Eur. Phys. J. E **19**, 303–310 (2006)
DOI 10.1140/epje/i2005-10071-1

THE EUROPEAN PHYSICAL JOURNAL E

Physics of RNA and viral assembly

R.F. Bruinsma[a]

Department of Physics and Astronomy, The University of California at Los Angeles, Los Angeles 90049, CA, USA

Received 11 August 2005 /
Published online: 23 March 2006 –

Abstract. The overview discusses the application of physical arguments to structure and function of single-stranded viral RNA genomes.

PACS. 87.15.Nn Properties of solutions; aggregation and crystallization of macromolecules – 61.50.Ah Theory of crystal structure, crystal symmetry; calculations and modeling – 81.16.Dn Self-assembly

The other contributions to this Focus Point discuss how physical descriptions of DNA and RNA can illuminate our understanding of cell functioning. Here, the focus will be on *dys*functional nucleic acids, specifically the genomes of infectious viruses. A virus is a shadowy citizen of a borderland between living and dead matter; a parasite that reproduces only in the environment of a host cell, in fact by the mobilization of the macromolecular machinery of the host. A virus also does not carry out any metabolic activity, which means that, unlike living cells, it effectively is in a state of *thermal equilibrium* with its environment; one reason why physics —and statistical physics in particular— can be a useful tool. Another reason is apparent from an image reconstruction of a virus. Many spherical viruses, such as the MS2 virus shown in Figure 1, exhibit *icosahedral symmetry*.

Note that the image is oriented along one of the twelve five-fold icosahedral axes. MS2 belongs to the $T = 3$ structural class, an extended group of small RNA viruses that also includes the polio and common cold viruses. The image only shows the outer protein shell, or capsid, that encloses the nucleic acid material. For a $T = 3$ virus this capsid consists of precisely 180 proteins. The pioneers of structural studies of viruses —some of which had a background in physics such as Francis Crick, Donald Caspar, and Aaron Klug— viewed the capsid as a curved, two-dimensional crystal closed on itself, and they borrowed concepts from crystallography to describe the different icosahedral tiling patterns of viral shells [1].

The genome enclosed by the capsid is composed of one or more DNA or RNA molecules that are either single stranded (ss) or double stranded (ds). The focus of this contribution will be on ss RNA genomes enclosed by a $T = 3$ capsid, and the MS2 virus will serve as our prototype. It is a "bacteriophage", *i.e.* a virus that infects bacteria (*E. Coli* in this case), and also something of a venereal-disease carrier since infection of an *E. Coli* bacterium can only take place when it exchanges genetic material with another bacterium. Figure 2 shows a schematic cross-section of a $T = 3$ virus like MS2.

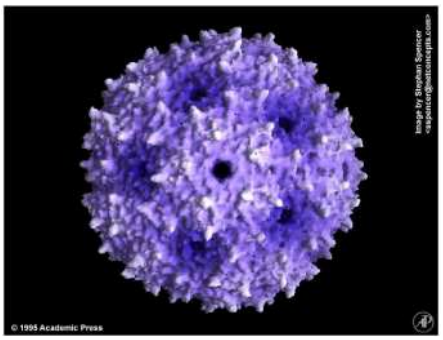

Fig. 1. Image reconstruction of the MS2 virus, viewed along a 5-fold icosahedral symmetry axis. From reference [2].

The outer radius R^* of the shell is in the 10 nm range and the inner radius is about 8 nm. The MS2 genome is composed of a single ss RNA molecule of about 4000 bases and contains four genes [3]. The "coat" gene is the code for the synthesis of a protein that is the main component of the MS2 capsid. A second gene encodes the "Replicase" protein, a molecular copying machine of viral RNA molecules. Infection starts when a viral RNA molecule is released out of the capsid and injected into the cytoplasm of a host cell. Next, a host Ribosome attaches to the RNA molecule and synthesizes a Replicase protein from the replicase gene template. This Replicase protein can make new copies of the viral genome. As this process is iterated again and again, more and more Replicase proteins and viral RNA molecules are synthesized, as well as capsid proteins. Manfred Eigen studied the population dynamics of this process and found a *hyperbolic divergence*

[a] e-mail: bruinsma@Physics.ucla.edu

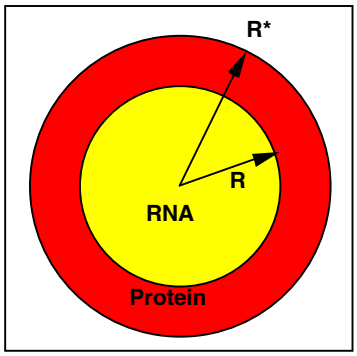

Fig. 2. Schematic cross-section of a $T = 3$, single-stranded RNA virus. The 180 protein capsid shell has an outer radius R^* of about 10 nm and an inner radius R of about 8 nm. It encloses a genome of one or more RNA molecules with a total of about 4000 bases.

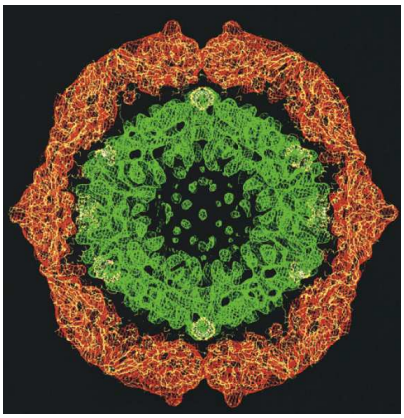

Fig. 3. A cross-section of the electron density map of the Flock-House Virus. The density shown in red on-line corresponds to the protein capsid, while density in green on-line corresponds to the RNA genome or to disordered protein segments. From reference [7].

of both the viral protein and the viral RNA molecules after a certain delay time [3]. Now, *in vitro* studies of solutions of MS2 capsid proteins and RNA molecules show that —under the right conditions of *p*H and salinity— they can *spontaneously self-assemble* into infective MS2 viruses [4]. A similar self-assembly process in a host cell would lead to the production of a swarm of new viruses. A third gene, "lysis", generates a protein that stimulates the dissolution of the host cell. The swarm is released into the environment, ready to infect other bacteria.

