q + (q-1)\ln(x/2) - 2.303\ \text{pH} - p(q)\Delta\mu_{H^+}^0]/n_H - V_{attr}/kT}\right)^2 \quad \text{(monomer-helix)}, \tag{14}$$

where $z_{p=q-1}$ indicates that we put $p = q - 1$ in Eq. 8.

Finally, we have the stability boundary between disks and helices, which follows from the condition $f_D = f_H$, which, if we presume the fraction free monomers to be negligible, leads to

$$c_s \approx \left(\frac{0.3a(n_H z_{p=q-1}^2 - n_D z_{p=0}^2)}{4V_{attr} - p(q)(\Delta\mu_{H^+}^0 + 2.303\ \text{pH})}\right)^2 \quad \text{(disk-helix)} \tag{15}$$

The reason that this stability limit does not depend on the concentration of CP subunits is that in our model, disks and helices are of the same size, and hence represent the same loss of translational entropy per monomer. Obviously, this is only approximately so.

The stability boundaries as predicted by Eqs. 13−15 are plotted in Fig. 4, which should be compared to the experimental stability diagram given in Fig. 1 *A*. We have also plotted, in the same figure, a cut through the diagram of states, showing the relative prevalence of the three species—monomer, disk, and helix—as a function of pH at constant ionic strength of 0.1 M. Clearly, Figs. 1 *A* and 4 are remarkably similar, at least if we assume that the actual boundary between disks and helices is indeed given by Eq. 15.

Although this is a gratifying result, it is important to point out that the ''phase boundaries'' shown in Fig. 1 are qualitative in the sense that they indicate where each species becomes detectable. This would be consistent with the arguments leading to Eqs. 13–15. Indeed, as can be seen in Fig. 4 (*right*), in the pH range where the helix fraction drops off from unity to zero, the three species of monomers, disks, and helices all appear in comparable quantities. So, in the pH range bounded by Eqs. 14 and 15, and indicated by M/H and M/D in Fig. 4, all three species are expected to appear in significant concentrations. This implies that the precise location of the phase boundaries is to some degree arbitrary, which makes it difficult to quantitatively compare experiment and theory.

Still, according to both Figs. 1 *A* and 4, the monomer-disk stability boundary lies at an almost constant ionic strength at high pH, whereas the helix-monomer stability boundary is at an almost constant pH, in particular at ionic strengths above, say, 0.1 M. The model reproduces the essential qualitative features of the stability boundaries and confirms the underlying physics assumed in the model: 1), the disk-helix transition is driven by the Caspar pair bonding; and 2), the monomer-disk transition is driven by the screening of the Coulomb interactions by the presence of inert salt.

Finally we note that the limiting value of the ionic strength at the theoretical monomer-disk stability boundary at high pH (as shown in Fig. 4) is smaller by almost a factor of 2 than the experimental value shown in Fig. 1 *A*. This should not worry us too much, because of the relative arbitrariness in the precise location of the phase boundaries as discussed.

Stability boundaries: impact of pH and temperature

Having dealt with how the acidity and the ionic strength regulate the relative concentrations of the monomers, disks