We will focus here on two questions: first, what are the structure and physical conditions of an RNA molecule inside a virus and, second, what is the physics of the amazing protein/RNA self-assembly process. To illustrate the fact that physical considerations can be useful in this respect, let us estimate the *maximum* number of genes of a $T = 3$ virus like MS2 [5]. Reasonably, the maximum density of an enclosed viral RNA genome is that of RNA molecules condensed into a *crystal* by condensing agents such as polyvalents ions. Molecular biologists choose to express packing densities in units of cubic angstroms per dalton (really the *inverse* of a density). Hydrated RNA crystals have a packing density $V_{\rm m}$ of about 2.2 cubic angstrom per dalton. It follows that we can pack a maximum of $(V/V_{\rm m})$ daltons worth of RNA in an enclosing volume V. For a $T = 3$ virus, this amounts to about 10^3 kilo daltons (kD), since $V = \frac{4}{3}\pi R^3$ is about 2×10^6 cubic angstroms. The molecular mass of one RNA nucleotide happens to be about 320 daltons, while the RNA code of one gene involves about 700–1000 nucleotides. The molecular mass of one gene is then about 200–300 kD so the molecular mass of four genes is about 800–1200 kD.

It follows that four to five genes is about the *maximum* amount of genetic information that can be packed on the genome of a $T = 3$ virus, an economy of design that compares rather favorably with the large amount of non-coding DNA we carry around. The key point here is that a purely physical constraint appears to be limiting the maximum size of the genetic program that a $T = 3$ virus can carry out [6].

So not only can we view the viral capsid as a *two*-dimensional protein crystal but the viral interior should be compared with a *three*-dimensional RNA solid. This RNA solid has unusual structural properties. We noted that the capsid has icosahedral crystal symmetry. This same icosahedral symmetry is imposed as well, to some degree, on the outer part of the genome. Figure 3 shows the reconstruction of the 4800–base-pair ss RNA genome inside the $T = 3$ Flock-House Virus (FHV), as obtained from an X-ray diffraction study by the group of J. Johnson [7].

Icosahedral symmetry is imposed on the outer part of the genome, but not on the inner part. Raman spectroscopy studies [8] indicate that the genome of $T = 3$ ss RNA viruses largely is in *double-helical* A-form (about 70%). Condensation apparently favors the paired state, presumably because of the reduced conformational entropy of unpaired bases in the crowded environment of the virus interior. Icosahedral genome patterns are only observed for *single-stranded* genomes. For instance, the linear, double-stranded genome of ds DNA bacteriophages is arranged in a spool-like, toroidal structure with no icosahedral symmetry at all.

It is interesting to compare these neatly folded, icosahedral RNA packages with the structure of the same viral RNA molecule *outside* the virus. Stretched out, an MS2 genome would have a total length of about 1000 nm. The actual size of the genome molecule in solution is smaller than that, not only because of the usual polymer entropic elasticity, but also because of secondary and tertiary structure formation. Scattering studies of the MS2 genome in solution under physiological conditions [9] indicate that it has the shape of a cylindrical particle with a length of about 300 nm and a lateral diameter of about 10 nm. The size of the particle is rather sensitive to the ambient Mg^{++} ion concentration, which is in fact an RNA condensing agent. The nucleotide density of viral RNA in solution is thus about ten times less than that of the interior of the MS2 virus, so a substantial level of RNA compaction is required during assembly.

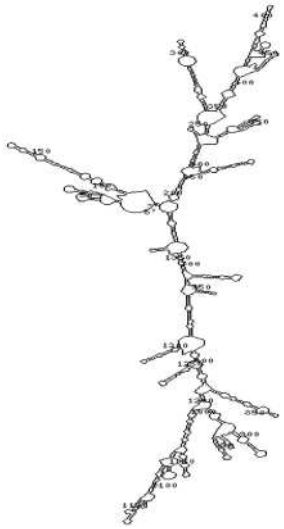

Fig. 4. Typical secondary structure of one of the two ss RNA molecules of the *Nodaviridae* computed using the Mfold program.

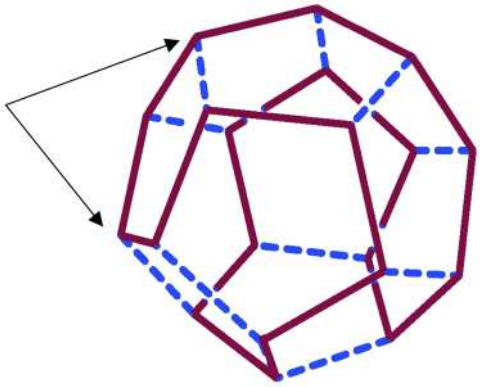

Fig. 5. Solid line: Hamiltonian cycle on a dodecahedral cage. Bubbles or branch points must be placed at the vertices (arrows) in order to accommodate sharp turns. Dashed lines: edges that are not covered by the main chain and that must be occupied by side branches.

The secondary structure of the viral RNA molecule plays an important role during assembly. Using a standard numerical package —"Mfold" [10]— one can compute secondary structures of the two ss RNA molecules of FHV. A typical result [11] is shown in Figure 4.

This particular structure is not special but only one member of a family of low-energy secondary structures, with a typical free energy difference of the order of $5\,k_{\rm B}T$. The secondary structure resembles a branched polymer. The branches consist of paired sequences alternating with unpaired "bubbles". The side branches terminate in *stem-loop sequences*. Certain particular stem-loop sequences have an affinity for the capsid proteins of the virus, and act as recognition labels for encapsidation [5]. This affinity would "anchor" these stem-loops to the capsid wall. Because of this stabilization effect, it seems likely that the branching topology of the RNA secondary structure in solution is largely maintained inside the virus, though we should expect an increased amount of double-helical sections as compared with Figure 4.

Now, it obviously is not possible to fold this —seemingly— randomly branched polymer into a *perfectly* icosahedral structure. In fact, for purely mathematical reasons, it is not even possible to fold a linear polymer into an icosahedral structure. However, it *is* possible to fold a branched polymer into a *partially* icosahedral structure. The key is to view Figure 4 as a main-chain with side-branches [11]. Take the main-chain and fold it over the edges of a *dodecahedral cage* (a dodecahedral cage has icosahedral symmetry). Demand that the main chain —assumed double-stranded— visits every one of the 20 vertices of the cage only once. It does not have to visit every one of the 30 edges, which would be possible only by multiple visits to the same edge. This we will forbid on the basis of self-repulsion between ds RNA strands, as discussed below. The resulting trajectory traced out by the main chain is known as a "Hamiltonian Path". If we also demand that the initial and final points are the same, then it would be called a "Hamiltonian Cycle". On a dodecahedral cage, there is only one Hamiltonian Cycle —the one shown in Figure 5— modulo trivial symmetry permutations.

In view of the bending rigidity of double-stranded polynucleotides ($k_{\rm B}T/500\,$Å), we must place bubbles and branch-points of the secondary structure at the vertices (arrows) in order to accommodate the necessary main-chain kinks. The final step is to decorate the remaining, unoccupied edges (dotted lines) with side branches, assuming these are properly spaced.

The product of this exercise is a partially icosahedral structure. Note that the edges of the cage are structurally homogeneous —ds RNA strands— but that the vertices are heterogeneous. Some vertices are branch points and others are branch end-points. Could this (speculative) idea be checked? A $T = 3$ virus with an exceptional amount of genomic icosahedral order is the *Pariacoto* virus. The image reconstruction is shown in Figure 6.

The dodecahedral organization is a prominent structural feature. The edges are well-ordered, double-helical strands, while the vertices are indeed more disordered. This is consistent with the model, but current high-resolution structure determination methods of viruses always rely on icosahedral averaging, in order to improve the resolution, so imaging of non-icosahedral features, such as the proposed vertex structure, is not yet possible. Note that each face of the dodecahedron forms the base of a pentagonal pyramid of 15 capsid proteins that are part of the capsid shell (blue on-line).

Let us turn to the physical condition of the genome. The ten-fold compression of the genome during the assembly process noted earlier is expected to produce a pressure inside a virus that is exerted on the capsid wall. If we consider the capsid wall as a semi-permeable surface that allows water molecules but not nucleotides to pass through, then we can view the pressure on the capsid wall as an *osmotic* pressure. Let us assume that the compressed ss RNA molecule is largely in duplex form. If an A form ds

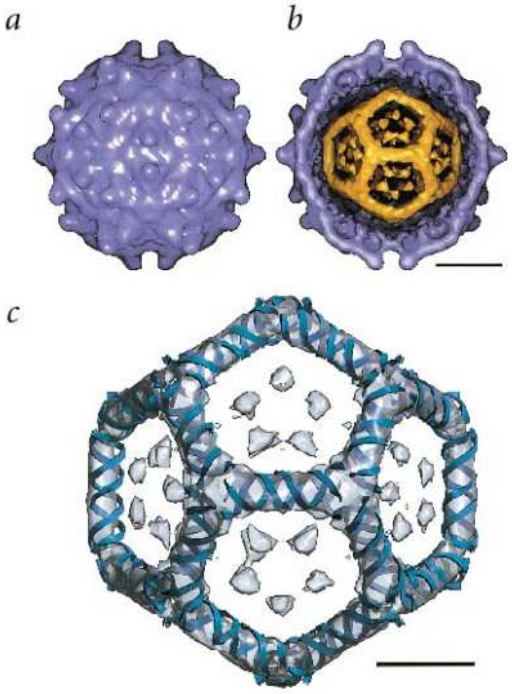

Fig. 6. Image reconstruction of the *Pariacoto* virus obtained from cryo-electron micrographs by Tang *et al.*, reference [12]. a) Capsid viewed along a 2-fold icosahedral axis. b) Same view of the genome structure (yellow on-line, bar is 10 nm). c) Double-helical sections of the genome arranged along the edges of a dodecahedral cage (bar is 5 nm).

RNA strand (helical rise of 2.7 angstrom) with a length L of 1700 base-pairs (half of the MS2 genome) is compacted homogeneously into a sphere of radius $R = 10\,\mathrm{nm}$, then the mean spacing D between the ds RNA strands can be obtained from the packing condition $(\pi/4)D^2L = V$. This gives a value for D of about 4 nm. The equation of state $\Pi(D)$ relating pressure to D spacing is unfortunately not known for ds RNA, but the equation of state of a hexagonal bundle of ds DNA was determined by Rau and Parsegian [13] by exposing the DNA bundle to a PEO solution with a known osmotic pressure and then measuring the D spacing by X-ray diffraction. Using their results to estimate $\Pi(D)$, one finds that for $D = 4\,\mathrm{nm}$, the osmotic pressure of the viral interior should be in the range 1–10 atm under physiological conditions. This pressure is due to a combination of electrostatic repulsion between the negatively charged phosphate groups of the nucleotides and counter-ion confinement [14]. In an elegant study by the UCLA group [15], the osmotic pressure of the ds DNA genome inside a λ phage virus was measured and found to be of the order of tens of atmospheres, consistent with the somewhat higher packing density.