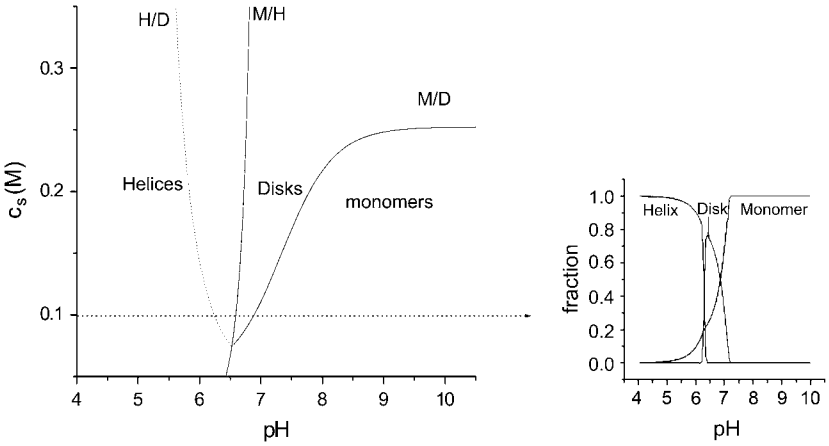

FIGURE 4 Calculated stability diagram of TMV coat proteins as a function of the salt concentration c_S and the pH, according to Eqs. 13–15 (cf. Fig. 1). The lines indicated by M/D, M/H, and H/D correspond to the crossovers between regimes dominated by monomers and disks, monomers and helices, and helices and disks, respectively. The total subunit concentration is $x = 5.4 \times 10^{-6}$ in mole fraction units, which is equivalent to that in Fig. 1, and the temperature is $T = 293$ K. The values of the quantities used to calculate the phase boundaries are $T_0 = 293$ K, $V_{attr}/kT = -8.5$, $a = 0.45$ nm^{-1}, and $\Delta\mu_{H^+} = -12.5$. The value of $s_0 = -0.06$ k was estimated from calorimetric data. See the main text for a discussion. At low ionic strength, our model predicts a direct transition from monomers to helices upon decreasing pH, i.e., without the appearance of disks. On the right, we show the fractions of monomers, disks, and helices as a function of pH at constant ionic strength of 0.1 M.

and helices at constant temperature, we now keep the concentration of added inert salt constant, and investigate in what way the pH and temperature shift the various stability boundaries.

Again we impose equal fractions of species in Eqs. 3–5 to find the crossovers between the monomer-, disk-, and helix-dominated regimes. Let $T_{\text{disk-mono}}$, $T_{\text{helix-mono}}$, and $T_{\text{helix-disk}}$ denote the crossover temperatures where disks and monomers, helices and monomers, and helices and disks coexist. We then obtain

$$T_{\text{disk-mono}} = \frac{T_0(\bar{V}_{\text{attr}}(T_0) - T_0\bar{s}_0)}{y_D - T_0\bar{s}_0 - az^2_{p=0}\kappa^{-1}}, \tag{16}$$

$$T_{\text{helix-mono}} = \frac{T_0(\bar{V}_{attr}(T_0) - T_0\bar{s}_0)}{y_H - T_0\bar{s}_0 - az^2_{p=q-1}\kappa^{-1}}, \tag{17}$$

$$T_{\text{helix-disk}} = \frac{T_0(\bar{V}_{\text{attr}}(T_0) - T_0\bar{s}_0)}{p(q)(2.303\,pH + \Delta\mu^0_{H^+})/4 - T_0\bar{s}_0 - a\kappa^{-1}(n_D z^2_{p=0} - n_H z^2_{p=q-1})/4}, \tag{18}$$

where for brevity we put $\bar{V}_{\text{attr}} = V_{\text{attr}}/kT_0$, $\bar{s}_0 = s_0/kT_0$, $y_D = (\ln q + (q-1)\ln(x/2)/n_D$, and $y_H = [\ln q + (q-1)\ln(x/2) - 2.303p(q)\text{pH} - p(q)\Delta\mu^0_{H^+}]/n_H$. Again, $z_{p=q-1}$ indicates that we put $p = q - 1$ in Eq. 8, with a similar prescription for $z_{p=0}$.

The phase boundaries that we obtain from these expressions are given in Fig. 5 for the representative case of a constant ionic strength of 0.1 M. The figure shows that at low pH and increasing temperature, there is a transition from monomers to helices, whereas at high pH, monomers form disks upon increasing temperature. This has indeed been found experimentally (6,11). Interestingly, at intermediate pH, our model predicts that first helices form but that, as the temperature is increased further, they disappear again in favor of the disks. This scenario is shown in the lower middle graph in Fig. 5. The reason for this behavior in a small pH window around pH 6.3 is the negative slope of the helix-disk boundary in the T-pH plane. This negative slope, in turn, is caused by the negative surface excess entropy of the hydrophobic patches, which in fact seems to be a generic property of hydrophobic surfaces (18,19).