The pressure inside the capsid will produce a (2D) lateral stress τ along the capsid surface. According to Laplace's Law $\Pi = \frac{2\tau}{R}$, this stress would be in the range 10^{-3}–$10^{-2}\,\mathrm{N/m}$. Interestingly, this osmotic stress is nearly comparable to the *rupture stress* of a $T = 3$ viral capsid, measured in an AFM study [16], that lies in the range 10^{-2}–$10^{-1}\,\mathrm{N/m}$. The *work* W required to compress the solution RNA coil into the enclosing capsid against the electrostatic self-repulsion also can be computed from the equation of state, as $W = \int \Pi(V)\mathrm{d}V$ integrated under the constraint $V(D) = (\pi/4)D^2L$. It is of the order of 100–1000 $k_\mathrm{B}T$, again using the Parsegian and Rau data.

For the case of the ds DNA phages, we know that this work is performed by a powerful packaging motor that inserts the DNA strand into a pre-assembled capsid head. The energy source of the motor is ATP hydrolysis. Measurement of the force-velocity curve [17] of this motor shows that the insertion velocity slows down when the shell starts to fill up. The stalling force of the motor is thus probably determined by the osmotic pressure of a filled capsid. For a λ phage, the stored work of genome compression is an energy source for the injection of the phage genome into the host cell. Note that DNA λ phage genome molecules appear to be the rather passive subjects of a lot of pushing and pulling during viral assembly and disassembly.

There is no such packaging motor available (nor, perhaps, even possible) for MS2 and other viruses with flexible and polymorphic ss RNA genomes. RNA viral genomes actually play a surprisingly *active* role during their own encapsidation, and possibly in their release as well, which is in line with the fact that, unlike ds DNA, certain RNA molecules —known as Ribozymes— can carry out enzymatic activity. The free-energy source for ss RNA encapsidation is not ATP or GTP hydrolysis but a non-specific affinity between capsid proteins and RNA molecules, usually of *electrostatic* origin, plus the specific interaction of the stem-loop sequences with the capsid proteins. Capsid proteins typically carry a net positive charge of 10–13 elementary charges. For example, crystallographic studies of the $T = 3$ CCMV plant virus [18] (CCMV stands for *Cowpea Chlorotic Mottle Virus*) show that the ordered part of the capsid protein has positively charged residues facing the viral interior. In addition, a disordered N terminus tail, which extends into the interior and which is composed of 26 residues, has a positive charge of 10 elementary charges. If one estimates the electrostatic condensation energy of a $T = 3$ virus like CCMV as the number of capsid proteins (180) times the number of positive charges per protein (10) times a few $k_\mathrm{B}T$'s (for a polar bond), one finds that the available non-specific electrostatic nucleo-protein affinity must be of the order of $10^3\,k_\mathrm{B}T$.

Electrostatic interactions also are likely to be responsible for the icosahedral patterning of the genome that we discussed earlier. If we put the positively charged proteins together into an icosahedral shell, we will produce an electrostatic potential inside the shell that has icosahedral symmetry as well. It follows from elementary electrostatics that this potential must be large along sharp folds of the shell. We expect that the negatively charged genome will try to fold itself into an icosahedral pattern with the highly charged duplex strands nested in the folds of the protein shell [19].

The role of electrostatics in ss RNA controlled viral assembly is actually rather subtle, as illustrated by the

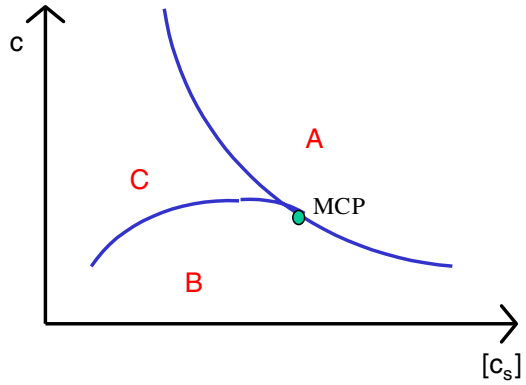

Fig. 7. Theoretical co-assembly phase diagram of charged polyelectrolytes with oppositely charged proteins. The vertical axis is the protein concentration and the horizontal axis the salt concentration. B Phase: polyelectrolyte and protein molecules do not assemble. A Phase: protein molecules assemble into empty capsids, through hydrophobic attraction, as the protein concentration increases. C Phase: polymer-filled capsids self-assemble as the protein concentration increases. The multicritical point MCP corresponds to the polyelectrolyte desorption threshold.

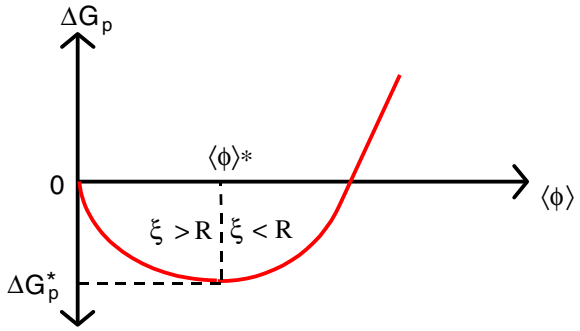

Fig. 8. Gibbs free energy $\Delta G_{\rm p}$ of a polyelectrolyte-filled protein capsid as a function of the polyelectrolyte monomer concentration $\langle\phi\rangle$. The free energy first decreases and then reaches a minimum at $\langle\phi\rangle = \langle\phi\rangle^*$. At this minimum the charge of the polyelectrolyte material is double that of the capsid charge while the correlation length ξ of the semi-dilute polymer solution is comparable to the radius R of the capsid. The osmotic pressure is close to zero for $\langle\phi\rangle$ less than $\langle\phi\rangle^*$, while for $\langle\phi\rangle$ greater than $\langle\phi\rangle^*$, the osmotic pressure starts to rise.