In conclusion, we predict that within a small range of pH values helices appear and subsequently disappear upon increasing the temperature. At an ionic strength of 0.1 M, this small range amounts to ~0.1 pH unit around pH 6.2. This is a testable prediction.

Excess heat capacity

A powerful experimental tool allowing access to the thermodynamic properties of supramolecular assemblies, including virus capsids, is calorimetry. For our model of the self-assembly of TMV CPs, the excess heat capacity per monomer subunit is given by

$$\Delta c_p = \Delta h \frac{\partial\langle n\rangle}{\partial T}, \tag{19}$$

at least if we ignore the contribution from the ''breathing'' or ''phonon'' modes of the assemblies. Here, Δh is again the enthalpy per bond and $\langle n\rangle$ the number of bonds per monomer unit averaged over all assemblies. If we insert Eqs. 11 and 12 into Eq. 19, we obtain the results plotted in Fig. 6 for the same pH values as of the experimental curves of Fig. 5. The peaks have a shape similar to those of the experimental ones given in Sturtevant et al. (11) and reprinted in Fig. 1 *B* of this article, albeit the theoretical peaks are a little broader.

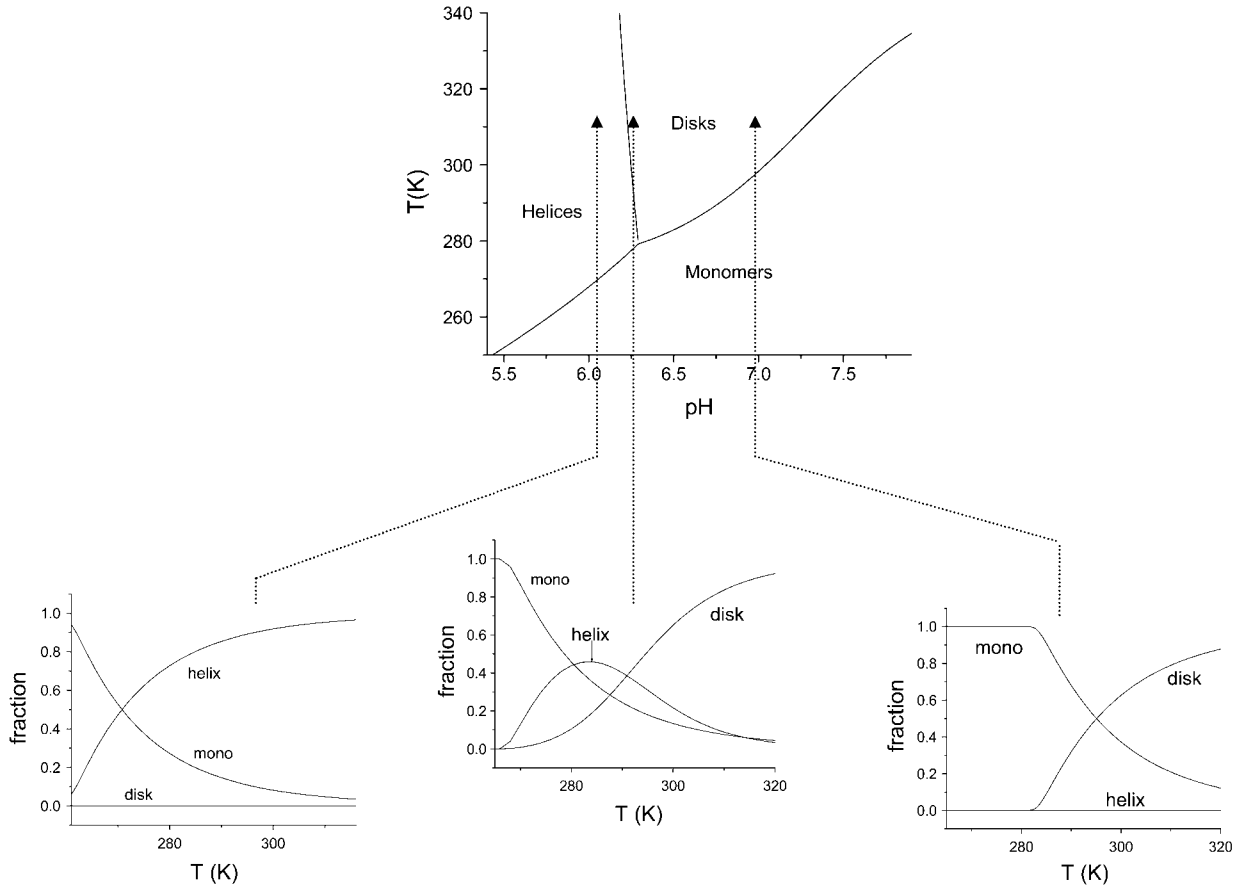

FIGURE 5 Stability diagram of TMV coat proteins as a function of the temperature T and the pH, at a constant ionic strength of 0.1 M. Shown also are the fractions of monomers, disks and helices as a function of temperature at pH 6.1, pH 6.3, and pH 7. At a low pH of 6.1, monomers directly transform into helices, whereas at higher pH (pH 7), a transition of monomers into disks is observed, in agreement with experiments (6,11). However, both disks and helices appear in appreciable concentrations at the intermediate pH 6.3. The helix fraction goes through a maximum as a function of T. Note that the stability lines at $T < 273$ K only make sense provided the aqueous solutions do not freeze.