CCMV assembly phase-diagram [20]. Under conditions of physiological *p*H and salinity, CCMV capsid proteins do not assemble into empty capsids, at least not at low concentrations. The electrostatic *repulsion* between the positively charged capsid proteins exceeds any hydrophobic attraction between the proteins. However, infectious viral particles form after CCMV RNA genome molecules are added to the solution. Exceptionally, CCMV proteins are promiscuous polymer packers. They co-assemble not just with viral RNA molecules, but also with non-viral RNA molecules and even with *generic* anionic polyelectrolyte molecules.

The electrostatic *attraction* between the genome molecules and the proteins apparently suffices to overcome both the electrostatic repulsion between capsid proteins and the electrostatic self-repulsion of the genome molecules that we discussed earlier. We thus should expect the free energy of assembly of a CCMV virus to be of the order of the estimated $10^3\, k_{\rm B}T$ electrostatic affinity between genome and capsid. An *in vitro* study of CCMV assembly confirms this estimate [21]. Note the elegance: electrostatics both prevents the assembly of *empty* protein shells and drives the assembly of *filled* shells.

Could we use the theory of aqueous electrostatics to actually predict the osmotic pressure inside an RNA virus that assembled under equilibrium conditions, *i.e.* with a nucleotide density that minimizes the free-energy minimization? This is complicated because of the secondary and tertiary structure of the RNA molecule but we can consider the simpler case of the packing of generic linear polyelectrolyte molecules. The adsorption of polyelectrolyte molecules on an oppositely charged surface is a classical problem of polymer physics. A thermodynamic analysis [22] —that borrows methods from polymer physics— of the encapsidation of generic polyelectrolytes by oppositely charged capsid proteins under conditions of thermodynamic equilibrium produces the phase diagram shown in Figure 7.

The vertical axis (c) represents the capsid protein concentration and the horizontal axis [c] the ambient salt concentration. At high salt concentrations, electrostatic interactions are screened out and the hydrophobic protein-protein attraction is the dominant interaction. With increasing capsid protein concentration, empty capsids ("A Phase") form out of a phase ("B Phase") with only protein oligomers. As the salt-concentration is lowered and the strength of electrostatic interactions increased, a multi-critical point is reached beyond which polymer-filled capsids form with increasing protein concentration ("C Phase"). The amount of encapsidated polyelectrolyte material corresponds to about twice the total surface charge. This charge inversion is an example of the well-known "overcharging" phenomenon of aqueous electrostatics [14]. Perhaps by coincidence, the net RNA charge of a $T = 3$ virus —about 4000— really is roughly twice the net capsid protein charge of a $T = 3$ virus —about 1800 though this rule does not hold for larger viruses with $T > 3$.

The multi-critical point corresponds to what is known in polymer science as a "desorption threshold" [23]. That is the threshold where the adhesion energy of a polymer to a surface is exceeded by the conformational entropy loss due to surface confinement, so the polymer will not adhere anymore to the surface.

The surprise is that, according to this thermodynamic analysis, the osmotic pressure Π of the polyelectrolyte material inside the shell should be *close to zero*. We can illustrate this important point by a plot of the Gibbs free energy $\Delta G_{\rm p}$ as a function of the monomer (*i.e.* nucleotide) concentration $\langle\phi\rangle$ trapped inside the shell (Fig. 8).

The negative of $\Delta G_{\rm p}$ is the polymer contribution to the capsid assembly energy as it appears in the Law of Mass Action for capsid formation out of a solution of protein oligomers so if $\Delta G_{\rm p}$ is positive, then self-assembly would

produce predominantly empty capsids. As the nucleotide concentration $\langle\phi\rangle$ increases from zero, $\Delta G_{\rm p}$ is negative and decreases. The correlation length ξ of the semi-dilute polymer system decreases as well, as $1/\langle\phi\rangle^{1/2}$. The free energy reaches a minimum when the correlation length is comparable to the capsid radius R, and then starts to increase. Eventually, $\Delta G_{\rm p}$ turns positive. Beyond $\langle\phi\rangle^*$ the osmotic pressure $\Pi \equiv -{\rm d}\Delta G_{\rm p}/{\rm d}V|_N$ —with N the number of monomers and V the capsid volume— starts to rise and essentially corresponds [24] to the slope of $\Delta G_{\rm p}$ with respect to $\langle\phi\rangle$. Aggregates formed by thermodynamic self-assembly are in general dominated by structures at or near a maximum of assembly free energy -$\Delta G_{\rm p}$. Since the slope is zero at that point, filled capsids should have negligible internal osmotic pressure.

Could this "$\Pi = 0$" result for a model self-assembly system hold for actual $T = 3$ ss RNA viruses? The notion certainly contradicts our earlier 1–10 atm estimated osmotic pressure, but an intriguing experiment by the UC Irvine group [25] indicates that it may not be unreasonable. They found that the ss RNA genome of the *Satellite Tobacco Mosaic Virus remained condensed even after all of the enclosing capsid proteins were enzymatically digested*, leaving behind only the peptide sequences of the capsid proteins that were associated with the RNA genome (*e.g.*, the N termini of CCMV capsid proteins). This would be consistent with an internal pressure close to zero and it indicates that the N terminus tail of a capsid protein can act as a powerful RNA condensing agent. It is known that the osmotic pressure inside ds DNA bundles can drop to zero in the presence of powerful condensing agents [13].