We verified whether or not the apparent increase of the ''baseline'' observed in the experimental data shown in Fig. 1 *B* (11) is caused by the ''phonon'' term $\langle n\rangle\partial\Delta h/\partial T$, which we neglected in Eq. 19. This term does indeed lead to a linear increase of the baseline at temperatures beyond the peak of the excess heat capacity, but not over the whole range of temperatures as seen experimentally. We therefore speculate that the baseline drift must be caused either by contributions from the CPs themselves, e.g., from the electric double layer, by the (buffered) aqueous solvent itself, or by a combination of these. (For a discussion, see, e.g., Gallagher and Sharp (20)).

Titration curves

Information on the pH dependence of the number of adsorbed protons per protein subunit, Γ_{H}^{+}, is obtainable from acid-base titration measurements. In our model, it can be calculated straightforwardly, because

$$\Gamma_{\mathrm{H}^+} = 1 - \alpha + f_{\mathrm{h}}, \tag{20}$$

where we assumed that as a result of the presence and formation of Caspar carboxylate pairs, a maximum of two protons can be taken up by the subunits. One proton is taken up by the intra-CP carboxylate pair and one by the inter-CP pair (see also the discussion in the preceding section). The maximum of two protons agrees with experimental observations reported in Butler et al. (10) and Scheele and Lauffer (21).

Our theoretical titration curves are shown in Fig. 7, which should be compared with the experimental ones of Fig. 1 *C*. At the highest temperature shown in Fig. 7, the jump is steep, and coupled to the sudden occurrence of the helices. The shapes of the curves in Fig. 7, as well as the trend that the jumps of the curves shift to higher pH upon increasing the temperature, are in qualitative agreement with experiment (10). Moreover, the adsorption of two protons in two consecutive steps agrees with the experiments.

Note, however, that there are differences in the details of the titration curves of Figs. 7 and 1 *C*. For example, the influence of temperature on the steepness of the curves as found experimentally is not being reproduced by theory. In