A functional objection against the $\Pi = 0$ idea is the fact that certain $T = 3$ viruses such as FHV are known to release their RNA genome by extrusion through a pore in one of the twelve pentamers of the capsid shell [5]. If the genome of ss RNA viruses really is not pressurized, then why would the RNA molecules want to leave the shell? In other cases, viruses seem to disassemble in the same physico-chemical environment as where they were assembled, *i.e.*, the cytoplasm of the host cell. Generally, why would an RNA molecule that drives virus assembly by free energy minimization *spontaneously emerge* from the capsid after host infection when the assembled state constitutes a free-energy minimum of the sort shown in Figure 8? We do not really know, but part of the answer may be that the virus that releases its genome is *not* the same as the virus that formed earlier by self-assembly. Many viruses undergo a *maturation* process. For FHV for instance, that involves the scission of a covalent bond of the capsid proteins [5]. It may be that this reduces the nucleo-protein affinity sufficiently to produce a non-zero osmotic pressure that could drive genome release, which raises the intriguing possibility that ss RNA viral genomes actively *upregulates* the osmotic pressure inside the virus before release. Another part of the answer may be that the assembly and disassembly environment is *not* the same. For instance, viral assembly can take place in membrane-enclosed compartments budded of the Endoplasmatic Reticulum or the Golgi apparatus. These compartments could have their own physico-chemical conditions favorable for assembly, for instance in terms of the acidity level. Many $T = 3$ viruses, such as CCMV and the Tymoviruses, will spontaneously assemble in a certain pH/salinity range but spontaneously disassemble in a different pH/salinity range.

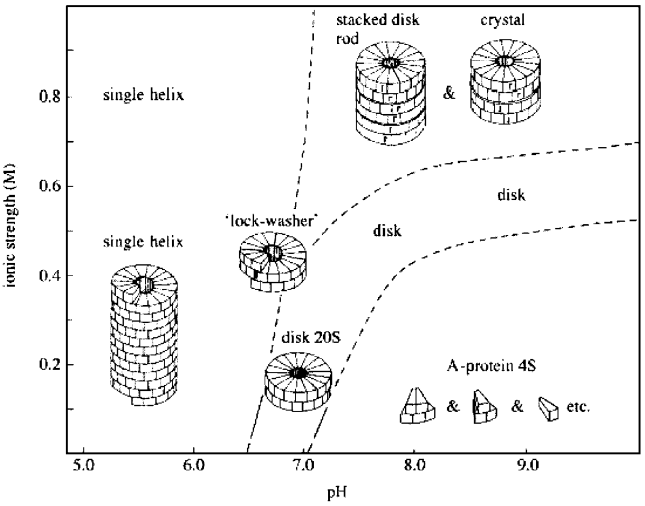

Fig. 9. Phase diagram of *Tobacco Mosaic Virus* capsid proteins with no RNA (from Ref. [28]). The wedge-shaped protein monomers can assemble into oligomers at high pH, or 20S double-disks at physiological pH (ionic strength = 0.1 M; pH = 7). At lower pH levels, helical cylinders form with the same structure as the capsids of *Tobacco Mosaic Virus* viruses.

We have seen that viral ss RNA is an active player in the theater of viral assembly: it provides the energy source for its own encapsidation but our arguments so far involved only the generic affinity between RNA and capsid proteins. The structural organization of RNA inside a capsid indeed appears to be mostly determined by this non-specific interaction. For instance, the dodecahedral cage of the *Pariacoto* viral genome forms just well when *non-viral* RNA is encapsidated [26]. On the other hand, viral encapsidation of RNA is highly selective. RNA molecules of similar viruses infecting the same cell are correctly packaged in their proper capsids [27]. Indeed, when the critical stem-loop sequences we discussed earlier are removed, the packaging efficiency is greatly reduced. So how can structural features, like these stem-loop sequences, that do not seem to affect the free-energy balance very much have such a dramatic effect on the assembly process? To address that question we must turn to the *kinetics* of viral assembly.

In his classical study [28] of *Tobacco Mosaic Virus* (TMV), Aaron Klug proposed a model for assembly initiation. He measured the self-assembly phase-diagram of TMV capsid proteins, see Figure 9, and showed that under physiological conditions only double-disk protein oligomers form. Fully formed —but empty— helical TMV capsids will assemble as well but only at non-physiological (lower) pH levels. If, under physiological conditions, viral RNA is added to the solution then —just as for the CCMV case— infectious TMV viruses form.

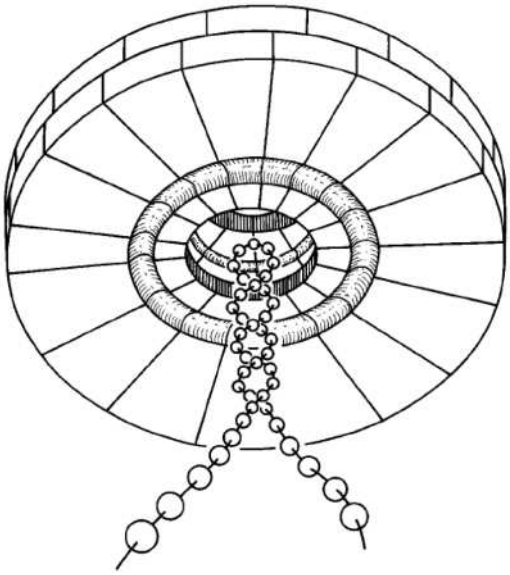

Fig. 10. Nucleation of TMV assembly is believed to start with the insertion of a stem-loop sequence in the central hole of double-layer oligomer (from Ref. [28]).