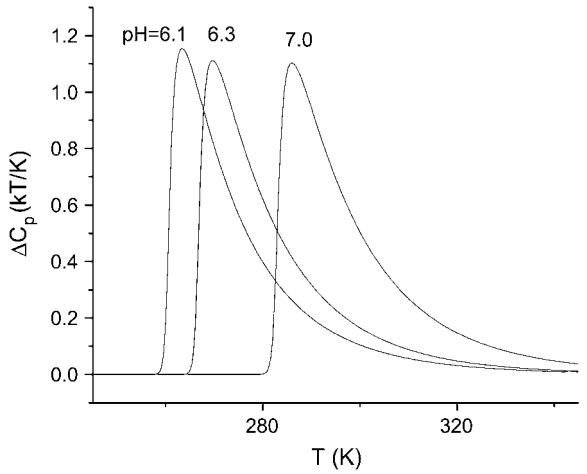

FIGURE 6 Excess heat capacity ΔC_p per protein subunit as calculated from Eq. 19 for the same three pHs as in Fig. 5. The shape of the curves agrees reasonably well with the experimental curves of Sturtevant et al. (11), reproduced in Fig. 1 *B* (but at different pH values). The curves have been calculated as described in the section Excess heat capacity.

fact, the trend in Fig. 7 even seems opposite to that in Fig. 1 *C*, but the situation is not completely clear, as the curve at highest temperature seems to deviate from all the others. Moreover, theory predicts that the transition from weak to strong pH dependence of the adsorbed protons is only weakly temperature dependent, whereas experimentally (Fig.1 *C*) the dependence on temperature seems much stronger.

Despite these quantitative differences, we suggest that the model does indeed capture the essential physics of the helix formation being coupled to the proton adsorption.

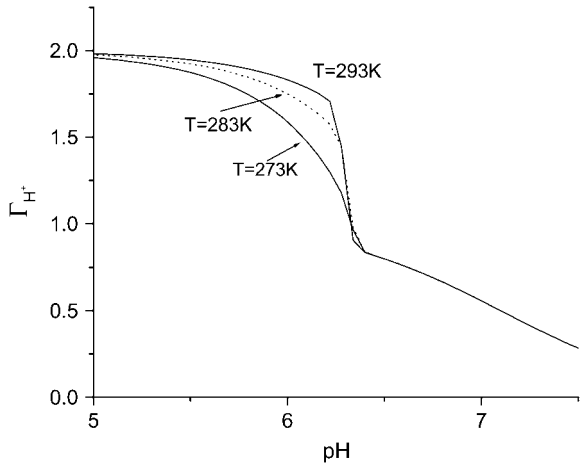

FIGURE 7 Titration curves under the same conditions as in the previous figures. Shown is the number of protons absorbed, Γ_{H+}, as a function of the degree of acidity, pH, for three different temperatures. The shape of the curves as well as their shifts upon increasing temperature are in qualitative agreement with experiments reported in Butler et al. (10) (cf. Fig. 1 *C*).

DISCUSSION AND CONCLUSIONS

In this work, we present a minimal model for the physical regulation of the in vitro self assembly of the coat protein of tobacco mosaic virus in the absence of its RNA. The main ingredients of the model are:

1. The principle of mass action producing disk-like and helical assemblies in solutions of CP subunits;
2. The presence of hydrophobic interactions between the coat proteins that drive the assembly into disks and helices and that predominate the temperature dependence of the self assembly;
3. The electrostatic repulsion between CPs in an assembly that couples to the ionic strength of the solution (through the effects of screening) and to the pH (by regulating the surface charge density on the CPs);
4. The coupling between the helix formation and the adsorption of protons by pairs of carboxylate residue located on neighboring CPs.

The model explains the main features of the in vitro aggregation behavior of TMV coat protein subunits as observed in experiment. Interestingly, to do that the model does not need to rely on any conformational changes of the CP subunits themselves. These have in the literature been suggested to be crucial to the assembly of CPs into capsids, see, e.g., Bruinsma et al. (22) and Ceres and Zlotnick (23).

Although CPs may or may not be subject to conformational switching upon assembly, we find that it is not necessary to explicitly take this into account in a theoretical description. This conclusion seems to hold for TMV as in fact it does for hepatitis B virus, as we showed in a previous work (12). In that work, we found that the competition of attractive hydrophobic and repulsive Coulomb interactions determine whether capsids of the HBV coat proteins form or not. This seems to be the case for TMV CPs too, except for the additional role of the Caspar pairs. Interestingly, the interaction parameters that we extracted for HBV are very similar to those we obtained for TMV, as shown in Table 1. It is tempting to speculate that these values are universal, that is, may be about the same for all viruses.

Novel in our approach is that we explicitly account for the anomalous titration behavior of the Caspar carboxylate pairs, and that we recognize their role in stabilizing the helical assembled state of TMV. They influence the stability of assembled species in two ways, through electrostatics and through hydrogen bonding. Both effects add to the net binding free energy of CPs in an assembly. It is important to point out that there are indications that Caspar pairs play a similar role in stabilizing other types of virus capsid, such as that of the icosahedral cowpea chlorotic mottle virus, to mention but one (see Tama and Brooks (24)).

Obviously, our model is only a minimal model in that it ignores many of the details of the actual interaction between the CPs of TMV, details that may be important under certain

conditions. Indeed, specific ionic components are known to strongly influence the aggregation behavior of TMV coat protein subunits (as for other viruses). For example, much work has been done on calcium ion binding (see, e.g., Caspar and Namba (1), Gallagher and Lauffer (25), and Einspahr and Bugg (26). It seems reasonable to suggest that carboxylate pairs also play an important role here, as they have strong affinity for calcium ions (26). In principle, the effect of calcium or other specific components that bind to them can be included in our model. For simplicity, we have in this work limited ourselves to nonspecific components such as protons and inert salt.

We also ignored in our model oligomers such as trimers. Inclusion of oligomers will not influence the stability boundaries between monomers, disks, and helices, but may influence the sharpness of the transitions between these states, e.g., in Figs. 4 (*right*) and 5 (*lower*) as a function of pH and temperature, respectively.

A moot point in the model is our tacit presumption that helices in the lock-washer state, as shown in Fig. 2 *A*, can actually become stable under the right conditions. This need not be so. An alternative scenario that could explain the stability, in particular, of long helices is that if the lock washer is metastable, helices can become stable by an explosive growth mechanism (17). The reason is that if a helix grows beyond the size depicted in Fig. 2, the number of subunit contacts per monomer grows with it (27). However, this scenario does not naturally explain the coupling between helix formation and proton adsorption that results in the ''anomalous'' titration behavior of CP subunits. Although we do not rule out the possibility that interactions are stronger in longer helices, this should not qualitatively change the picture that we sketched. Note also that calorimetric data do not indicate that the excess enthalpy per monomer is very different in helices compared with disks (11).

Perhaps more convincingly, the model presented here reproduces the well-known but hitherto unexplained observation that if we let the temperature go up, monomers become helices if the pH is low, but form disks at high pH (6). A crucial test would be the experimental verification of the helix fraction going through a maximum, which, according to our prediction, should occur as a function of temperature within quite a narrow pH range (Fig. 5). Clearly, a more quantitative validation of the theory is only possible if the fractions of the various species of assembly are known experimentally, which at present they are not. Our hope is that this article encourages workers in the field to reinvestigate in more detail the assembly behavior of TMV CPs.