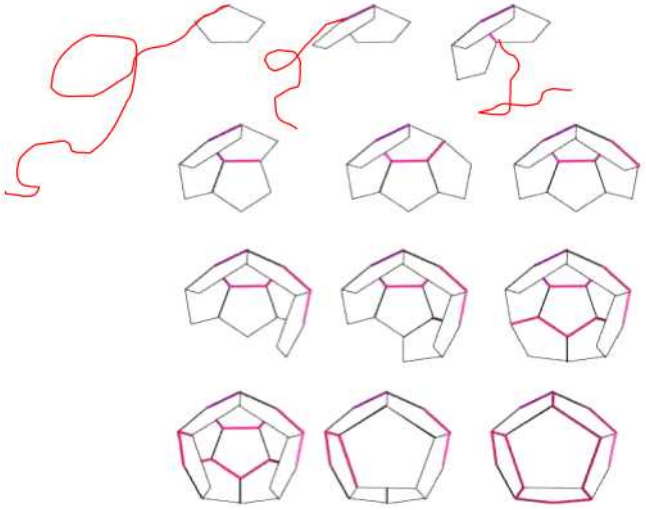

Fig. 11. Growth scenario of a viral capsid, constructed from five-fold symmetric protein units (*e.g.*, pyramids of 15 proteins), from a linear template (red online line). The template can be the main chain of a branched ss RNA molecule.

Klug proposed that one or more of the stem-loop sequences of the secondary structure of TMV RNA molecules must have a high affinity for the hole at the centre of the protein disks, which is lined by positively charged residues. After the hairpin is inserted into the disk, as shown in Figure 10, electrostatic repulsion between this disk and a second (RNA-free) disk is reduced sufficiently for a two-disk assembly to form.

After all, the phase diagram of Figure 9 shows that under physiological conditions we are close to the phase-transition line separating the "disk phase" from the "helical phase" which means that the free energy difference (per protein) between the two structures must be small. After the second disk has attached, the RNA molecule is threaded through the central hole of the growing helical capsid, followed by the attachment of a third disk. The process continues until the full length of the RNA molecule has been consumed. The length of the RNA molecule determines the length of the TMV cylinder so the RNA molecule acts as a *linear growth template* for the assembly of the TMV virus. The "nucleation" step is controlled by the interaction of a specific stem-loop sequence with a protein oligomer, while the "growth" step is controlled by the non-specific electrostatic affinity between generic RNA nucleotides and capsid proteins. In the spirit of Ribozyme activity, viral RNA manipulates the *activation barrier* of its own encapsidation!

Can the RNA genome of spherical $T = 3$ ss RNA viruses as well as act as a growth template? We should again distinguish initiation from growth. Assembly initiation depends in that case not only on the stem-loop recognition sequences. RNA molecules of many plant and retroviruses have, at their $3'$ end, a sequence that resembles a host tRNA molecule [29] (sometimes described as a *pseudo-knot* that ties up the end of the RNA molecule). This tRNA sequence is the initiation site for RNA duplication by the Replicase protein that we discussed earlier, and it can also function as a telomere. It has been discovered that this tRNA sequence plays a crucial role during viral assembly [30]. Specifically, it appears to act as a *condensation surface* for the formation of capsid proteins oligomers (pentamers). This role of the tRNA segment is only of a transient nature: inside an assembled capsid it appears to be no longer associated with the capsid wall. In other words, it only lowers the activation barrier for the initiation of capsid assembly but it does not contribute to the final assembly free energy. We see that $T = 3$ viral RNA molecules with tRNA sequences really act as Ribozymes in the sense that they carefully control the activation barrier for assembly initiation.

Now about the growth step. Recall that if the secondary structure of a ss RNA viral genome has a main-chain/side-chain organization, then a partially icosahedral folding pattern can be generated by letting the main chain perform a Hamiltonian walk or cycle along the edges of a dodecahedral cage (see Fig. 5). This folding pattern actually provides a scenario for the main chain to act like a linear growth template. Let us assume for simplicity a solution of RNA molecules and five-fold symmetric oligomers, such as the pentagonal pyramids of FHV. Assume that a pentameric pyramid has attached to or nucleated at the $3'$ end of the genome. Because of the (partial) electrostatic neutralization of the pyramid, a second pyramid can attach to the first pyramid with the main chain of the genome lying along the common shared edge (see Fig. 11), directly analogous to the disk-disk assembly of TMV.

We can then add a third pentagonal pyramid, with the main chain covering either one of the two shared edges between the third pyramid and the two earlier pyramids. A side-branch would have to cover the other shared edge (not shown). We can continue this assembly game provided we watch out that the main chain does not run into itself, which would terminate the process. This is

avoided by demanding that the main chain visits every vertex only once. For the genome, this leads to either a Hamiltonian path or the Hamiltonian cycle shown in Figure 5. For the proteins, this assembly scenario produces a compactly growing, curved shell that eventually closes on itself as shown in Figure 11. Interestingly, the Hamiltonian cycle scenario produces nearly the same growth intermediates as a thermodynamic model for assembly intermediates of spherical capsids built from five-fold symmetric units without any considerations involving the viral genome [31].

In summary, viral RNA molecules have emerged from this overview as macromolecules with an amazing ability to manipulate their own physical condition. The physical considerations presented in the overview have been of a rather naive and qualitative level. More sophisticated physical arguments could produce interesting results. For instance, one of the other contributors to this "Focus Point" has shown [32] that a *designed* RNA structure, with a certain fraction of specific complementary base-pair sequences holding a specific secondary structure in place, will undergo a continuous melting transition into a high-entropy molten globule state as the pairing specificity is reduced. The dodecahedral cage of Figure 6 actually may be such a designed structure, held in place by a certain amount of complementary pairing. The condensation of the genome during encapsidation should perhaps be described as a continuous *freezing* transition from a high-entropy molten state to a low-entropy designed structure.