REFERENCES

1. Caspar, D. L. D., and K. Namba. 1990. Switching in the self-assembly of tobacco mosaic virus. *Adv. Biophys.* 26:157–185.
2. Klug, A. 1999. The tobacco mosaic virus particle: structure and assembly. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 354:531–535.
3. Butler, P. J. G. 1999. Self-assembly of tobacco mosaic virus: the role of an intermediate aggregate in generating both specificity and speed. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 354:537–550.
4. Fraenkel-Conrat, H., and R. C. Williams. 1955. Reconstitution of active tobacco mosaic virus from its inactive protein and nucleic acid components. *Proc. Natl. Acad. Sci. USA.* 41:690–698.
5. Durham, A. C. H., J. T. Finch, and A. Klug. 1971. States of aggregation of tobacco mosaic virus protein. *Nat. New Biol.* 229:37–42.
6. Durham, A. C. H., and A. Klug. 1971. Polymerization of tobacco mosaic virus protein and its control. *Nat. New Biol.* 229:42–46.
7. Caspar, D. L. D. 1963. Assembly and stability of the tobacco mosaic virus particle. *Adv. Protein Chem.* 18:37–121.
8. Lauffer, M. A. 1966. Polymerization-depolymerization of tobacco mosaic virus protein. VII. A model. *Biochemistry.* 5:2440–2446.
9. Smith, C. E., and M. Lauffer. 1967. Polymerization-depolymerization of tobacco mosaic virus protein. VIII. Light-scattering studies. *Biochemistry.* 6:2457–2465.
10. Butler, P. J. G., A. C. H. Durham, and A. Klug. 1972. Structures and roles of the polymorphic forms of tobacco mosaic virus protein. *J. Mol. Biol.* 72:1–18.
11. Sturtevant, J. M., G. Velicelebi, R. Jaenicke, and M. A. Lauffer. 1981. Scanning calorimetric investigation of the polymerization of the coat protein of tobacco mosaic virus. *Biochemistry.* 20:3792–3800.
12. Kegel, W. K., and P. van der Schoot. 2004. Competing hydrophobic and screened-Coulomb interactions in hepatitis B virus capsid assembly. *Biophys. J.* 86:3905–3913.
13. Lu, B., G. Stubbs, and J. N. Culver. 1996. Carboxylate interactions involved in the disassembly of tobacco mosaic tobamovirus. *Virology.* 225:11–20.
14. Wang, H., A. Planchart, and G. Stubbs. 1998. Caspar carboxylates: the structural basis of tobamovirus disassembly. *Biophys. J.* 74:633–638.
15. Wohlfahrt, G., T. Pellikka, H. Boer, T. T. Teeri, and A. Koivula. 2003. Probing pH-dependent functional elements in proteins: modification of carboxylic acid pairs in *Trichoderma reesei* Cellobiohydrolase Cel6A. *Biochemistry.* 42:10095–10103.
16. Culver, J. N., W. O. Dawson, K. Plonk, and G. Stubbs. 1995. Site-directed mutagenesis confirms the involvement of carboxylate groups in the disassembly of tobacco mosaic virus. *Virology.* 206:724–730.
17. van Gestel, J., P. van der Schoot, and M. A. J. Michels. 2003. Role of end effects in helical aggregation. *Langmuir.* 19:1375–1383.
18. Israelachvili, J. 1992. Intermolecular and Surface Forces. Academic Press, San Diego.
19. Claesson, P. M., R. Kjellander, P. Stenius, and H. K. Christenson. 1986. Direct measurement of temperature-dependent interactions between nonionic surfactant layers. *J. Chem. Soc. Faraday Trans. l.* 82:2735–2746.
20. Gallagher, K., and K. Sharp. 1998. Electrostatic contributions to heat capacity changes of DNA-ligand binding. *Biophys. J.* 75:769–776.
21. Scheele, R. B., and M. A. Lauffer. 1967. Acid-base titrations of tobacco mosaic virus and tobacco mosaic virus protein. *Biochemistry.* 6:3076–3081.
22. Bruinsma, R. F., W. M. Gelbart, D. Raguerra, J. Rudnick, and R. Zandi. 2003. Viral self-assembly as a thermodynamic process. *Phys. Rev. Lett.* 90:248101.
23. Ceres, P., and A. Zlotnick. 2002. Weak protein-protein interactions are sufficient to drive assembly of hepatitis B virus capsids. *Biochemistry.* 41:11525–11531.
24. Tama, F., and C. L. Brooks. 2002. The mechanism and pathway of pH induced swelling in cowpea chlorotic mottle virus. *J. Mol. Biol.* 318: 733–747.
25. Gallagher, W. H., and M. A. Lauffer. 1983. Calcium ion binding by tobacco mosaic virus. *J. Mol. Biol.* 170:905–919.
26. Einspahr, H., and C. E. Bugg. 1984. Crystal structure studies of calcium complexes and implications for biological systems. *In* Metal Ions in Biological Systems. Marcel Dekker, NY. 51–97.
27. Oosawa, F., and M. Kasai. 1962. A theory of linear and helical aggregations of macromolecules. *J. Mol. Biol.* 4:10–21.