We have focused in this overview on a special group of viruses (ss RNA $T = 3$ viruses), which does not do justice to the great diversity of encapsidation and release schemes encountered in the viral kingdom. Particularly interesting in this respect is the bizarre natural history of the HIV-1 virus, the focus of much current research. The diversity of schemes employed by different viruses underscores the structural and functional plasticity of ss RNA molecules, and suggests the possibility of an early "RNA World", where RNA molecules both were the repository of genetic information and central actors on the stage of molecular evolution.

I have greatly benefited from extensive interactions with my collaborators William Gelbart, Charles Knobler, Joseph Rudnick, and Paul van der Schoot as well as from discussions with S. Grosberg, C. Henley, J. Johnson, L. Lavelle, D. Nelson, R. Phillips, D. Reguera, N. Toan, R. Zandi, and A. Zlotnick. I would like to thank the NSF for their support under DMR Grant No. 0404507.

References

1. T.S. Baker, N.H. Olson, S.D. Fuller, Microbiol. Mol. Biol. Rev. **63**, 862 (1999).
2. R.G. Webster, A. Granoff, *Encyclopedia of Virology Plus CD-ROM* (Academic Press, 1995).
3. C.K. Biebricher, M. Eigen, J. McCaskill, J. Mol. Biol. **231**, 175 (1993).
4. P.P. Hung, C.M. Ling, L.R. Overby, Science **166**, 1638 (1969).
5. J. Johnson, R. Rueckert, in *Structural Biology of Viruses*, edited by Wah Chiu, Roger M. Burnett, Robert L. Garcea (Oxford University Press, New York, 1997).
6. Not only do small RNA viruses have nucleotide densities that are comparable to that of crystalline RNA, but there are examples where the nucleotide density even *exceeds* that of crystalline RNA, such as the *Rhinovirus* which has a packing density of 1.69 cubic angstrom per dalton.
7. H. Tsuruta, V.S. Reddy, W.R. Wikoff, J.E. Johnson, J. Mol. Biol. **284**, 1439 (1998).
8. T. Li, Z. Chen, J.E. Johnson, G.J. Thomas, Biochemistry **131**, 6673 (1992).
9. G. Ribitsch, R. De Clercq, W. Folkhard, P. Zipper, J. Schurz, J. Clauwaert, Z. Naturforsch. C **40**, 234 (1985).
10. M. Zuker, Nucleic Acids Res. **31**, 3406 (2003).
11. J. Rudnick, R. Bruinsma, Phys. Rev. Lett. **94**, 038101 (2005).
12. L. Tang, K.N. Johnson, L.A. Ball, T. Lin, M. Yeager, J.E. Johnson, Nature Struct. Biol. **8**, 77 (2001).
13. D.C. Rau, V.A. Parsegian, Biophys. J. **61**, 246 (1992).
14. For a review see W. Gelbart, R. Bruinsma, P. Pincus, V.A. Parsegian, Phys. Today, September (2000) p. 38.
15. A. Evilevitch, L. Lavelle, C.M. Knobler, E. Raspaud, W.M. Gelbart, Proc. Natl. Acad. Sci. U.S.A. **100**, 9292 (2003).
16. C. Knobler, private communication.
17. D.E. Smith, S.J. Tans, S.B. Smith, S. Grimes, D.L. Anderson, C. Bustamante, Nature **413**, 748 (2001).
18. J.A. Speir *et al.*, Structure **3**, 63 (1995).
19. Actually, the edges of the dodecahedral cage in Figure 6 are *low-curvature* borders between the pyramids.
20. K.W. Adolph, P.J. Butler, Philos. Trans. R. Soc. London, Ser. B **276**, 113 (1976).
21. J. Johnson, J. Tang, Y. Nyame, D. Willits, M. Young, A. Zlotnick, Nano Lett. **5**, 765 (2005).
22. P. van der Schoot, R. Bruinsma, Phys. Rev. E **71**, 061928 (2005).
23. F. von Goeler, M. Muthukumar, J. Chem. Phys. **100**, 7796 (1994).
24. The physical meaning of the negative slope for $\langle\phi\rangle$ less than $\langle\phi\rangle^*$ in terms of the pressure is that, for ξ less than R, the Gibbs free energy scales as the *surface area* of the shell. By Laplace's Law, this effective negative surface tension indeed should produce a negative pressure. However, in reality, the negative tension produced by the polymer adsorption is exactly cancelled by repulsion between the capsid proteins leading to a zero total surface tension (as in the Shulman criterion for surfactant bilayers). The actual pressure is thus zero for $\langle\phi\rangle$ less than $\langle\phi\rangle^*$. For ξ greater than R, the Gibbs free energy no longer scales as the surface area and the capsid wall develops a true tension.
25. J. Day, Y.G. Kuznetsov, S.B. Larson, A. Greenwood, A. McPherson, Biophys. J. **80**, 2364 (2001).
26. K.N. Johnson, L. Tang, J.E. Johnson, L.A. Ball, J. Virol. **78**, 11371 (2004).
27. C.M. Ling, P.P. Hung, L.R. Overby, Virology **40**, 920 (1970).
28. A. Klug, Philos. Trans. R. Soc. London, Ser. B **354**, 531 (1999).
29. M.H. Kolk, M. van der Graaf, S.S. Wijmenga, C.W. Pleij, H.A. Heus, C.W. Hilbers, Science **280**, 434 (1998).
30. P. Annamalai, A.L. Rao, Virology **332**, 650 (2005).
31. A. Zlotnick, J. Mol. Biol. **59**, 241 (1994).
32. R. Bundschuh, T. Hwa, Phys. Rev. Lett. **83**, 1479 (1999).